Open Access to Grey Resources

INIST-CNRS
Nancy, France
5-6 December 2005

SEVENTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

Conference Proceedings

ISSN 1386-2316

Host and Sponsors

INIST-CNRS, France
The Institute for Scientific and Technical Information;
The National Centre for Scientific Research

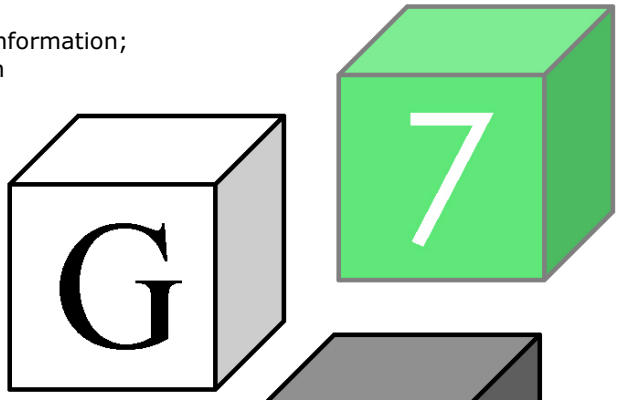
EU Bookshop, Luxembourg
Office for Official Publications of
the European Communities

CORDIS, Luxembourg
Community Research & Development
Information Service

EBSCO, USA
EBSCO Information Services

NYAM, USA
The New York Academy of Medicine

City of Nancy, France
Region of Lorraine



Program Committee

INIST-CNRS	Institute for Scientific and Technical Information	France
CNR-IRPPS	Consiglio Nazionale delle Ricerche	Italy
GreyNet	Grey Literature Network Service	Netherlands
JST	Japan Science and Technology Agency	Japan
OSTI-DOE	Office of Scientific & Technical Information; Dept. of Energy	USA
UCI	University of California, Irvine	USA
UiB	University of Bergen	Norway
UvA	University of Amsterdam	Netherlands

TextRelease
GL7 Program and Conference Bureau
Beysterveld 251, 1083KE Amsterdam, Netherlands
www.textrelease.com ▪ info@textrelease.com ▪ T/F +31 (0)20-672.1217

CIP

GL7 Conference Proceedings
Seventh International Conference on Grey Literature : Open Access to Grey Resources, 5-6 December 2005 / compiled by D.J. Farace and J. Frantzen. – Amsterdam : TextRelease, January 2006. – 210 p. ; 30 cm. – Author Index. – (GL-conference series, ISSN 1386-2316 ; No. 7)

The conference proceedings contain the full text of some twenty-eight papers presented during the two-days of conference sessions, panels and roundtables. The print copy has been standardized in font, size and format for general layout and appearance. Included is a List of Participating Organizations and an Author Index.

ISBN 90-77484-06-X €70,00

Foreword

Open Access to Grey Resources

Open access to information is the key to knowledge, both in its generation and transfer. This is based on the principles of science, the economics of free enterprise, and the management of valued resources in a global environment. Recently, open access to grey resources has been mandated by the grey literature community. And, this mandate must be analysed and understood in its multiple aspects – thus enabling implementation in the information policies and activities of organizations in both public and private sectors.

The Seventh International Conference on Grey Literature focuses on this theme and approaches it via the lines of communication and logistics. Often, the elements of communication are so embedded and pervasive that they are taken for granted; however, GL7 must look at information systems and networks, partnering and OAI, as well as, curriculum development and current research in this field. Likewise, the logistical components require rethinking and adaptation. Here, we must focus on repositories and collections, information retrieval and document delivery, as well as quality assessment issues.

Dr. Dominic J. Farace
Grey Literature Network Service

Amsterdam,
January 2006

GL7 Conference Sponsors

INIST
Institut de l'Information Scientifique et Technique;
Centre National de Recherche Scientifique, France

City of Nancy, France
Region of Lorraine, USA

CORDIS, Luxembourg
Community Research & Development Information Service

EBSCO, USA
EBSCO Information Services

EU Bookshop, Luxembourg
Office for Official Publications of the European Communities

NYAM, USA
The New York Academy of Medicine

GL7 Program Committee Members and Organizations

Dr. Joachim Schöpfel *(Chair)*
Institut de l'Information Scientifique et Technique, France

Anne Asserson
University of Bergen, Norway

Takaharu Atago
Japan Science and Technology Agency, Tokyo, Japan

Prof. Dr. Albert K. Boekhorst
University of Amsterdam, Netherlands

Deborah Cutler
Dept. of Energy; Office of Scientific & Technical Information, USA

Dr. Dominic J. Farace
Grey Literature Network Service, Netherlands

Julia Gelfand
University of California, Irvine, USA

Daniela Luzi
National Research Council, Italy

Contents

OPENING SESSION

Chair: J. Schöpfel, Institut de l'Information Scientifique et Technique, INIST/CNRS - France

SESSION ONE

Chair: D. Luzi, National Research Council, CNR/IRPPS - Italy

PANEL ONE

Chair: J. Gelfand, University of California, Irvine, UCI – United States

SESSION TWO

Chair: Anne Asserson, University of Bergen, UiB - Norway

PANEL TWO

Chair: C. de Blaaij, Library of Zeeland, Netherlands

R O U N D T A B L E S

THESES AND DISSERATIONS

Facilitator: C. Stock, Institut de l'Information Scientifique et Technique, INIST - France

REPOSITORIES AND COLLECTIONS OF GREY LITERATURE

Facilitators: K.G. Jeffery, CCLRC – Rutherford Appleton Laboratory - United Kingdom
G.E. Siegel, Portland State University, PSU – United States

QUALITY ASSESSMENT OF GREY LITERATURE

Facilitator: P. De Castro, Istituto Superiore di Sanità, ISS - Italy

Welcome Address

Raymond Duval, Director INIST-CNRS
Institut de l'Information Scientifique et Technique, INIST
Centre National de Recherche Scientifique, CNRS France

Good morning,

Ladies and Gentlemen,

On behalf of the French National Scientific Research Centre CNRS and the Institute for Scientific and Technical Information INIST, I would like to welcome you to the 2005 conference on "Open Access to Grey Resources".

Dissertations, conference proceedings and reports are an important part of scientific information. Nevertheless, they are often difficult to find and to obtain. As you will know, our institute is deeply involved in the collection and dissemination of grey literature, and our holdings contain some hundred thousands of grey documents.

Today, the challenge for scientific information is the transition from print to digital resources. The technological and legal environment undergoes a rapid change, and a new economic model emerges.

The traditional role of INIST was to organize the access to scientific information through databases and document supply for the French public and corporate scientific community. In order to preserve and develop our position in the new digital environment, we have to adapt our technology, services and organisation.

Today, we host scientific portals and negotiate licences for online resources and digital archives. We develop new products in e-publishing, multimedia, terminology and scientometrics, we participate in national and international standardization, and we help other structures and colleagues through professional training, consulting and assistance.

On the other hand, we play a central part in the national initiative for open access to scientific information. INIST signed the Berlin Declaration in behalf of the National Research Centre, and we contribute to the development of the French institutional research repository.

Open access to all kind of grey literature became a major topic of your discussion and work since the Amsterdam conference two years ago. How to archive reports, proceedings and so on in institutional repositories? Which metadata should we select to facilitate their retrieval? How to make them available and accessible for the end user? These are your questions, and this is our concern and business, too.

Therefore, I think it was an excellent idea to choose the theme of open access to grey resources for the seventh conference on grey literature, and it is fitting that this conference should be held at Nancy, a city that is to become a national "pole of competence and excellence" in scientific and technical information.

I am sure that you will find the discussions over the next two days to be stimulating and informative, and I welcome you to this conference!

Open archives and SIGLE participation in Italy: Is there a subtle link between the two experiences?

Daniela Luzi, Rosa Di Cesare and Roberta Ruggieri
Istituto di Ricerche sulla Popolazione e le Politiche Sociali , IRPPS-CNR, Italy

Abstract
This paper proposes on the one hand to outline Italian initiatives regarding open archives, with a qualitative rather than quantitative analysis, focussing in particular on data providers and on organisation and initiatives that lead to integration that encourage the development of IRs. On the other hand, it proposes to analyse the 20 years of Italian input to SIGLE in order to verify if and to what extent the Italian producers participating in SIGLE have contributed to the fostering of new ways of sharing and diffusing information sources produced within their organisations.

1. Introduction

Why have we decided to verify the existence of the weakest of links between the creation of Italian open archives and the Italian participation in the European SIGLE system? There are numerous differences between the two documentation systems, including the date of their development, the methods used for data collection and the technology that has been employed.

The SIGLE project (System of Information for Grey Literature in Europe) got underway in 1980 and was funded and launched by the then European Community. EAGLE (European Association for Grey Literature in Europe) came into existence in 1985 when the database was fully operative [1]. SIGLE adopted an operational system that was cutting edge for its time, with the collection and indexing of GL in the hands of national reference centres. The data was successively collected and elaborated centrally by a technical unit that provided a fee-based access to the database. Nowadays this would be termed a distributed system with centralised access to data.

Open archives, on the other hand, were the initiative of individual researchers (the most famous being Ginsparg, Harnard, Krichel), who, in a brief period of time, managed to collect a large number of documents from colleagues and make them freely accessible via Internet [2]. This gave rise to a movement, a new philosophy for the diffusion of scientific literature, known as the Open Archive Initiative (OAI). In effect, one of the most significant results of the development of e-print archives has been the consensus received from scientific institutions and the support of libraries that belong to them leading to the establishment of the Institutional Repositories (IRs) [3]. In fact, institutional commitment to making available and preserving the scientific documentation produced by each institute along with the active participation of libraries and that of academics can create ideal conditions for access to scientific literature without economic and technological barriers.

However, the SIGLE initiative and open archives - IRs in particular - have some elements in common. These can be traced above all in their common objectives, but extend to initiatives to guarantee the success of both documentation systems. Looking first at the objectives, the open archives - and in particular the Institutional or disciplinary repositories - aim to make available the internal documentation of an institute or a discipline. Similarly, the various national reference centres aim to provide a GL production framework that faithfully reflects the production in each member state. Setting aside the opportunity of a national information policy, historically lacking in Italy, conditions necessary for the successful performance of the two documentation systems are very similar. Both require the setting up of a network that is, particularly for the repositories, appealing to the various scholarly communities and that *convinces* them of the advantages of depositing their work in open archives. For the SIGLE national reference centres too, it was necessary to identify the GL producing institutes and either have them provide a constant stream of documents and/or make them actively participate in GL cataloguing. In order to do so, GL producers had to be convinced of the information value of GL and of the necessity for access to such documents via participation in the SIGLE European database. In both cases organisational and cultural changes are needed in order to assure the success of the initiatives.

A further common element of both documentation systems is precisely that both distribute GL. SIGLE only manages GL, while the open archives deal with GL together with conventional literature. While Gelfand [4] is correct in affirming that GL has found its home in the repositories, it is exactly because of the co-existence of GL and conventional literature that repositories provide an *ideal, complete coverage* of the research output of an institute or disciplinary community. It is in this way that repositories can become an alternative and/or complementary source to commercial publishers, responding to request for free access to scientific products.

In Italy, IRs are still in an initial phase, though several projects exist for the development of both data and service providers. The Italian backing for the OAI comes mostly from Universities [5], which almost unanimously adhered to the Berlin Declaration [6]. Support was co-ordinated by the Libraries Panel of the Forum of Deans of Italian Universities (Conferenza dei rettori delle università italiane - CRUI) and this is an important sign for future development of national initiatives.
As regard SIGLE, whose existence spans more than twenty years, the changes of information diffusion over recent years have certainly influenced both the Italian contribution and the composition of the entire European database from whom important countries, the UK and the Netherlands, have withdrawn.

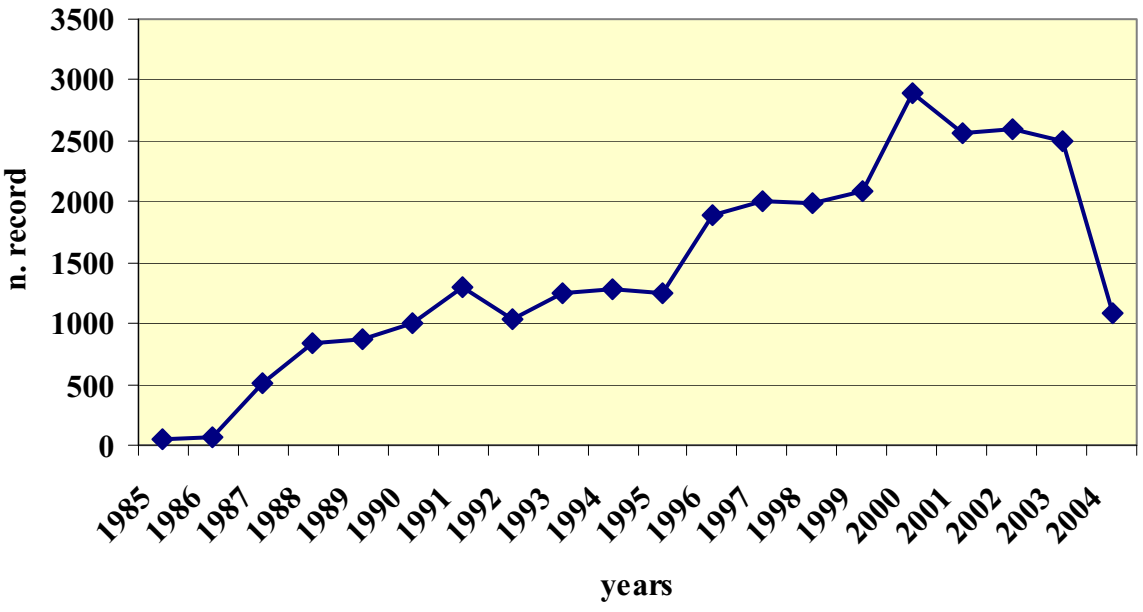
This paper proposes on the one hand to outline Italian initiatives regarding open archives, with a qualitative rather than quantitative analysis, focussing in particular on data providers and on organisation and initiatives that lead to solutions and integrations that encourage the development of IRs. On the other hand, it proposes to analyse the 20 years of Italian input to SIGLE in order to verify if and to what extent the Italian producers participating in SIGLE have contributed to the fostering of new ways of sharing and diffusing information sources produced within their organisations.

2. Italian participation in the SIGLE system

The SIGLE database was established in order to trace and make available GL produced in Europe. When, in 1985 the EC support came to an end, EAGLE was set up and to this day still handles the database via the National Reference Centres of the member states. The Italian National Reference Centre (INRC) is the Central Library of the Italian National Research Council (CNR), which joined SIGLE in 1985 [7].

The Italian GL producers in SIGLE constitute the reference point for observing analogies and/or differences with the recent establishment of Open archives. Consequently, this analysis firstly offers a synthetic view of the Italian contribution from 1985 to 2004, proceeding to focus on the characteristics of Italian producers; on the dimensions of their contribution and on the types of documents they produce. The data under analysis were obtained from the SIGLE CD-Rom, updated to December 2004 [8].

Fig. 1 - Italian input to SIGLE data base



The total Italian contribution to the SIGLE database amounts to 29,063 documents. Figure 1 shows that, after an initial increase, the rate of contributions stabilised. The number of documents sent to the database started to increase again in 1995, peaking in the year 2000 with an input of 2899 documents. Thereafter input declined, gradually at first, but falling sharply to a mere 1082 documents in 2004, the lowest figure for a decade.

2.2 The expansion of the database: from regular to occasional producers

The aims underlying the INRC policy are the result of twenty years' activity and elaboration in accordance with EAGLE members. This activity, in turn is the result of collaboration between the INRC and Italian GL producers. The contribution of the Italian Library Association (AIB), in backing the database should not be overlooked. In 1989 the AIB, amongst other things, promoted a survey on GL producers [9, 10], which revealed that the most prolific producers of GL were fundamentally public research institutes that were better organised in collecting and managing GL documents. These institutes constitute the historical nucleus of the Italian database and guarantee constant input (cf. Tab. 1). The initial work of the INRC was based on these institutes. Broadly speaking, this phase came to an end in 1995 [11, 12].

Table 1 shows the growth rate in the number of documents submitted to the Italian database and groups producers according to their activity and institutional mandate. The Italian contribution to the database increased by 1.7 from 1995-2004 compared to the previous decade. This increase refers above all to documents produced by universities (34.6), with a minor increase on the part of government institutes (19.0). Research institutes' contribution grew a negligible 0.2, but this has to be seen in the light of their contribution during the initial phase, when their input was higher than that of other institutes. The data for research institutes is not indicative of a *physiological decline* in their participation in the database; the explanation is to be found elsewhere and will be discussed further below.

Tab. 1 - Italian input by GL producers in the period 1995-2004 compared to1985-1994

GL Producers	1985-1994	1995-2004	growth rate
Universities	306	10,906	34.6
Research institutes	7,702	9,450	0.2
Governmental institutes	39	781	19.0
Foundations and research centres	153	682	3.5
Associations & Societies	423	413	0.0
Total	**8,196**	**22,303**	**1.7**

Tab. 2 - Italian input by GL range of documents and producer types

N. of documents	Universities	Research institutes	Governmental institutes	Associations & Societies	Foundations & Res. centres	Other	Total	%
< 10	10	20	93	78	32	29	262	71.8
11 -200	37	9	21	5	5	0	77	21.1
201-400	11	1	0	0	1	0	13	3.6
401-1000	6	1	0	0	0	0	7	1.9
1001-4000	1	3	0	0	0	0	4	1.1
> 4001	0	2	0	0	0	0	2	0.5
	65	**36**	**114**	**83**	**38**	**29**	**365**	**100.0**

The database includes 365 Italian producers (Tab. 2) [13]. It is worth noting that both universities and research institutes contribute a significant number of documents. Combined, they distribute over 1000 documents via SIGLE. Nevertheless, the majority of institutes (71.8%) contribute less than 10 documents to the Italian SIGLE. This figure is to be attributed to the expansion of the database to include a larger number of producers and new categories. It is no coincidence that the period 1995-2004 (cf. Tab. 1) saw the INRC seeking new producers and sources of information. In accordance with the EAGLE policy, institutes' involvement in the database was encouraged. Turning to new sources of information, the INRC began to include data from bibliographies such as those in PhD theses that became available at that time. The INRC also made use of the web, which in that period became an important source of GL documents [14]. The results of this pro-active approach can be seen both in the expansion of the database to include new producers and in the considerable increase in the numbers of universities involved.

2.3 The network of Italian producers

Table 3 shows that in terms of GL producing institutes, the composition of the Italian database is predominantly technical and scientific. The contribution of research institutes and universities is significant (respectively 56.2% & 36.8%). All the same, these two categories participate in different ways. Research institutes were present in the initial phase and remained constant up to 1994. An inversion took place over the last five years, with the number of documents amounting to 30.9%. This is considerably fewer than in previous periods, such as 1985-1989 for example, when their contribution reached 99.1%. Conversely, in the initial phase, universities contributed little, but have since greatly upped their input, from approx. 5.1% in the period 1990-1994 to 62.2% from 2000-2004. As to the other categories of producer, their presence is greatly inferior to that of research institutes and universities. Government institutes and Foundations have the same figures (2.7%), while Associations and Societies account for less than 1%.

Tab. 3 - Italian input by GL producers (1985-2004)

GL producers	1985-1989		1990-1994		1995-1999		2000-2004		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Universities	5	0.2	301	5.1	2,892	30.7	8,014	62.2	11,212	36.8
Research institutes	2,315	99.1	5,387	91.8	5,465	58.0	3,985	30.9	17,152	56.2
Governmental institutes	13	0.6	26	0.4	347	3.7	434	3.4	820	2.7
Foundations & Research centres	0	0.0	153	2.6	471	5.0	211	1.6	835	2.7
Associations & Societies	2	0.1	3	0.1	202	2.1	211	1.6	418	1.4
Others	0	0.0	0	0.0	43	0.5	28	0.2	71	0.2
Total	**2,335**	**100.0**	**5,870**	**100.0**	**9,420**	**100.0**	**12,883**	**100.0**	**30,508**	**100.0**

Tab. - 4 Italian input by document types (1985-2004)

Types of documents	1985-1989		1990-1994		1995-1999		2000-2004		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Acad. /progress reports	2,001	85.7	5,319	90.7	6,005	65.1	4,649	39.9	17,974	61.8
Dissertations	2	0.1	31	0.5	1,678	18.2	6,363	54.6	8,074	27.8
Conference Proceed.	238	10.2	385	6.6	838	9.1	131	1.1	1,592	5.5
Miscellaneous	0	0.0	0	0.0	644	7.0	490	4.2	1,134	3.9
Others	93	4.0	129	2.2	54	0.6	13	0.1	289	1.0
Total	**2,334**	**100.0**	**5,864**	**100.0**	**9,219**	**100.0**	**11,646**	**100.0**	**29,063**	**100.0**

Data on the document type confirms the technical and scientific nature of the Italian contribution to SIGLE (Tab. 4). Academic or progress reports amount to 61.8% of the total documents present in the database. Nevertheless, these have declined significantly recently, falling from a peak of 90.7% in the period 1990-1994 to 39.9% in the last five years. As has been mentioned, this can be explained by the declining participation of research institutes in the database. Conversely, in the first five years, the proportion of theses to the total number of documents is negligible, but has grown progressively over time. This can be explained by the CNR making use of the Bibliography of PhD theses first published in 1995 [15]. It was in this period that a new norm made it obligatory to deposit PhD theses at the two national central libraries in Rome and Florence.

Tab. 5 - Italian GL producers with over 200 documents (1985-2004)

GL Producers	1985-94	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Total
Ferrara Univ.	0	0	31	0	9	32	11	53	44	21	3	204
Bari Univ.	0	0	0	1	29	40	57	59	25	21	2	234
EUI	57	47	2	75	57	1	1	0	0	0	0	240
Turin Polytech.	0	8	0	1	14	8	37	41	65	50	15	239
Turin Univ.	0	0	0	1	19	47	62	63	29	33	7	261
Pisa Univ.	0	3	0	3	33	39	66	62	54	31	3	294
Enel	33	0	0	147	113	0	0	1	0	0	2	296
Palermo Univ.	0	0	0	2	29	23	34	26	49	67	67	297
Pavia Univ.	3	0	24	0	19	38	25	54	91	45	14	313
Milan Polytech.	0	0	3	0	16	11	60	27	136	65	3	321
Genoa Univ.	0	0	0	0	16	46	48	76	78	59	30	353
Bank of Italy	56	23	36	39	9	26	8	25	52	87	12	373
Catania Univ.	0	0	0	0	14	47	39	58	25	120	71	374
Naples Fed II Univ.	0		23	0	58	61	117	74	94	18	2	447
Florence Univ.	3	0	1	0	47	60	86	98	161	78	11	545
Siena Univ.	65	30	38	22	36	70	30	59	100	70	37	557
Padua Univ.	50	38	37	39	37	97	78	77	82	35	15	585
Milan Univ.	0	0	1	1	111	78	83	124	231	131	38	798
ISS	382	54	60	43	52	10	60	86	11	61	24	843
Bologna Univ.	102	27	34	20	111	111	147	102	120	149	74	997
Rome Sapienza Univ.	14	8	20	5	132	98	158	152	259	227	43	1116
INFN	700	92	86	185	69	74	67	113	0	0	0	1386
ENEA	755	29	43	77	48	37	69	75	99	149	33	1414
ISAS	941	77	103	91	67	110	1042	246	0	0	0	2677
IC	2807	395	274	229	187	261	26	303	116	160	108	4866
CNR	1889	243	785	678	332	186	129	156	261	160	131	4950

The list of institutes inputting more than 200 documents is included in Tab. 5, which seeks to:

- Compare overall input from the first decade (1985-1994) with that of the following period on an annual basis;
- Check whether and to what extent changes occurred in the database along with the development of the web;
- Outline the evolution of input from the year 2000, when various Italian institutes undertook OA initiatives.

We will consider first of all, the decline in contributions from research institutes which, broadly speaking applies to all institutes. The CNR and the International Centre for theoretical Physics (IC) significantly decrease their contributions (respectively from 243 in 1995 to 131 in 2004 and 395 to 108). Turning to the National Institute of Nuclear Physics (INFN) and the International School for Advanced Studies (ISAS), which are among the main Italian physics research institutes, we are looking at a total withdrawal from the database, given that no contributions have been sent since 2002. As will be seen in the following paragraph, in this period ISAS developed its own repository and prior to this, internal documents had been deposited in arXiv e-print archive [16]. INFN too had contributed to the establishment of e-print archives [17], both at institutional and individual researcher level. As for the other research institutes, the decline is probably due to documents being made available on their own home pages. Considering the institutes that were original contributors to the database, such as Istituto Superiore di Sanità (ISS) and ENEA, it can be seen that there was no major variation in their contribution. This confirms their adherence to SIGLE, despite a change in GL diffusion policy with full text documents now available on the web.
The data for universities confirms that, in most cases, input is concentrated in the last few years. It is also worth noting that most of the universities with a significant number of documents reported in table 5 have, as will be seen, established their own IRs.

Concluding this section on Italian participation in SIGLE, it is important to stress the INRC contribution to the bibliographic standardisation of GL documents, to the identification of GL producers as well as their part in monitoring the difficulties in building network of producers.

3. The Italian Institutional repositories

As already mentioned, open archives started out with the e-print archives, an initiative from the base that was organised and developed by groups of researchers. The aim was to accelerate and expand diffusion of their work (aims shared in the production of GL). The automatic document archiving system designed by Ginsparg at the Los Alamos National Laboratory, is an autonomous and self-sufficient model that places emphasis on direct communication without intermediaries between authors and readers of scientific documents. This method of communication radically changed the traditional linear phases of the scientific communication chain by circumventing the role up to then played by commercial publishers as well as libraries. However, the e-print archives expanded with the consensus of the open access movement, culminating in the Berlin Declaration on the part of the scientific institutions and libraries involved in the archives. This led the way to new forms of co-operation and integration of information resources and specialised know-how and fuelled the development of the IRs, that constitute the institutional commitment needed for access to and preservation of the scientific documentation produced by each institute. Although still in the project phase, the active involvement of libraries points to the ever-greater integration of traditional library information sources (opac, metaopac, digital libraries) with those of the open archives [18]. The integration is further enhanced by the constitution of archives based on a common metadata schema that enables integration with external information sources by means of the Open Archive initiative protocol for metadata harvesting (OAI-PMH) [19, 20].

Compared to other surveys [21, 22], our analysis has been based on direct consultation of open archives that have been developed in Italy. The main information source for access to these archives was Pleiadi [23, 24, 25] (Portale per la Letteratura scientifica Elettronica Italiana su Archivi Aperti e Depositi Istituzionali), a portal for Italian electronic scientific literature on open archives and institutional repositories. The portal was developed by the Italian Universities consortia CASPUR (Inter-university Consortium for the Application of Super-Computing for Universities and Research) and CILEA (Italian Lombardy Inter-university consortium for automatic elaboration). They are active promoters of the open archive movement and provide technical and scientific back up for the development and implementation of new systems. As can be seen, we started from a valid, constantly updated source, a service provider that operates as a link between the Italian initiatives and which provides useful information on international developments. Consequently, it was not necessary to select or check whether archives could be included in the definition of repositories [26, 27]. The emphasis was on organisational and operational models, the roles played by different actors and on the initiatives and projects that are both characteristic of the situation in Italy and merit attention. The systems were consulted in depth and this at times threw up incongruence in data or in their presentation and malfunctions but also revealed innovative and efficient ways of displaying and distributing the scientific output of an institute or a given discipline. As a result, the survey does not provide a quantitative analysis of the data, but seeks to stimulate reflection with a view to encouraging the development of the still young but promising open access initiative.

3.1 Italian data providers

There are 18 open archives operating in Italy, 14 of which are Institutional repositories, and 4 are disciplinary repositories. These initiatives are promoted by 10 universities, including one international university at Florence; three state research institutes (the CNR research area at Bologna, the National Institute of Geophysics and Volcanology and the Institute of Social Medicine); one management studies institute (Istituto di studi per la direzione e gestione di Impresa); and one international disciplinary repository, to which Italy is an active contributor. Considering that there are 77 universities and over 40 state research institutes and that the number of deposited documents fails to reflect the scientific output of each institute, it is clear that the situation in Italy is embryonic. Data for the expansion of the archives is not available and is difficult to verify. Nevertheless, our experience shows that since the initial consultations in the period from April-June, the number of documents has grown significantly [28]. Furthermore, as will be seen in following paragraph, some archives have concentrated so far on the collection of a single type of document.

Considering the quantity and type of documents, the situation is fairly heterogeneous. Although still in their early stages, the archives to be found range from those that already resemble genuine institutional archives, to those that have without doubt implemented systems following OAI criteria, but that still lack adequate visibility within their institute.

One of the most successful IR has been developed by the University of Bologna [29]. It is one of the first Italian universities that have implemented an institutional repository, divided in three separate archives One of these specialises in the diffusion of didactic programmes and materials (AMS Campus), the second specialises in the collection of "contributions deriving from the university's research activity" (AMS Acta)

and, the third collects "cultural contributions on a variety of subjects, deposited and distributed directly by the authors". Taken as a whole, this archive is the nearest to Lynch's broad definition of a repository [27], who distinguishes "*two views*" of IRs, and sees the second "as broadly housing the documentation of the intellectual work - both research and teaching - of the institution, records of its intellectual and cultural life" and which therefore includes datasets, video, learning objects, etc. among its documents.

Among those considered, the repository, which differs most from the others, is E-Lis [30]. This has been included in the list of Italian repositories as the Italian contribution is considerable, both technically and in terms of contributions. It is an international disciplinary repository on librarianship, information science, and technology, and has evolved from ReLIS/DoIS (Research in library and information science and documents in information science), the library science portal run by the MIMAS consortium of Manchester [31, 32]. It is supported by Spanish Ministry of Culture and is hosted by the AEPIC team CILEA. E-Lis has been running since 2003 and enables the self-archiving of electronic documents from 32 countries, not only from Europe and North America, but also from Asia, Africa, Central and Southern America and Oceania. The documents can be written in any language, but must include abstracts and keywords in English.

3.2 Archive consistency and type of documents

Data on the size of repositories provides useful insight on the extent to which the archive is representative of institutional research output and the degree of self-archiving. The number of documents deposited (Tab. 6) confirms that the Italian repositories are still in the initial phase. Even focussing on the larger repositories, the data are not encouraging. The two with more than 2000 items are an international university based in Italy [33] (European University Institute) and the E-Lis disciplinary repository that operates at international level. Worse, the former provides bibliographic descriptions but not the full text. Turning to E-Lis, the positive news is that more than 500 of the 2000 documents are self-archived by Italian personnel or directly inserted by researchers. ISAS [34], another international institute based in Italy, has a repository with over 1000 documents, but the link to the full text often connect to the publisher's homepage where papers can be accessed on subscriptions basis or charged singly. For a while ISAS and its researchers have been active in a wider diffusion circuit of internal documents via the e-print arXiv.gov. Consulting the repository one can often find two report numbers, which indicate that documents are deposited in both archives.

Tab. 6 - Documents contained in repositories

Number of documents	No. of Repositories
< 60	5
60 - 200	3
201 - 400	3
401 - 600	3
601 - 900	0
901 - 2000	2
> 2001	2

In all likelihood, the repositories with a medium number of documents (from 200 to 600) are the ones that have developed efficient organisational models and are fully operative. As elsewhere, the Italian situation is undergoing constant evolution. Since the international survey [21] and our first phase of data collection, the number of documents has increased in more than one institution. Moreover, two new repositories have recently been instituted (University of Pisa [35] and the National Institute of Geophysics and Volcanology [36]) which have a small number of documents, but which represent an important development as they have only been operative for a few months.

Clearly the validity of a repository depends not only on the number of documents it contains, but also on the extent of self-archiving carried out directly by the authors. For the latter it is difficult to obtain reliable data consulting the repositories themselves. In the recent international survey [21, 22] the percentage of authors personally depositing documents was not estimated to be very high. Consulting the

repositories, we noticed that when this information was available, the name of the author frequently did not correspond to that of the person who had deposited the document. This would seem to be the repositories' weak point. In Italy and elsewhere repositories need to advocate the advantages of academics and researchers practising self-archiving.
The type of document is a complex and fundamental issue. Every institute can decide on the type of document to be included in its archive. This is possible both with Dspace software, with which each community selects what collections to make available, and with GNU e-print software, which allows the inclusion of various document types along with the corresponding bibliographic descriptions. In the latter case there is clearly a difference between the type of document envisaged in a single repository and that, which is effectively used. Furthermore, GNU e-print software requires contributors to specify whether the document has been published, whether the document is in print or submitted and whether it has been peer reviewed. These are significant options which can provide useful insight on the extent to which repositories consider themselves free access channels, operating separately from commercial publishing, and the extent to which they reflect the quantity of GL produced inside the institute. Nevertheless, not all the repositories use or make this information mandatory. Consequently, it is often the case that when consulting an archive, no correspondence can be found between the data displayed and the real number of items contained in the archives. We consider this to be significant given that the challenge faced by repositories and their chances of success depend both on the presence of GL and conventional literature. A prevalence of GL would, on the one hand, further marginalize this types of document, on the other, fail to live up to the free access expectations of the published scientific literature.

Tab. 7 - Types of document held in Italian repositories

Repositories	Monogr./ b.chapters	Journal articles	Conference Proceed.	Reports	Pre-print	Thesis	Didactic materials	Others	Total	
									NO.	%
Bologna University Ams campus	--	--	--	--	--	--	94.6	11.5	464	5.7
Bologna University Acta	50.0	0.1	41.9	--	10.7	0.3	--	--	990	12.3
Bologna Univ. Miscellanea	0.2	0.1	0.3	0.4	3.6	--	--	5.2	25	0.3
Florence University	23.9	4.0	2.9	14.4	1.8	8.6	--	17.7	497	6.2
Messina University	--	6.4	1.9	--	--	--	--	--	269	3.3
Naples Federico II Univ.	1.1	--	0.1	--	25.4	--	--	--	49	0.6
Trento University	0.8	1.1	0.6	70.4	0.6	2.6	--	5.2	583	7.2
E-Lis	9.7	48.3	48.5	10.9	57.4	7.6	2.9	43.8	2,952	36.6
CNR Bologna research area	--	--	1.0	0.1	0.6	0.1	--	1.0	22	0.3
Institute of social medicine	13.8	2.5	2.2	--	--	--	2.5	15.6	222	2.8
STOA'	0.4	0.1	0.7	3.6	--	1.4	--	--	52	0.6
ISAS (two communities)	--	37.4	0.1	--	--	8.3	--	--	1,434	17.8
Padua University	--	--	--	--	--	42.9	--	--	309	3.8
Pisa University	--	--	--	--	--	7.2	--	--	52	0.6
Rome La Sapienza University	--	--	--	--	--	20.8	--	--	150	1.9
TOTAL	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**8070**	**100.0**
%	**5.8**	**45.5**	**21.7**	**8.8**	**2.1**	**8.9**	**5.9**	**1.2**	**100.0**	

Note: the National Institute of Geophysics and Volcanology, the University of Parma and the European University Institute are excluded from the list of repositories as data was not available in the archives.

Archives using GNU e-print contain an average of 10 different document types, with a maximum of 25 different types for E-Lis. In the archives which specialise in didactic, as is the case for the Bologna AMScampus, specific document types are identified and included such as programmes and bibliographies,

handouts, notes, lessons etc. In order to give an overview, the data in Tab. 7 has been elaborated by grouping similar types of documents [37]. Once again the results show considerable disparity between one archive and another. Articles from periodicals are above all to be found in the E-Lis (48.3%) and ISAS (37.4%) repositories. Otherwise, considering monographs and book chapters along with articles only Bologna University (50% of monographs and book chapters) and Florence University (23% and 4% of journal articles) provide a significant number of published documents. Proceedings constitute a significant type above all in the E-Lis archive and in that of the University of Bologna. Three universities (Padua [38], Pisa [35] and Rome [39]) have implemented their repositories by including only theses. Considering the different types of GL, (reports, preprint, theses e didactic materials), along with a part of the data for proceedings, not always commercially published, it emerges that GL is significantly present and tends to be equally distributed among the different archives.

3.3. Technical and organisational aspects

Similarly to IRs in other countries, Italian libraries play a central role in the development and management of open archives and, in some cases, avail themselves of the collaboration of interuniversity information technology units. In addition, CILEA and CASPUR are the two interuniversity consortia that in Italy provide technical backup for the launching of numerous initiatives. It may seem obvious that libraries are the main actors, particularly considering that user-friendly and open source software is used, yet the difficulty above all for IRs, is not so much in adapting the software to institutional situations, but more in supplying the repository so that it sufficiently represents the scientific activities of the institute. This implicates the selection of the documents produced and of potential interest for open access, encouraging researchers to submit their work, where necessary providing support for the submission of documents.

The library service running the depository is usually responsible for checking the accuracy of the metadata that are inserted in order to obtain a standardised bibliographic description that facilitates harvesting and seamless archive interoperability. The E-Lis disciplinary repository makes use of an editor, who manually checks the submitted documents "to confirm they fit our policy" and "ensure that inappropriate papers are not included". In this context too, there will probably be developments of new professional profiles and librarian specialisations. Consulting the archives, in cases where the identity of the person submitting the work is provided, confirms that library staff is involved in the submission of documents. This is clearly dependent on repositories still being in an early phase and the filling of a new archive can incentive authors to insert their documents directly.

As regard software, the majority of repositories, normally those established first, use the free GNU e-print software, whereas another four repositories use the MIT's Dspace. Rome University uses CDSware, Pisa NDTL. To shift from one repository to another encountering similar interfaces with similar search options and comparable help and instruction pages is user-friendly. This gives an impression of standardisation that suggests a combined initiative regardless of variations in operation.

3.3 Outlook for developments in evaluation and integration of data resources.

The development of IRs encounters difficulties in various fields connected with the ongoing uncertainty regarding intellectual property and copyright as well as peer-review and evaluation of self-archived publications. In this paragraph we briefly mention a few initiatives, some of which are still not beyond the planning phase, that address these issues. Some elements deserve a lengthier consideration that would be beyond the aim of this paper. These considerations could be taken as examples of good practices that would be useful for the growth of IRs.

Trento University is establishing the Polaris database [40] that is run by university library staff and that enables academics and researchers to input their own curricular data along with bibliographic descriptions of the work produced. The aim is to enable evaluation of scientific output by the competent bodies (Ministries or external refereeing bodies). GL is included among the types of document that academics can upload and in this way takes its rightful place among the evaluated scientific documentation of an institute or researcher.
Other universities envisage faculty committees or linking with external referees. Bologna University has signed an agreement with the Florence National Library for the voluntary depositing of its electronic scientific and didactic publications evaluated by a specific scientific committee. In this way the University is ahead of the new law on copyright, and, at the same time provides an extra, certified channel for the distribution of its publications.

Other examples of evaluation of publications concern projected or implemented electronic university

publishing systems, which is significant above all in Italy, where there is no consolidated tradition of University Presses comparable to that found in countries such as the UK and the US. This could encourage the distribution of electronic, scientific-academic publications freely available and/or at low cost, thereby reducing the problems connected both with the "journal crisis" and the "permission crisis" [41]. The recent establishment of the Florence University Press [42] constitutes a successful example of just such an initiative and, in addition, is closely connected to the university repository that, amongst other documents, distributes its output. Similar initiatives are underway in other universities, usually those that are more sensitive to the question of open access. Some of them are jointly developing projects aimed at the creation of national quality electronic periodical platforms with free access [43].

A final consideration concerns the service providers that gather and index metadata enable the retrieval of documents spread across different data providers. Currently in Italy, the already mentioned Pleiadi system is operative. There are other projects and prototypes that are usually run by research institutes (CNR with Cyclades [44] and Sail-eprint [45]; ISAS with the Torrii project [46]) that are developing systems and services that apply OAI-PMH and in the process give greater visibility to the various data providers.

4. Concluding remarks

Results of this analysis show that there is a certain connexion between the Italian participation in SIGLE and the setting-up of IRs, despite specific characteristics. Apart from the common technical-scientific content, both document systems share the same information producers. These are mainly universities and research institutes that from time to time have used different channels to distribute internally produced documents. This can be seen in the changes to the Italian SIGLE database, that above all concern the period when GL began to be distributed on the web and when open archives began to be developed. These changes are more evident in the case of research institutes that used to contribute a large number of documents to the Italian SIGLE. Universities are present thanks to the indexing of theses, which constitute a nucleus of documents for IRs too and which have often formed the starting point for the filling of new archives.

As in other countries, Italian IRs are in the growth phase. The number of documents in each repository is not conspicuous and the process of involving departments and academics is still only just underway. Nevertheless, considering the expansion of the Italian SIGLE database, it could be said that the IRs are starting from a favourable position, given that in less than three years they have already made more than 5000 documents available (excluding the large scale producers), with the majority of these being accessible in full texts. GL forms a significant part of the IRs and retrieval not only of bibliographic data, but also of full texts is one of the winning features of IRs. All the same, given the nature of IRs and the request for free access to scientific literature, it is essential that they give equal coverage to both conventional and non conventional literature.

What is the future for SIGLE in the light of the changes affecting the distribution of information on the web and the growth of IRs? We feel certain proposals can be made so as not to lose the documentary and organisational heritage that has been built up over the last twenty years. It is clear that, OAI compatible formats must be used to enable harvesting of data on GL. Furthermore, national reference centres could, for example, promote open access to GL produced by small institutes and occasional producers, and could exploit their knowledge networks of GL producers to stimulate the collection of GL following OAI standards.
We would be happy if these conclusions were to stimulate debate among information specialists leading to common and fruitful solutions for the freest and most openly accessible distribution of scientific information.

References and Notes

[1] Pagamonci, Armida (1987), "Il SIGLE e l'EAGLE: un percorso di politica comunitaria e di cooperazione europea in materia d'informazione scientifica", *Bollettino d'informazione AIB*, 1987; 27 (3/4): 348-363.
[2] Harnard Stevan (1999), "Free at Last: The future of Peer-reviewed Journals", D-Lib Magazine, 5 (12), 1999, http://www.dlib.org/dlib/december99/12harnard.html.
[3] Clifford Lynch A. (2003), "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age", ARL Bimonthly Report 226, Feb. 2003. http://www.arl.org/newsltr/226/ir.htm.
[4] Gelfand Julia (2004), "Knock, Knock: "Are Institutional Repositories a Home for Grey Literature?". 6th International Conference on Grey Literature Work on Grey in Progress, New York, 6-7 Dec. 2004. Conference Proceedings. Amsterdam: GreyNet/TextRelease, 2005, pp. 10-15.
[5] Workshop nazionale "Gli atenei italiani per l'Open Access: verso l'accesso aperto alla letteratura di ricerca, Messina, 4-5 Nov. 2004, available at: http://www.aepic.it/conf/program.php?cf=1.
[6] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Available at: http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.org.

[7] Pagamonci, Armida (1993), "Italian grey literature in SIGLE in the scientific and technical fileds". First International Conference on Grey Literature: Weinberg report 2000, Amsterdam 13-15 Dec. 1993. Conference Proceedings. Amsterdam: Transatlantic, 1994, pp. 97-106.
[8] SIGLE: System for Information on Grey Literature in Europe, Boston Mass, Silver Platter, 1980-.
[9] Alberani Vilma; De Castro Pietrangeli, Paola (1994), "The initiatives of a library association for the promotion of grey literature", 1.st International Conference on Grey Literature, Amsterdam, 13-15 Dec. 1993, Conference Proceedings. Amsterdam: Transatlantic, 1994, pp. 50-55.
[10] LG Informazioni. Newsletter of the AIB study groups of grey literature. N; 1 (1988) -N. 5 (1991).
[11] Di Cesare Rosa; Lazzari Gianni (2000), "Toward integration of information sources on grey literature: a case study. *Journal on Grey Literature,* 1 (4), 2000.
[12] Di Cesare Rosa (1996), Il centro Nazionale di Riferimento SIGLE nella gestione della letteratura grigia prodotta in Itlaia: ruolo e prospettive. 2° Convegno Nazionale sulla Letteratura Grigia, Roma, 20-21 Maggio 1996, Atti del Convegno a cura di V. Alberani, P. De Castro Pietrangeli, D. Minutoli, *ISTISAN Congressi*, 48, 1996, pp. 41-50.
[13] The database consultation system does not allow for access to the institutes' Authority file, therefore the number of institutions are probably more than that reported in the table.
[14] Luzi Daniela (1997), "The Internet as a new distribution channel of scientific Grey Literature: The case of Italian WWW servers", *Publishing Research Quarterly*, 13 (2), 1997.
[15] Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bilbiografiche, "Bibliografia Nazionale Italiana. Tesi di Dottorato", Roma, 1995-.
[16] arXiv.org e-Print archive: http://arxiv.org
[17] Luzi Daniela (1996), Modalità di diffusione e distribuzione di informazione grigia dei WWW server degli enti scientifici italiani, 2° Convegno nazionale sulla Letteratura Grigia, Roma, 20-21 Maggio 1996, Atti del Convegno a cura di V. Alberani, P. De Castro Pietrangeli, D. Minutoli, *ISTISAN Congressi*, 48, 1996, pp. 163-169.
[18] Hunter Philip, Day Michael (2005), "Institutional repositories, aggregator services and collection development", available at: http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/coll-development/coll-development.pdf.
[19] Lagoze Carl, Van de Sompel Herbert, "The Open Archives Initiative Protocol for Metadata Harvesting", available at http://www.openarchives.org/OAI/openarchivesprotocol.html.
[20] Lynch Clifford A. (2001), "Metadata Harvesting on the Open Archive Initiative", *ARL Bimonthly Report*, no. 217, pp.1-9, http://www.arl.org/newsltr/217/mhp.html.
[21] van Westrienen Gerard, Lynch A. Clifford (2005), "Academic Institutional Repositories. Deployment Status in 13 Nationas as of Mid 2005", *D-Lib Magazine*, 11 (9), http://www.dlib.org/dlib/september05/westrienen/09westrienen.html.
[22] Data submitted by each nation are available at: http://www.surf.nl/download/country-update2005.pdf .
[23] Pleiadi, (Portale per la Letteratura scientifica Elettronica Italiana su Archivi Aperti e Depositi Istituzionali) available at http://www.openarchives.it/pleiadi.
[24] Comba Valentina (2003) "Il progetto AEPIC (Academic E-publishing infrastructures - CILEA): Le esigenze degli autori e lo stato dell'arte dei serivizi" Bilbliotime, 6 (3), 2003, http://didattica.spbo.unibo.it/bibliotime/num-vi-3/comba.htm.
[25] Gargiulo Paola, Mornati Sussana, Contino Ugo, Tajoli Zeno (2005) Pleiadi, a portal solution for scholarly literature", Proceedings Open culture: accessing and sharing knowledge, Milan, available at: http://eprints.rclis.org/archive/0004403/Pleiadi_AICA2005_rev.pdf.
[26] Lynch Clifford A., (2003)"Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age", *ARL Bimonthly Report* 226, Feb 2003. Http://www.arl.org.newsltr/226/ir.html.
[27] Lynch Clifford A., Lippincott Joan K., (2005) "Institutional Repositories Deployment in the United States as of early 2005", *D-Lib Magazine*, 11 (9), http://www.dlib.org/dlib/september05/09lynch.html.
[28] The survey reports data on the last consultation on 2nd November 2005
[29] Ams Acta: http://amsacta.cib.unibo.it.
[30] E-LIS E-prints in Library and Information Science: http://eprints.rclis.org.
[31]. The archive is the result of the RCLIS (Research in Computing, Library and Information science) project, and is based on the model of the highly successful RePEc (Research Papers in Economics), an open archive for economics and related disciplines.
[32] De Robbio Antonella (2003), "E-LIS: un open archive in Library and Information Science", *Bibliotime,* 6 (1), 2003, http://www.spbo.unibo.it/bibliotime/num-vi-1/derobbio.htm.
[33] Cadmus IUE: http://cadmus.iue.it/dspace/index.jsp.
[34] SISSA Digital Library: http://digitallibrary.sissa.it/index.jsp.
[35] UNIPI. Sistema ETD: http://etd.adm.unipi.it.
[36] INGV, Earth-prints: http://www.earth-prints.org/index.jsp.
[37] The document type *reports* includes: reports, technical reports, research reports, departmental reports, working papers, project business plan. The document type *journal articles* includes: journal article (Online/unpaginated), journal article (Print/paginated), Newspaper/magazine article, article. The document type *proceedings* includes: conference proceedings, conference paper, conference presentation, conference poster, conference & workshop items. The document type *didactic materials* include: programmes, bibliographies, handouts, notes, exercises, syllabus, guide/manual.
[38] UNIPD Archivio gestione tesi: http://tesi.cab.unipd.it:8500.
[39] UNIROMA Padis: http://padis.uniroma.it
[40] Polaris: http://polaris.unitn.it.
[41] Suber Peter (2003), "Removing the barriers to research: An introduction to Open Access for librarians", http://eprints.rclis.org/archive/00001027/01/acrl.htm.
[42] UNIFI Archivio e-prints: http://eprints.unifi.it & Florence University Press: http://epress.unifi.it.
[43] District Architecture for Networked Edition (DAFNE), available at: http://dafne.cab.unipd.it.
[44]An Open Collaborative Virtual Archive Environment (CYCLADES), available at: http://www.ercim.org/cyclades.
[45] SAIL-eprints (Search, Alert, Impact and Link), available at: http://eprints.bo.cnr.it/main.html.
[46] Torii. Access the digital Research Community, available at: http://tips.sissa.it/docs/booklet.pdf.

Acknowledgement
We are grateful to the library of the Istituto Superiore di Sanità and the Central Library of CNR for making available the SIGLE databases.

Author Information

Daniela Luzi is researcher of the National Research Council at the Institute of research on populations and social politics. Her interest in Grey Literature started at the Italian national reference centre for SIGLE at the beginning of her career and continued carrying out research on GL databases, electronic information and open archives. She has always attended the International GL conferences and in 2000 she obtained an award for outstanding achievement in the field of grey literature by the Literati Club. Email: d.luzi@irpps.cnr.it

Rosa Di Cesare was born in Civita d'Antino (AQ) in 1952 and graduated from "La Sapienza" University in Rome in 1982. She received her diploma in Librarianship from the Vatican Library in 1996. She worked in the Central Library of National research council (CNR), where she started to become involved in research activity in the field of Grey literature (GL). Member of the Technical Committee for the SIGLE database from 1995 to 2001, she is presently responsible for the Library at the Institute of research on population and social policies (IRPPS) of the National research council. Her studies have focused on citation analysis and on the use of GL in scientific publications.

Roberta Ruggieri was born in Taranto in 1974 and graduated from "La Sapienza" University in Rome in 2002. She is presently a student at the same University in Information science course. She is having training at the Library of the Institute of Research on Population and Social Policies of the National Research Council.

Grey in the R&D Process

Keith G. Jeffery
CCLRC Rutherford Appleton Laboratory, United Kingdom

Anne Asserson
University of Bergen, UiB, Norway

Abstract
The rate of acquisition of data, its structuring into information and its interpretation as knowledge is increasing rapidly. There are more active researchers now than ever and the output of white publications per researcher is increasing. The output of grey publications is orders of magnitude greater. Past techniques of experts (librarians) cataloguing manually with metadata each publication do not scale. The problem is to find ways to manage this resource.

The hypothesis is in 4 parts: (a) that the R&D process itself provides some context for managing the information; (b) that linking the records of the process to the publications provides this context; (c) that questions of curation and provenance are addressed automatically in such an environment; (d) that such an environment integrates grey and white literature and other R&D outputs such as software, data, products and patents.

At UiB the emphasis of the work has been on assessment of the research output - especially publications - linked in context with records of the researchers, their organisational units, and related CRIS (Current Research Information System) information (the FRIDA system which is mostly CERIF-compatible).

At CCLRC the emphasis of the work has been on the production of an open access repository of publication outputs from the organisation (ePubs), linked to the CERIF-compatible CDR (Corporate Data Repository) CRIS and thus to other research outputs with associated metadata.

The recording of the data provides the context including the workflow of the R&D process, history and provenance. Grey documents produced as early ideas are captured in a temporal and organisational context, just as well as white publications, via the linked repository. CERIF allows, in a multidimensional framework, deduction or induction of relationships between documents – for example between a grey internal report and a white published paper - and with other research outputs. Furthermore, relationships between documents can be expressed explicitly: references and / or citations can be recorded. In this way a rich context for understanding the R&D output is provided, including versions, history and provenance.

Recording facts once in a structured R&D process environment and then re-using them in many ways reduces - by automated provision assistance - the need for user input of metadata to describe research outputs (especially grey literature) and thus addresses the scalability problem.

The costs including all staff, overheads, equipment, software etc for CCLRC are as follows:
(a) Development of the CDR (Corporate Data Repository CERIF-compatible): (i) Pilot phase: 30kEUR; (ii) Production phase (~ 1 year): 80kEUR; (iii) Annual maintenance including integration with the ePubs system and with our workflow environment: 30kEUR
(b) Development of the ePubs Open Access Institutional Repository: (i) 80kEUR (~ 1 year); (ii) Annual maintenance including integration with CDR 50kEUR
We are now (with integrating CDR, ePubs and workflow) developing re-engineered business processes: this programme of work is estimated at ~ 300kEUR per year for 2 years and relies on CDR and ePubs. We believe the benefits in improved effectiveness and efficiency will run at ~ 4mEUR per year.

The costs for the University of Bergen are as follows: The total cost of developing our national system Frida is 2.250.000 NOK for the 4 universities and the cost of UiB (only for developing) is approx 500.000 NOK.

1 BACKGROUND

1.1 The Problem

Research outputs – publications, patents, products – are produced by research activities or projects. The volume is increasing rapidly, partly due to the 'publish or perish' paradigm. Researchers are under increasing pressure to produce outputs, and to record them. The metadata describing the research outputs, and the source materials themselves, may not be adequately recorded. There are three major problems:

1. the information is difficult to collect, partly because the end-user interface to available systems presents a formidable threshold barrier;
2. the end-user commonly works in an environment that is not well-structured and so the tasks to record the research output information (usually the metadata including a URL to the source material) are not done, or done without sufficient attention, or simply forgotten;
3. commonly the information is available progressively; e.g. as a publication is submitted, accepted and published. At each stage metadata are available and should be recorded. Without a structured environment for the progressive collection of data the request for information input consists of a demand for a large amount of information - rather than building progressively the metadata corpus using small incremental data input steps - and this demand for a large amount of information is inhibiting to the end-user.

The aim of this paper is to address those problems.

1.2 The Research Process

The research process, considered in overview, is the set of steps – with alternatives and conditions, and some steps omitted or repeated – that produces research outputs such as publications, patents or products. The overall process can be represented (Figure 1):

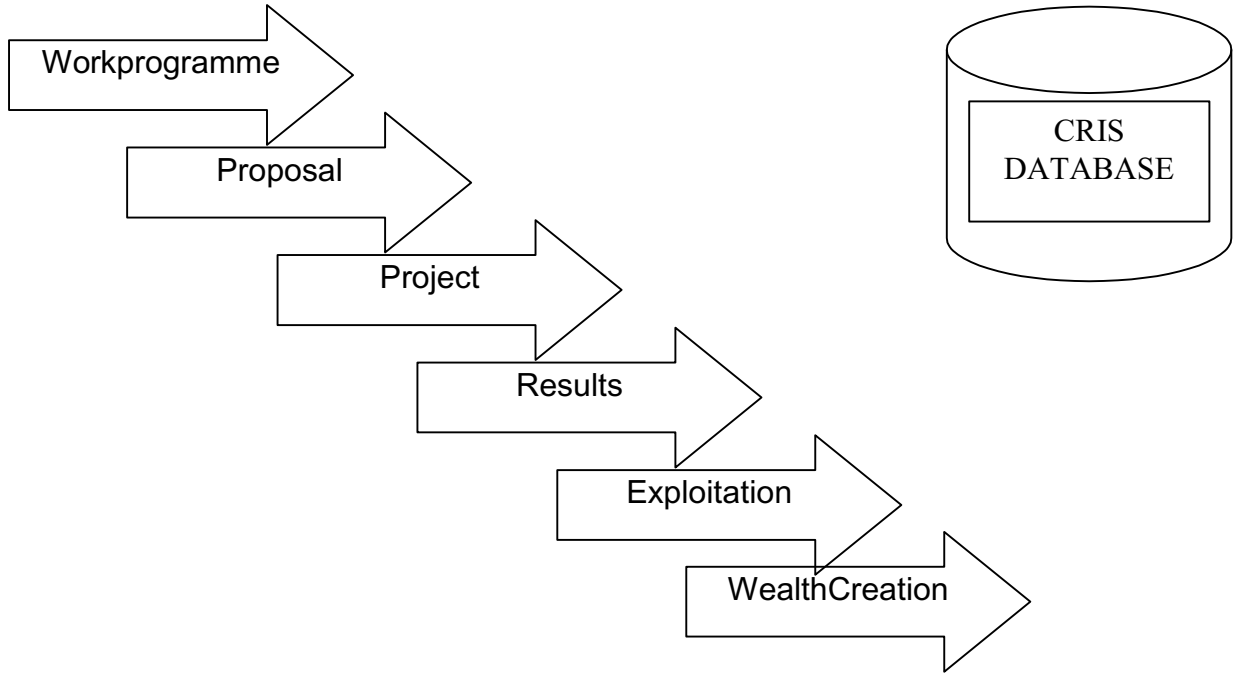


Figure 1: The Research Process

It should be noted that the metadata describing the research outputs (and indeed appropriate information collected at each step of the research process such as project title, project abstract, person(s) doing the research, organizational unit(s) of the persons etc.) are stored in a CRIS (Current Research Information System). The EC (European Commission) standard for recording research information recommended to member states is CERIF (CERIF). The advantages of relating output publications (and other outputs) to a CERIF-CRIS have been demonstrated (Jeffery and Asserson 2004) and concern mainly the value of contextual information in understanding the research publication and the maintenance of integrity and data quality. As an aside, the requirement for formalised metadata to describe research output publications thus allowing machine-understanding in addition to machine-reading was demonstrated (Jeffery 1999). Formalised publication metadata within or associated with the

formalised CERIF structure of a CRIS provides an information environment suitable for precise querying, improved constraint checking at data input and knowledge-based techniques to resolve heterogeneity and to provide end-user assistance. The knowledge-based techniques of deduction and induction can also be applied (Jeffery and Asserson 2004).

1.3 The Research Process and Grey Literature

Grey literature is an important research output. Commonly grey literature forms the documented 'know-how' of an organisation and may also be shared with partner organizations or persons. Grey literature commonly is stored in a repository, more particularly an institutional repository which may have open access if the organisation perceives business benefit (including public relations benefit) from making the material openly and freely available. The metadata describing the grey literature may be stored in the CRIS or in the repository or both; the source material (the grey literature publication) is stored only in the repository. Thus, the collection of metadata concerning the grey literature publication should be progressive in the research process and – being linked with the CERIF-CRIS – be associated with contextual data about the research.

1.4 The Hypothesis

The major hypothesis of this paper is that: (a) that the R&D process itself provides some context for managing the information; (b) that linking the records of the process to the publications provides this context; (c) that questions of curation and provenance are addressed automatically in such an environment; (d) that such an environment integrates grey and white literature and other R&D outputs such as software, data, products and patents.

2 RESEARCH PROCESS, PUBLICATIONS AND CONTEXT

2.1 The Process

The process was presented above in overview. It is clear that, knowing that a particular publication has been produced from a research project and submitted for publication, that the CERIF-CRIS provides additional contextual information adding value to the metadata concerning the publication – for example the purpose of the research, the persons in the team, the organizational units (universities, departments, groups) involved, the facilities used, the particular equipment utilized etc.

Furthermore, at the time of publication submission all this information should have been collected incrementally at previous steps in the research process so the burden on the researcher to input information at the time of submission is lessened considerably. The relevant information at any one time period (present or past) is easily determined because CERIF includes linking relations with role and date/time stamps linking the base entities (Jeffery 1999) so that a snapshot of database state at any time can be obtained. This is particularly helpful in tracking provenance and in data curation.

Here, each step of the research process is considered in turn with the (meta)data that is collected and the input information that is required for the step to be executed.

2.2 Research Process Steps

The major outputs and inputs at each step can be represented in the following table:

PROCESS STEP	SUB-PROCESSES	INPUT	OUTPUT
Workprogramme	Economic, Societal, Technology Foresight	World/Country state, world/country models, technology predictions, solicited advice	ProgrammeName Funding OrgUnit Person responsible Workprogramme document
Proposal	Idea, review previous work, Objectives, Method, Resources and Dependencies	Previous results, previous projects, finance, human resources	Title Abstract Person(s) OrgUnit(s) Proposal Document
Project	Project management	Previous results,	Title

		previous projects, finance, human resources	Abstract Person(s) OrgUnit(s) Funding Project Plan
Results	Initial result, internal review, peer review, registration or publication	Previous results	Person(s) OrgUnit(s) Project(s) Product(s) Product Description (e,g, publication reference metadata)
Exploitation	Results, business plan, finance, marketing, production, selling	Marketing information, economic information	Person(s) OrgUnit(s) Business plan Finance Data Marketing Data Production Data Sales Data
Wealth Creation	Marketing, employment, production	Marketing information, economic information	Person(s) OrgUnit(s) Annual Reports/Accounts Employment Records Dividends Records

From the above it is obvious that much (meta)data required in later steps can – and should - be recorded in earlier steps.

2.3 Research Output Publications

Research output consists of publications, patents and products. Here we consider only publications. The metadata required was defined (Jeffery 1999; Jeffery and Asserson 2004). However, in addition, for published (white) literature the full reference metadata is required. This includes author names as they appear in the publication (as well as the usual title, abstract), but also - for the channel of publication - the name, series, volume, part, ISSN or other identifier, pages within the bound publication containing the publication of the author(s) and other relevant information that can be (re)structured into Vancouver, APA or other formats.

It is clear that some of the metadata required is available at the time the publication is drafted: title, abstract, authors. More metadata is added through internal review: a URL to location in an institutional repository, linkages to the research context (using a CERIF-CRIS) and annotations by reviewers. Again, further metadata is added when it is submitted for external publication via peer review: publication channel name. Finally, when published, the full metadata can be completed – but most of the input is already done and only the publication channel details need to be added.

2.4 Grey Literature

The process for grey literature is little different. The early sub-process steps are the same, including usually internal review. However, the changes come in the later sub-process steps since there is no external peer-reviewed publication channel. However, it is common for an internally-reviewed grey literature publication to become (possibly with modifications) an externally peer-reviewed white publication. CERIF-CRIS provides the means to record this using the linking relation publication-publication with role and date/timestamps.

3 EXPERIENCE

The experiences at CCLRC and UIB are described.

At UiB, there has been an historical development, from FORSKDOK – an information retrieval system using semi-structured simple research data - through FDOK – a CERIF-based system developed at UiB - which created the national demand for FRIDA (FRIDA) – a system developed at University of Oslo but

used by 4 universities and with a steering committee and project group representing them. FRIDA is essentially CERIF-compatible but concentrates on entities persons, organisational units and publications and the entity project is being implemented. The representation of publications required extensions over the CERIF2000 model but these extensions are now being incorporated into CERIF2005.

The use of FRIDA allows each university to quantify its white research output in a form suitable for the funding ministry and thus assures appropriate funding allocations. However, in addition each university now has a consistent record of its intellectual property including grey literature, the importance of which is being realized progressively. FRIDA acts as a hub; from each university personnel system data are imported into it. To ensure completeness and to overcome the threshold barrier of metadata input, records are imported from the Thomson ISI system and the national library bibliographic system, BIBSYS (BIBSYS). FRIDA is providing a component to link to the institutional OA repositories of the four universities providing access to full-text or hypermedia representations of the research output.

At CCLRC, the CRIS (named the CDR: Corporate Data Repository) has several functions:

- it links together many other systems including those managing finance, human resources, projects and e-documents among many;
- it links to the open access institutional repository containing metadata (and in many cases also the source material) for CCLRC publications – both grey and white literature;
- it links to portals managing access to research datasets (with their own metadata) and is being extended to link to open source software for research;
- it provides the repository record and source data to drive the workflow-supported business processes of the organization including research processes such as research proposal submission and publication submission and management or administrative processes such as travel claim management or authorising leave.

Its central role is evident; it provides the hub for interoperation and linkage making replacement of other systems and system components relatively easy and allowing development and evolution in line with the CCLRC business requirements. It allows recording of the information related to the research process. It also provides the central mechanism for curation of CCLRC information.

4 THE WAY FORWARD

The GRIDs environment (Jeffery 2001, 2004, 2004a) provides an infrastructure to benefit greatly the research process. The major concept – distinguishing GRIDs from the GRID (Foster & Kesselman 1999) is that the user interacts intelligently with the GRIDs environment to determine the request and the GRIDs environment then proposes a 'deal'; composing the necessary resources of computation, networking, data stores, information provision, detector-based additional data collection, appropriate software for analysis and visualization and ambient intelligence as necessary for the results to reach the end-user on her device of choice. The user accepts or rejects the proposed 'deal; if accepting the GRIDs environment then executes the request. This level of virtualization of all the resources required frees the researcher from tedious work finding the data, interpreting it as information, running computer models, executing statistical analyses, generation visualizations, collecting more data etc etc – in order to leave the researcher free to think and do research. Of course these same virtualization virtues have equal benefits in a commercial business or industrial environment, in healthcare, in environmental management and in the humanities.

5 REFERENCES

(ArXiv) www.arxiv.org
(Asserson Jeffery et al 2002) Asserson,A; Jeffery,K.G; Lopatenko,A: 'CERIF: Past, Present and Future' in Adamczak,W & Nase,A (Eds): Proceedings CRIS2002 6th International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 33-40 2002 (available under www.eurocris.org)

(Asserson and Jeffery 2004) Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 29-40 (available under www.eurocris.org)

(Berners-Lee 1999) Berners-Lee,T; 'Weaving the Web' 256 pp Harper, San Francisco September 1999 ISBN 0062515861

(BIBSYS) www.bibsys.no

(CERIF) www.eurocris.org/cerif

(DC) http://www.oclc.org:5046/research/dublin_core/
(Foster and Kesselman 1998) Foster I and Kesselman C (Eds). The Grid: Blueprint for a New Computing Infrastructure. Morgan-Kauffman 1998

(FRIDA) http://frida.uio.no

(Jeffery 1999) Jeffery, K G: 'An Architecture for Grey Literature in a R&D Context' Proceedings GL'99 (Grey Literature) Conference Washington DC October 1999 http://www.konbib.nl/greynet/frame4.htm

(Jeffery, 2000). Jeffery, K G: 'Metadata': in Brinkkemper,J; Lindencrona,E; Solvberg,A (Eds): 'Information Systems Engineering' Springer Verlag, London 2000. ISBN 1-85233-317-0.

(Jeffery 2001) Jeffery,K.G.;'GRIDs: Next Generation Technologies for the Internet' Invited Keynote Presentation (Abstract in Proceedings) Eds Wang,Y; Patel,S; Johnston,R.H; OOIS2001 Conference, Calgary August 2001, page 1, Springer, ISBN 1-85233-546-7

(Jeffery 2004) Jeffery, K.G.; 'GRIDs, Databases and Information Systems Engineering Research' in Bertino,E; Christodoulakis,S; Plexousakis,D; Christophies,V; Koubarakis,M; Bohm,K; Ferrari,E (Eds) Advances in Database Technology - EDBT 2004 Springer LNCS2992 pp3-16 ISBN 3-540-21200-0 March 2004

(Jeffery 2004a) Jeffery, K.G.; 'The New Technologies: can CRISs Benefit' in A Nase, G van Grootel (Eds) Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 77-88 (available under www.eurocris.org)

(Jeffery and Asserson 2004) K G Jeffery, A G S Asserson; Relating Intellectual Property Products to the Corporate Context; Proceedings Grey Literature 6 Conference, New York, December 2004; TextRelease; ISBN 90-77484-03-5

(OAI) www.openarchives.org
(W3C) www.w3.org

Author Information

Keith Jeffery is currently Director, IT of CCLRC (Council for the Central Laboratory of the Research Councils), based at Rutherford Appleton Laboratory in UK and head of Business and Information Technology Department with a staff of 140 supporting over 360000 users, developing software for business and science and doing leading edge R&D. The department hosts the UK and Ireland Office of W3C and develops and supports the largest OA (Open Access) institutional repository in UK. Keith is a Fellow of both the Geological Society of London and the British Computer Society. He is a Chartered Engineer. He is an Honorary Fellow of the Irish Computer Society. He is president of euroCRIS (www.eurocris.org) and vice-president of ERCIM (www.ercim.org) and holds three honorary professorships. He has extensive publications and has served on numerous programme committees and research grant review panels. He has particular interests in 'the research process' and the relationship of hypotheses, experiments, primary data and publications based on research in information systems, knowledge-based systems and metadata. Email: k.g.jeffery@rl.ac.uk

Anne Asserson holds a Cand. polit. with a Masters in Information Science from the University of Bergen, UiB. She has been working with Research Documentation, and has participated in substantial parts of CRIS developmental work, locally and nationally. Anne Asserson has been part of the establishing and implementing of a Research Documentation system, Fdok http://www.ub.uib.no/fdok/sok/, at the UiB. For several years she was the chairwoman of the Steering Group of the national CRIS system and project secretary of a National system for academic administration. Anne Asserson is presently representing UiB in the national group that is implementing a new national research documentation system, FRIDA. She has also participated in The CORDIS funded European-wide project on " Best Practice" 1996 . She was a member of the working group set up 1997 that produced the report CERIF2000 Guidelines (1999) www.cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS is now the custodian of the CERIF model www.eurocris.org. Anne Asserson is a member of the Best Practice Task Group.

Assisting scientists to make their research results world wide freely available: An experience begun in the 90's

Stefania Biagioni
Institute of Information Science and Technologies, ISTI
The Italian National Research Council, CNR, Italy

Abstract
The ERCIM Technical Reference Digital Library service (ETRDL) was promoted by the European Research Consortium for Informatics and Mathematics (ERCIM) with the aim of managing grey literature produced by scientists working in the areas of information science and applied mathematics, making their research results immediately world wide available and also building a test-bed for their research activities. In this paper ETRDL is described and its future development presented. Section 1 traces the history of ETRDL from its conceptualization to the first experimental service given by the early ETRDL prototype. This was implemented to satisfy the requirements of the European IT scientific community, although realized as a part of NCSTRL (the US Networked Computer Science Technical Reference Library). Section 2 describes how further developments made ETRDL a system presently capable to manage DLs for very different types of literature. Section 3 discusses issues related to migration of ETRDL managed repositories to OpenDLib, an advanced digital library service system developed by ISTI-CNR. Advantages and difficulties of this migration are considered in the conclusions.

1. ETRDL history: an experience begun in 90s

In the 1996 the European Research Consortium for Informatics and Mathematics (ERCIM) [1] recognized the importance of the digital library technologies and thus decided to establish an ERCIM programme for R&D in Digital Library (DL) sector. The Italian National Council of Research (CNR) [2] was appointed as coordinator of the ERCIM Digital Library Initiative (DLI)[3].
The aim of DLI was to promote the development of DL technologies in Europe. Such Initiative had three objectives: to support long-term research activities, develop large digital collections, collaborate with the US Digital Library Community. Since 1996, a series of research-oriented activities, mainly sponsored by the DELOS Working Group [4], were thus organized, e.g. workshops, conferences, collaborative studies on DL-related research issues. Within this context, towards the end of 1997, ERCIM decided to undertake an implementation activity by setting its own digital library up to provide open access to grey literature: the ERCIM Technical Reference Digital Library (ETRDL) [5][10].
During the preliminary meeting held in Budapest (1996) and in Pisa (1997), ERCIM Librarians and computer scientists decided to build a DL infrastructure that should satisfy particular needs of different European Institutions and contemporarily participate to the international research context. The first step towards the development of ETRDL was a survey of above requirements. Two main objectives were identified: (i) to implement functionality to satisfy specific local requirements regarding the language of the user interfaces; (ii) to assist ERCIM scientists to make their research results world wide immediately available by providing them with on-line facilities for a self-publishing service. The Project, funded by ERCIM and the Delos WG on DL - Esprit Long Term Research Programme - LTR n. 21057 5, was a collaborative effort among the following National Institutions: CNR-Italy, CWI-The Netherlands, CRCIM-Czech Republic, FORTH-Greece, GMD-Germany, INRIA-France, SICS-Sweden, SZTAKI-Hungary.
The system was developed by ISTI-CNR (Pisa) in the context of collaboration with Cornell University DL Group. This group had already realized an on-line grey literature distributed service in the Computer Science domain, called NCSTRL (the US Networked Computer Science Technical Reference Library) [6][9].
The system employed by NCSTRL was Dienst [7][8], an open conceptual architecture for digital libraries, an open protocol for communication in the architecture, and a reference software system implementing the architecture, an open system that provides internet access to a distributed, decentralized, multi-format documents collection. ERCIM DL should have become a node of the NCSTRL federation and adopt the infrastructure Dienst: in this way ETRDL would form part of an international collection of grey literature.

1.1. The Dienst Architecture and the ETRDL extended functionality

The Dienst architecture was built on the notion of individually defined services that when combined together create a distributed digital library. Services and resources in a Dienst digital library may be located anywhere on the Internet. The functionality of a Dienst digital library includes storage and access to resources, deposit of new resources, discovery and browsing of those resources.
Communication with and among individual Dienst services takes place via an open protocol, which makes it possible to build other service layers on top of the existing basic Dienst services.
The services defined in the protocol are as follows:

- A *Repository Service* that provides the mechanisms for storage of and access to the digital documents;
- An *Index Service* that accepts queries and returns lists of documents identifiers matching those queries.
- A *Meta Service* that provides a directory of locations of all other services.

Human interaction with these services and their protocols is mediated by a *User interface service*. Server implements a service.
A Dienst Standard Site (DSS) instances the functionality of the Repository Service, the User Interface Service and the Index Service for its own digital documents. The Meta Service functionality is instantiated by distributed servers: the Master Meta Server (MMS) and the Regional Meta Servers (RMS). In order to improve the overall connectivity by reducing the number of interconnections between servers DSS are partitioned into *regions*.

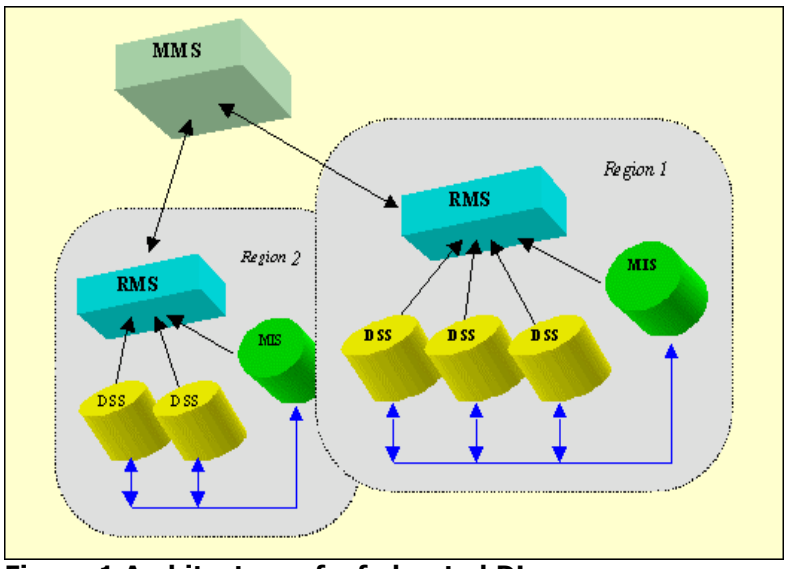


Figure 1 Architecture of a federated DL

A region comprises DSSs that have a good connectivity among them, an RMS and Merged Index Server (MIS) that act as back-up server.
A set of intercommunicating DSS, DLS, RMS MIS and a MMS constitute a *federated digital library* (see Figure 1).
The Dienst reference implementation released when the development of ETRDL begun was Dienst version 4.1.9. This version provides primarily a search service offering the user the possibility to perform a simple monolingual free-text search over the entire collection or to enter query terms in three fields: author, title and abstract. This primitive service did not satisfy completely the requirements of the ERCIM users. The main aspects lacking regarded: (i) subject classification; (ii) on-line document submission and deletion; (iii) multiple language indexing and search; (iv) on line documents administration.
In order to overcome these lacks, we implemented ETRDL by extending the Dienst reference implementation [11].
From the architectural point of view ETRDL consists of a federation of extended DSS belonging to the same region, adopting the same naming schema (ercim.xxx), and enabling the same core set of functionality. Each ERCIM institution (publishing authority) that participated in the project instanced a modified DSS as a local server, managing one or more local collections, and also providing locally developed functionality. At the same time, however, each publishing authority member of ETRDL was also a part of the NCSTRL federated digital library.
The extended DSS instances the functionality of an Extended Repository Service, an Extended User Interface Service and an Extended Index Service for its own digital documents. It also instances the new functionality of a Submission/Withdraw Service and Administration Service. The extended and the new services are accessible via a Dienst protocol - by the way of a set of new services requests - that defines the public interface to the services. Great care was taken to maintain compatibility between NCSTRL and ETRDL. This means that an NCSTRL user can access the documents of the ETRDL collection by selecting the ETRDL local institutions and using NCSTRL-implemented functionality; similarly all the ETRDL collections can be queried simultaneously through the ETRDL services, or separately through the respective local services.

The functionality provided by the extended and new services are the followings [12][13]:

Multilingual common information access
- User interface localization (local language as well as English)
- Different character set manipulation (complete Latin-1 char. set -ISO_8859-1)
- Cross language search
- Metadata extensions (from RFC1807 vs. Dublin Core metadata set)

Search and Browse
- Browsing extensions by subject (ACM, MSC, Free keywords)
- Search extensions by subject, date, language, type
- More complex query structures

Publishing
- Submission, withdrawal
- Administration

1.2. A first experimental service

Many ERCIM scientists and technicians were involved in the setting up of the first experimental technical reference service. The operational DL service was released in 1998 after a one-year period of testing and refining by ERCIM Librarians and developers.
Each project partners set up its own pilot server with a local language user interface besides an English one, as is showed in Figure 2.

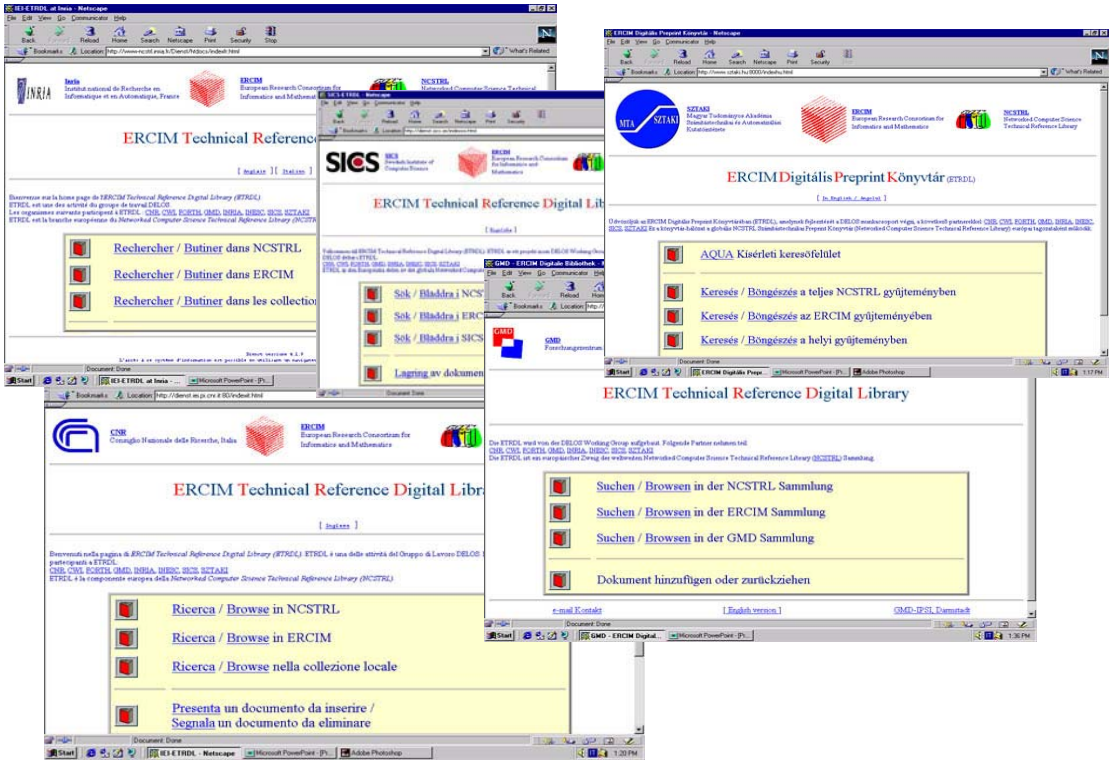


Figure 2 Samples of ETRDL User Interface

The technical reference service was working for two years till the end of the project in 1999. Its collection consisted in all kinds of grey literature (technical reports, pre-prints, proceedings of conferences or workshops available free, theses, project deliverables, etc.).

2. ETRDL to day

The ERCIM Project finished in 1999, but the CNR decided to maintain the service working for the Italian partners. This implied the managing of the entire Dienst system infrastructure and the ETRDL digital library by the ISTI-CNR. However, ISTI-CNR proceeded in developing its software, so that ETRDL can today be considered a DL management system capable of different applications. Therefore, from this point on the name "ETRDL" will be used both to indicate the software for specific applications and the digital library of CNR.
At the present, the ETRDL system delivers a self-publishing service for some Italian CNR institutions. Its metadata are harvested to exchange information with other initiative as the centralized data-base of the

CNR Grey Literature. Further, the CNR User Interface allows accessing to the new NCSTRL[1] [22] and to all Institutions currently working.
A special instantiation of ETRDL, named PUMA DL (Publications Management DL), was set up to allow the self-archiving of the published works by ISTI researchers. PUMA manages the access rights of the documents archived. Both ETRDL and PUMA DL contribute in managing of a number of institutional repositories. Their metatada are harvested for delivering a set of bibliographic services.
Different instances of the ETRDL system, regarding different fields of science, with personalized extensions were developed by ISTI-CNR within collaboration with Vallisneri Library of the University of Padova, the Costal Marine Environment Institute of CNR in Taranto. The aims was to make materials, that often remains unpublished and thus hardly retrievable, accessible by the scientific community engaged in the environmental study and conservation:
As a result of such a collaboration, two Digital Libraries [15] are today working in order to collect a wide amount of documents such as reports, technical reviews, graduate and undergraduate thesis, maps etc.

- The first, a digital thematic library called *LVDL Laguna di Venezia Digital Library*[2] [15], collecting published and unpublished documents arising from scientific studies about the Lagoon of Venice;
- The second, created following the example of LVDL and the prototype pattern, called *MeCME Mediterranean Costal Marine Environment Digital Library*[3] [16],collecting multidisciplinary documents concerning scientific studies in the Mediterranean Costal areas and in particular the Southern one.

Both DLs cover up similar fields of interest and classify the documents according to the same semantic descriptors (ASFA Thesaurus, Library of Congress Classification).
Further steps in the collaboration will be: 1) to create a common web site and a common interface to access digital libraries concerning the area of marine biology in order to permit to different users an easy, integrated search and retrieval of the documents; 2) to make the DLs managed by the more advanced system software OpenDLib (described in the next paragraph); 3) to promote new collaborations with public and private partners in step with increase in number of collections.

3. ETRDL evolution: migration to the OpenDLib service system

In this paper, we have described the implementation of ETRDL developed as part of NCSTRL network, employing and adapting the Dienst infrastructure. Moreover we have described how we have extended and specialized this service in order to support different types of literature and how ETRDL is currently working in the context of many scientific Libraries.
However, as so often happens in the computer science world, the ETRDL system needs a more sophisticated user-oriented service in order to be compliant with current requirements. Such requirements are fulfilled by a new software system, OpenDlib [17][18][19], developed at ISTI-CNR. Thus ETRDL documents will migrate into this system.
OpenDLib consists of an architectural infrastructure and a federation of services customizable to meet the requirements of different communities. It allows either loading or harvesting the content to be managed and supports an innovative document model to represent multi-edition, structured, multimedia documents that can be disseminated in multiple manifestation formats.
OpenDLib provides a number of interoperating services that implement the basic functionality of a digital library, such as acquisition, description, storage, search, browse, selection and dissemination of documents, and other advanced services such as collection services. These functions can handle a wide variety of document types with different format, media and structure, as well as new types of documents that have no physical counterpart, such as composite documents consisting of slides, video- and audio recording, etc. The OpenDLib powerful document model, i.e., the DoMDL model, permits this. Further, OpenDLib implements user management services, such as registration, authorization and authentication, and personalization services, such as construction and dynamic modification of personalized virtual views of the information space.
The OpenDLib architecture has been designed to be highly interoperable with other libraries. In particular, an OpenDLib library can act both as an OAI-PMH [23] data and service provider. This implies that the metadata maintained by any OpenDLib digital library can be open to other libraries and, vice-versa, the OpenDLib services can access the metadata published by any other OAI-PHM compliant library.
The first step in migrating ETRDL repositories to OpenDLib was defining the metadata mapping from RFC 1807 format [20] to the Qualified Dublin Core [21] in order to reuse metadata and documents

[1] NCSTRL has been re-implemented by the Old Dominion University Digital Library Research Group jointly with Virginia Tech and the University of Virginia, to an OAI-based framework. This moves NCSTRL from its previous Dienst-based distributed searching paradigm to one of the metadata harvesting. It has currently provided as a Public Service by ODU Digital Library Research Group and made still accessible by the user interfaces of ETRDL.
[2] Partners: Vallisnery Library, Palazzo Grassi and Hydrobiological Station in Chioggia, Geology, Paleontology and Geophysics Library (University of Padova), Naturalistic Observatory of the Lagoon and the Natural History Museum of Venice, Department of Environmental Sciences of the Ca' Foscari University of Venice, Italy.
http://laguna.isti.cnr.it/
[3] Library of the Institute Talassografico "A. Cerruti"- Costal Marine Environment Institute of CNR, Taranto Italy.**http://mecme.isti.cnr.it**

associated. The new DL was implemented using the standard base configuration of OpenDLib, however, we have to study the more appropriate configuration of the system from architectural and semantic point of view in order to meet the requirements of the ETRDL digital libraries that are currently working.

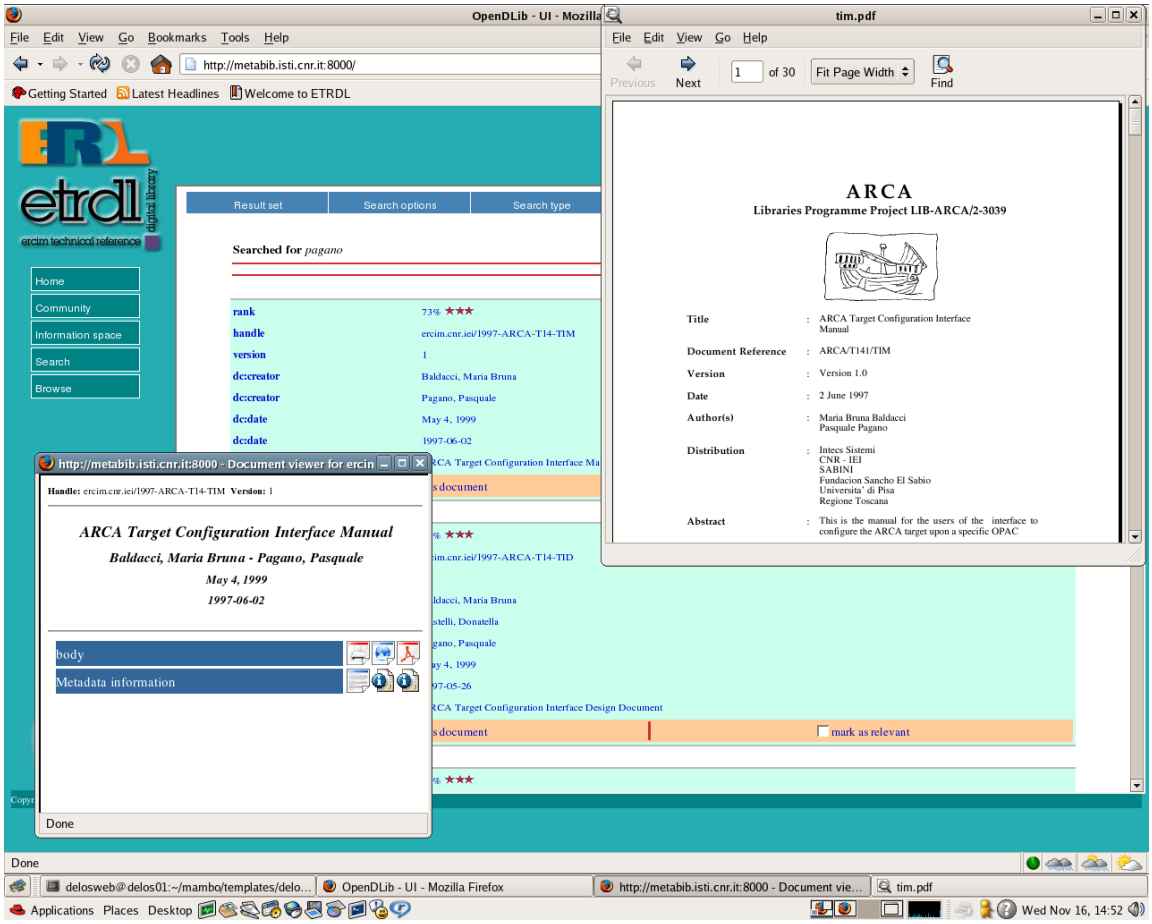


Figure 3 Sample of the new ETRDL search user interface

The new organization of ETRDL is now in an experimental phase (see a sample in Figure 3). The following decisions are still under consideration:
1) Which collections and which services have to implemented, 2) which additional metadata formats will be supported, 3) which document types with different format, media, structure and manifestations we are going to use; 4) which search options we need: google-like or fielded (with fields selected from a variety of known metadata formats); single or cross-language; with or without relevance feedback.

Conclusions

ETRDL is the result of a long lasting experience during which ETRDL staff, both researchers and librarians, became acquainted with all problems inherent in the conception, realization and development of very new library management systems. This process can be considered under various positive aspects. First of all, European research people and technical staff, as well as librarians, began working in a very new field while exploiting advanced systems realized in the U.S. This event has given rise to the most advanced DL community in Europe, i.e., the Delos community. Another important result is that the ETRDL prototype has been continuously operational since its first installation, although it has been enriched with new functions during its life. This has offered, and is still offering meaningful advantages, i.e., all the documents published by the different authorities during the entire life time of ETRDL are still available; many authors and librarians have become acquainted with self-publishing and self-archiving functions, while research people can presently use the content of ETRDL repositories to experiment new DL prototypes.
In the future, the planned migration to OpenDLib will assure the OAI compliancy, and a set of new advanced services. These services will preserve metadata and digital objects against changes in technologies that should make them inaccessible or meaningless.
The development of the ETRDL system software required a lot of work but its overall functioning presented no difficult problem. Most difficulties have been encountered, instead, in the administrative and social sphere. When the original project by ERCIM was concluded, the CNR staff began an activity, parallel to the activity for research and development, to diffuse information about digital libraries and related technologies through courses, seminars and demonstrations. Although most librarians proved to

be very much interested, academic people opposed many objections against self-publishing and institutional open archives. This is the reason why most of the authorities presently managed by ETRDL are research institutions. As is known, the interest in the OA themes started in the academic field very few years ago. We hope that the consensus recently obtained by the Pleiade national project, and the interest in the DL technologies by industrial and commercial organizations will make the future of open archives easier.

References

[1] European Research Consortium for Informatics and Mathematics http://www.ercim.org
[2] Italian National Research Council http://www.cnr.it
[3] Thanos, C.: *The Ercim Digital Library Initiative.* Ercim News, n. 27, October 1996.
[4] DELOS Network of Excellence http://delos-noe.iei.pi.cnr.it
[5] Ercim Technical Reference Digital Library http://dienst.isti.cnr.it (http://etrdl.isti.cnr.it)
[6] Networked Computer Science Technical Reference Library http://www.ncstrl.org
[7] Lagoze, C., Shaw, E., Davis, J.R., Kraft, D.B.: *Dienst: Implementation Reference Manual*. Cornell Computer Science Technical Report TR95-1514.
(http://cstr/cs/cornell.edu:80/Dienst/UI/2.0/Describe/ncstrl.cornell/TR95-1514)
[8] Lagoze, C., Davis, J.R.: *Dienst: an Architecture for Distributed Document Libraries*. In: Communications of the ACM, 38 (4) (April 1995), 45.
[9] Lagoze, C., Fielding, D.: *Defining Collections in Distributed Digital Libraries*. In: D-Lib Magazine, (November 1998).
[10] Biagioni S., Borbinha J., Ferber R., Hansen P., Kovacs L., Roos F., Vercoustre A.M. - *The ERCIM technical reference digital library*. Research and Advanced Technology for Digital Library: Second European Conference, ECDL 98. Proceedings (Heraklion, Crete, Greece, 1998), 905-906. Lecture Notes in Computer Science 1513. C. Nikolaou, C. Stephanidis (Eds.). Springer Verlag, 1998.
[11] Andreoni A., Baldacci M. B., Biagioni S., Carlesi C., Castelli D., Pagano P., Peters C., Pisani S. - *Developing a European technical reference digital library*. Research and Advanced Technology for Digital Libraries. 3rd European Conference:ECDL'99. Proceedings (Paris, September 22-24 1999), 343-362. Lecture Notes in Computer Science 1696. S. Abiteboul and A.M. Vercoustre (eds.). Springer Verlag, 1999.
[12] Andreoni A., Baldacci M. B., Biagioni S., Carlesi C., Castelli D., Pagano P., Peters C., Pisani S. - *The ERCIM technical reference digital library: meeting the requirements of a European community within an international federation*. In: D-Lib Magazine, 5, n. 12, December 1999.
[13] Baldacci, M.B., Biagioni, S., Carlesi, C., Castelli, D., Peters, C.: *Implementing the Common User Interface for a Digital Library: The ETRDL experience.* In: Proceedings of Eighth DELOS Workshop on User Interfaces in Digital Libraries. (Kista, Sweden, 1998). DELOS Working Group Report No.99/W001, (1998) 63-72.
[14] Biagioni S., Busatto E., Carlesi C., Filippi M., Ortolan M., Zane A. - *LV and MeCME digital library: two open doors to the aquatic sciences documentation*. EURASLIC 11: Open Waters - Open Sources, 11th Biennal Conf. of the Europen Association of the Aquatic Science Libraries and Information Centres, (Split, Croatia, May 4-6, 2005).
[15] Biagioni S., Busatto E., Carlesi C., Locatello L., Ortolan M., Zane A. - *Providing access to the inaccessible: a new thematic digital library for the Venice lagoon*. 10th EUNIS 2004 IT Innovation in a Changing World (Bled, Ljubljana, Slovenia, June 29 - July 2, 2004). Proceedings, pp. 106-109. Viljan Manic and Bostjan Vilfan (eds.). University of Ljubljana, Faculty of Computer and Information Science, 2004.
[16] Biagioni S., Carlesi C., and Filippi M. - *MeCME Digital Library (Mediterranean Costal Marine Environment Digital Library).* CNR- ISTI 2004-PR-11.
[17] OpenDLib http://www.opendlib.com
[18] Castelli D., Pagano P. - *OpenDLib: a digital library service system*. Research and Advanced Technology for Digital Libraries. 6th European Conference, ECDL 2002 (Rome, Italy, September 2002). Proceedings. Lecture Notes in Computer Science 2458, p. 292-307. Maristella Agosti, Costantino Thanos (eds.). Springer Verlag, 2002. (1999).
[19] Castelli D., Pagano, P. *A System for Building Expandable Digital Libraries*, Joint ACM-IEEE Conference on Digital Libraries (Houston, Texas, May 2003). Proceedings, p. 335-345. C. C. Marshall, Geneva Henry and Lois Delcambre (eds.). IEEE (2003).
[20] Lasher, R., Cohen. D., *RFC 1807 - A Format for Bibliographic Records*. Stanford, June 1995. http://www.faqs.org/rfcs/rfc1807.html
[21] Dublin Core Metadata Initiative: Dublin Core Metadata element set, version 1.1: Reference description. (http://dublincore.org/documents/dces/)
[22] An OAIbased NCSTRL. http://dlib.cs.odu.edu/index.html
[23] Open Archive Initiative. http://www.openarchives.org/

MetaGrey Europe, A Proposal in the Aftermath of EAGLE-SIGLE

Joachim Schöpfel
Institut de l'Information Scientifique et Technique, INIST
Centre National de Recherche Scientifique, CNRS France

Abstract
In the aftermath of EAGLE, the objective of the MetaGrey Europe project is to organize a new network of European leading organisations for the collection and dissemination of grey research literature, to facilitate the access to European grey literature through a unique web-interface (metasearch engine), to increase its international visibility and to safeguard the former SIGLE records and integrate them in the new network in an OAI-PMH compliant format. The communication describes the project scope, assesses the resources, contains a risk analysis and an outline of costs and benefits and provides information about project management, communication strategy and planning.

Introduction

2005 is a crucial year for the European grey literature. The EAGLE concept doesn't work any longer. The input to the SIGLE database definitely stopped in April, and the EAGLE association is to be liquidated.

EAGLE was a co-operative network for identification, location and supply of grey research literature. A non-profit association formed by the national centres participating in SIGLE, it produces and provides access to the bibliographic database SIGLE. All documents referred to can be obtained through the national SIGLE centres.

The decline of the EAGLE concept has four reasons:

(1) Internet: The 1985 concept of the SIGLE database has not been able to keep up with the rapid development with the Internet and the online resources. Even if FIZ Karlsruhe developed a web-version of SIGLE, the database production didn't adjust to the possibilities offered by the new technologies of information:

- No direct online cataloguing in a shared database,
- no metadata harvesting,
- very few records of e-documents,
- no direct access to full text,
- no links to other resources.

(2) Organisational structure: The highly centralized structure (national input centres, a central operating agent) was unable to open the network easily to new and/or important producers of grey literature on the national or international level. Furthermore, the formalized functional structure was not adapted to the need of flexible and rapid decision taking and marketing.

(3) Coverage: For most of the EAGLE countries, the national input became less and less representative. Often, even for the national input centres input criteria and workload excluded a significant part of their grey collections (dissertations, reports, conference proceedings).

(4) Economic model: The EAGLE model was based on flat membership fees and revenues from servers. Access to the database remained "traditional" with an annual subscription to the host or else a pay-per-view for each record with a credit card. With important members retiring from input and membership and the success of open access based initiatives, the economic model came to an end and needed revision.

The EAGLE members started a strategic debate on the future of the concept in 2001. The result is the present project: instead of a further development of the 20-year-old concept, the German and French EAGLE members propose a radical change that integrates the latest web-based technologies and the economic environment of open access to scientific and technical literature.

The following communication is an updated and augmented version of the project brief discussed at the 2005 General Assembly of EAGLE in Karlsruhe. Acknowledgments to Silke Rehme (FIZ Karlsruhe) and Dr. Irina Sens (TIB Hannover) for their helpful comments.

Projects aims and objectives

The main objective of the MetaGrey Europe project is to translate the key ideas of EAGLE in terms of 21st century information technology:

1. To organize a network of European leading organisations for the collection and dissemination of grey research literature.
2. To facilitate searchability of European grey literature through a unique web-interface.
3. To increase its international visibility.
4. To enhance online access to the full-text of European research grey literature whenever possible.
5. A complementary objective is to safeguard the SIGLE records (file data: 833,000 records November 2003) and integrate them in the new network.

The target group of MetaGrey Europe are scientists, academics, students, and library and information professionals.

Project scope

The MetaGrey Europe project's objectives are declined into three rationales:

1. *Development of a metasearch engine for grey literature ("MetaGrey").*

The MetaGrey search engine is hosted on a central server and maintained by a designed member of the network. It provides a simple and intuitive search interface adapted to the bibliographic and metadata of the main types of grey literature. The metasearch engine technology developed since about 10 years has arrived at a stage of maturity that allows specific developments as for this project.

2. *Transformation of the SIGLE database in an open archive ("OpenSIGLE") hosted on an open archive server and OAI-PMH compliant.*

The OpenSIGLE database is hosted on an OAI server and maintained by a designed member of the new network.

3. *Connexion of the most representative and important collections of European grey literature and of the "OpenSIGLE" archive to this metasearch engine.*

Grey literature key collections (databases, catalogues, archives) and other search engines are connected to the MetaGrey search engine by their institutions, with assistance by one or more members of the new network. There is no need for secondary or complementary cataloguing as for the SIGLE records.

The quality of the search engine and the overall projects depends also on the careful selection of grey collections and reliable partner organisations that guarantee the quality of data and documents. Conditions of access to and dissemination of grey literature are maintained by and remain under responsibility of the participating institutions. The overall structure of MetaGrey is decentralized.

The project rationale can be compared to the GrayLit portal of US Federal Agencies' grey literature. Other projects in the MetaGrey environment are the gateway "Science.gov" to authoritative selected science information provided by U.S. Government agencies, including research and development results, the new Google portal "Scholar" (access to scholarly literature from academic publishers, learned societies, universities and institutional repositories), the technology and concepts developed by the Hannover Search Engine Lab (with the leading German metasearch engine "MetaGer.de" and the German ResearchPortal "forschungsportal.net") or the metasearch engine "Virtueller Katalog Kunstgeschichte" funded by the DFG and developed with the MPG (History of Art, 14 organisations), different open access initiatives (the LARA project of INIST-CNRS; the former MAGiC project of the British Library and Cranfield University; the CERN server; the CCSD archives).

Stakeholders

The project distinguishes three levels of participants:

1. A small group of core stakeholders for the project management, the communication strategy, the initial development of the "MetaGrey" software and the "OpenSIGLE" archive and the connexion

of the most important grey collections. Today, three institutions are part of this core group: the TIB Hannover, the FIZ Karlsruhe and INIST-CNRS.

2. A limited group of European national and special libraries and documentation centres with important grey collections. The following institutions have already been contacted and have shown interest for a participation in this project: the British Library and the Cranfield University (United Kingdom), the Dutch Royal Library (Koninklijke Bibliotheek), the CNR (Italy), the CERN (Switzerland), the VNTIC (Russia), the State Technical Library and Academy of Sciences of the Czech Republic.

3. An unlimited number of potential participants with significant grey collections and databases. For instance, the Slovak Centre of Scientific and Technical Information, the Spanish CINDOC-CSIC, the Portuguese Fundacao para a Ciencia e a Tecnologia (FCT) Serviço de Informação e Documentação or the French ABES, CCSD-CNRS, and Cyberthèses. Actually, contacts exist with organisations located in 16 EU countries.

It will be possible to include EU research reports (CORDIS) and specific published report collections. Other contacts were taken with structures in the Maghreb countries.

Resource assessment

A preliminary resource assessment of the preparation, development and realisation of the MetaGrey project identified four domains with specific funding, staff and technical needs:

(1) Development of a metasearch engine for grey literature ("MetaGrey"):
1. Definition of content (searchable data, classification schema) and connexion interfaces.
2. Development of the software (with access to document delivery services).
3. Implementation on a server.

(2) Transformation of the SIGLE database in an open archive ("OpenSIGLE") hosted on an open archive server and OAI-PMH compliant:
1. Downloading of the SIGLE records and transformation into a harvestable structure.
2. Uploading of the OpenSIGLE data on an OAI server.

(3) Connexion of collections of European grey literature and of the "OpenSIGLE" archive to this metasearch engine:
1. Selection of collections and institutions (limited group).
2. Connexion of databases and catalogues (limited group).
3. Connexion of other collections (unlimited number).

(4) Functioning of MetaGrey Europe:
1. User assistance.
2. Search engine maintenance.
3. OpenSIGLE maintenance.

Up to now, a more detailed funding estimation exists only for the 2nd domain, the transformation of the SIGLE database. The INIST-CNRS information systems department made the following proposal (cf. communication to all former EAGLE member of June, 2005):

The submission of a Specific Targeted Research Project (STREP) to the European Commission (6th EU Framework Programme for Research and Technology Development) for an initial funding in 2005 was abandoned because of the too short delay after the EAGLE GA (deadline March 2005).

Risk analysis

Initially, six potential risks for the project had been identified. In italics, updated commentaries:

(1) Initial investment: No initial funding by the European Commission. No or limited funding by the EAGLE association. Restricted funding by the leading institutions. *Effectively, no EU funding in 2005, no EAGLE funding because of the unwillingness of the former members to spend "their" money for this project. No engagement by the "core institutions" up to now.*

(2) Language: Impossibility to select collections and data with a core language (English). *Theoretical problem, no real analysis of this topic up to now.*

(3) Bibliographic data: Too high complexity of data definition of the different types of documents, records and collections. *This needs analysis; nevertheless, the former EAGLE members already faced this problem – successfully. So why not now?*

(4) Connexions: Technical difficulties and limited motivations of institutions of the "limited group". *May be a real problem for some organisations.*

(5) OpenSIGLE: Technical and/or legal problems (intellectual property of SIGLE records) with the transformation of the SIGLE database into an open archive. *All former EAGLE members declared last summer that "their" SIGLE records could be temporarily archived by the EAGLE operating agent, the FIZ Karlsruhe, to be used by a future non-for-profit OAI compliant database. The technical aspects seem relatively simple.*

(6) Liquidation of EAGLE: Delay. No solution for the intellectual property of the SIGLE records. *The delay is caused by the liquidation of the bank accounts and the suppression in the Luxemburg Register for Societies and Commerce. Apparently, no legal problems.*

Table 1: Risk analysis

	Potential risk	Likelihood	Impact on project	Evaluation	Contingency
1	*Initial investment*	High	Medium	Most important risk	Alternative funding? Self-funding?
2	*Language*	Medium	High	Important risk	Acceptance of multilingual data
3	*Data*	Low	High	Low risk	Reduction of complexity
4	*Connexion*	Low	High	Low risk	Communication, adding other interfaces?
5	*OpenSIGLE*	Low	Low	Low risk	No alternative
6	*EAGLE*	Low	Medium	Low risk	Legal problems?

To resume, two risks (# 5 and 6) no longer exist. Two other risks (#2 and 3) should be analysed more thoroughly, based on concrete data and not on abstract considerations; this may need some months and could be done while prepare the funding demand. The risk #4 should also be assessed during the initial phase of the project. The most important risk, the need for financial resources (risk #1) today has no solution. For different reasons, most of the former EAGLE members seem to be exposed to financial shortcuts and a "fragile" budget situation. Without funding, no development. But without project rationale and minimal consensus on what to do and with whom, no needs analysis and no funding demand. So, where to start?

Two other risks have been excluded: Actually, there seems to be no concurrent project with analogue objectives or scope. The technology risk can be neglected because of the maturity of the used software concept.

Outline of costs and benefits

The economic model of MetaGrey Europe is a sustainable, low-budget project, based on limited funding of the initial phase (software development and server, marketing and communication strategy) and use of existing resources (staff and budget of member organisations) for the functioning of the metasearch engine (maintenance, assistance) and the OpenSIGLE database (maintenance).

Apart from the connexion to the search engine, the participating organisations will not assume any financial obligation. Technical input may be necessary only in case the partners change their own parameters (database format, cataloguing standards etc.).

The benefits for the participating organisations are twofold: the metasearch engine increases the visibility and searchability of their special grey collections through a new and dedicated "access point"; and it potentially increases the use of their document supply and ILL services through the link to their own online interfaces.

Project management approach

The proposed steering committee is composed of four organisations, each of them with specific roles:

INIST-CNRS (France): Project coordination, submission of EU funding project, definition of searchable data and core collections, user assistance. Development and host of OpenSIGLE.
TIB Hannover (Germany): Development, host and maintenance of MetaGrey. Definition of searchable data and core collections.
FIZ Karlsruhe (Germany): Liquidation of EAGLE. Development of OpenSIGLE. Definition of searchable data.
GreyNet (Netherlands): Development and realisation of marketing and communication strategy, promotion of the new network.

The cooperation between these organisations could be facilitated by a listserv with restricted access and hosted by one of these organisations.

Marketing and communication strategy

The marketing and communication strategy is developed and coordinated by GreyNet. The multilevel strategy includes several vectors of communication and information:

EAGLE network: Information and promotion through the existing mailing list of former and actual EAGLE member organisations.
CNRS network: Information and promotion through the recently created CNRS network of direct scientific communication organisations.
GL conferences: Communications at the GreyNet international conferences on grey literature 2005 (France) and 2006 (United States).
Other conferences: Information and promotion at selected national and international conferences and events on scientific information and open access.
Listservs: Information and promotion through selected national and international listservs of library and information professionals and scientists (ADBS-Info, Biblio.fr, INETBIB, JISC listservs, SPARC-Europe).
Publications: Information and communication in selected LIS serials (for instance, *Documentaliste*, *Bibliotheksdienst*, *Interlending and Document Supply*, *The Grey Journal, Electronic Library*, *Aslib Proceedings: New Information Perspectives*, *Library Hi Tech News*, *Online Information Review* etc.).

Timelines and milestones

Seven key stages have been identified. The timelines mainly depend on the submission, evaluation and negotiation of the initial funding demand.

Table 2: Timelines

	Project phase	Content	Period	Commentary
1	Preparation (1)	Concept, project team, network	2004 November – 2005 January	
2	Preparation (2)	Elaboration and submission of a funding proposal	2005 January – 2005 March	Depends on FP6 calls for proposals
3	EAGLE	Liquidation of association	2005 January – 2005 June	
4	Development (1)	MetaGrey Europe	2005 April – 2005 December	Depends on #2
5	Development (2)	OpenSIGLE	2005 April – 2005 December	To be specified, depends also on #3
6	Connexion (1)	First group	2005 September – 2006 March	During #4
7	Connexion (2)	Unlimited number	2006 April – 2007 March	Following #4 and 6

One year after the first project draft, the initial timelines are no longer realistic and need to be updated. Nevertheless, the table shows clearly two "milestones": first, the need to find a minimal consensus on the concept (#1); second and based on this consensus, the need to find an initial funding (#2).

Concluding remarks, open questions

One year after the first MetaGrey project brief, there are only two "successes" of the former EAGLE network: the legal and financial liquidation of the European association, and the temporary safeguard of the SIGLE records by FIZ Karlsruhe.

Once the SIGLE input stopped, there is no longer any technical or strategical coordination between the former EAGLE centres. Most of our organisations seem preoccupied with their specific problems, putting their priorities elsewhere, trying to redefine their objectives and activities in the environment of the Open Access movement, the Google Print initiative and the EU intellectual property rights.

In some way, we are facing a similar situation as 20 years ago, before the creation of the EAGLE network. Will there be another major European initiative for grey literature? And if so, when? Will the European projects for open institutional repositories and digital libraries attract too many financial, human and technical resources without leaving space for another grey project? Do we need some more years to measure the real interest of such a project? Open questions. Are our organisations only willing to give an answer now?

Glossary

ABES: Agence Bibliographique de l'Enseignement Supérieur (France), a public agency that hosts the French national academic union catalogue.
CCSD-CNRS: Centre de Communication Scientifique Directe of the French CNRS with a number of open archives (preprints, electronic theses and dissertations, articles, proceedings).
CERN: European organisation for nuclear research (Switzerland).
CINDOC-CSIC: Centro de Informaçion y Documentaçion Cientifica, Centre of Scientific Information and Documentation of the Spanish Higher Council for Scientific Research.
CNRS: Centre National de la Recherche Scientifique, a public organisation for scientific and technological research under the authority of the French Ministry of Research (France).
CNR: Consiglio Nazionale delle Ricerche, public research organisation (Italy).
DFG: Deutsche Forschungsgemeinschaft, public research agency under the authority of the German Ministry of Research (Germany).
EAGLE: European Association for Grey Literature Exploitation (Luxemburg).
FIZ Karlsruhe: Fachinformationszentrum Karlsruhe, a non-profit scientific service institution that produces and markets scientific and technical information services (Germany).
Grey literature: "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing" (GL'97 Luxemburg Convention).
GreyNet: Grey literature network service (Netherlands).
INIST-CNRS: Institut de l'Information Scientifique et Technique (France), the CNRS documentation centre for scientific and technical information.
LARA: Libre Accès aux Rapports Scientifiques et Techniques, an open archive of French reports under construction by INIST-CNRS (France).
MAGiC: A former UK project sponsored by the British Library and the Research Support Libraries Programme and hosted by the Cranfield University to enhance awareness, access and utilisation of key collections of technical reports.
MPG: Max-Planck Gesellschaft, a public organisation for scientific and technological research under the authority of the German Ministry of Research (Germany).
OAI-PMH: Open Archive Initiative-Protocol of Metadata Harvesting.
RRZN: Computer Centre of the Lower Saxony, University of Hannover (Germany).
SEL: Search Engine Laboratory of the University of Hannover (Germany).
SIGLE: System of Information for Grey Literature in Europe.
TIB Hannover: Technische und Informationsbibliothek Hannover, most important German library for technical and scientific information.
VNTIC: Scientific and Technical Information Centre (Russia).

Author Information

Joachim Schöpfel graduated from the University of Hamburg in 1984. A research assistant and lecturer at the University of Hamburg, Department of Developmental and Educational Psychology, from 1985 to 1990, he obtained his Ph.D. from the same university in 1992. He is presently head of the library and document delivery department at the French Institute of Scientific and Technical Information and teaches Culture and Society (1992-2001) and Documentation (from 2001 on) at the University of Nancy. He is member of the UK Serials Group, associate editor of TGJ, a member of the Editorial Board for Interlending & Document Supply, and a member of the former EAGLE, European Association for Grey Literature Exploitation. Email: schopfel@inist.fr

Repositories, Tools for NGOs Involved in Public Health Activities in Developing Countries

June Crowe and Gail Hodge
Information International Associates, Inc., IIa, USA

I. Background

Information International Associates, Inc. (IIa), a woman-owned, small business specializing in information management, performs research for government and commercial clients. IIa's Research Division has researched over 60 studies in the area of public health in less developed countries and regions. The information needed to complete the studies covers a range of health system topics that include statistics for health personnel, infrastructures, disaster preparedness, health financing, and other factors that impact public health care. In our experience, the search for global public health information can be both complex and frustrating. Although this information is often considered "open source" in many countries, it may be difficult to obtain, especially if governmental web sites are not readily available or completely viable, either not functioning at all or only functioning intermittently. In addition, the health information available from the site may be out of date. Many developing countries experience catastrophic events that impact access to public health information. For example, if a country has experienced political instability, natural disaster, civil strife, or other events, the existing medical system may easily be overwhelmed, with resulting health information being minimal at best. Therefore, various resources are consulted for global public health information, including electronic journals, databases, web sites, reference sources, library catalogues, bookstores, newspapers, statistics, electronic books, maps, directories, and grey literature sources. Non-governmental organizations (NGOs) are one of the primary sources of grey literature used for researching healthcare information for developing countries.

In this publication we describe the role of NGOs in global public health information, elaborate on the problem with NGO grey literature, and describe a possible solution based on the repository concept.

II. Role of NGOs in Public Health Care

NGOs play an important role in global health activities and health research. It is difficult to quantify the number of such organizations. There are 43,500 development organizations listed in the 2005 edition of the Directory of Development Organizations.[1] According to the World Health Organization (WHO), between 70-95% of health services in emergency situations are delivered by NGOs.[2] The work of many NGOs overlaps, making it difficult to discern those that have a primary focus on health. For instance, NGOs with a focus on sustainable development may also be concerned with poverty, education, and health. In Ecuador, for example, Fundacion FEVI is a non-profit NGO that facilitates intercultural education and volunteer community service. FEVI arranges community service visits from people all over the world to small communities in Ecuador. They work with healthcare centers in addition to centers for elderly people, women's organizations, indigenous communications, human rights organizations, and public schools.[3]

NGOs play key roles in health systems of developing countries and are recognized for developing innovative initiatives and programs that address health issues. They possess extensive knowledge of local conditions and can provide baseline data on health infrastructure and personnel and major obstacles to improvement. NGOs are often able to reach segments of rural populations that governments neglect or do not target as a priority.[4]

NGOs have roles in public health from the grass roots level to the national and international levels. The WHO has created the following table depicting the health system functions and examples of roles of civil society organizations (CSO)—a type of NGO.[5]

Health System Function	Examples of Roles of CSOs
Health services	Service provision; facilitating community interactions with services; distributing health resources such as condoms, bed nets, or cement for toilets; and building health worker moral and support.
Health promotion and information exchange	**Obtaining and disseminating health information; building informed public choice on health; implementing and using health research;** helping to shift social attitudes; and mobilizing and organizing for health.
Policy setting	Representing public and community interests in policy; promoting equity and pro-poor policies; negotiating public health standards and approaches; building policy consensus, **disseminating policy positions**; and enhancing public support for policies.
Resource mobilization and allocation	Financing health services; raising community preferences in resource allocation; mobilizing and organizing community co-financing of services; promoting pro-poor and equity concerns in resource allocation; and building public accountability and transparency in raising, allocating, and managing resources.
Monitoring quality of care and responsiveness	Monitoring responsiveness and quality of health services; **giving voice to marginalized groups**, promoting equity; representing patient rights in quality of care issues; and **channeling and negotiating patient complaints and claims.**

Some of these roles already involve research and information dissemination as indicated in the highlighting of those functions in the table above. Although NGOs promote and advocate for public health, as well as performing other functions in the health systems, there is a need to more effectively include NGOs in the knowledge production and diffusion of public health information in developing countries and to better manage the knowledge output.

This can be accomplished through dedicated partnerships with appropriate organizations and agencies. These roles could easily be expanded to include more of a role in health research knowledge diffusion because they are "on the ground" and know what is happening firsthand. As a researcher needing access to current information, having ready access to reports, online newsletters, or blogs generated by these NGOs would be extremely valuable.

As a research organization, IIa and its clients need persistent access to documents from all organizations/agencies involved in health activities in these countries. We found that for a country study completed in 2003, 18% of the urls in the study are now dead, 3% have changed, 4% have moved or been re-directed, and 29% were no longer there. The reliability for older studies is even worse. A quick look at the urls from a study completed in 2000 revealed that only 30% of the urls were active and accessible, about 62% were dead, and about 8% had moved or were re-directed.

III. Repository Definition

Given the importance of NGO information and the problems mentioned with accessing this information, what could be done to improve the situation? A repository is one possible solution to the problem of locating NGO public health information, particularly reports and studies. What is a repository? A repository is a digital collection that captures and preserves the intellectual output of an institution, agency, or organization. The development of repositories has principally been by universities to collect and manage the output of students and faculty, however, they could easily be developed and used by NGOs.

Benefits of a Repository:

The benefits to researchers of having one or several resources for locating and accessing this grey literature are obvious. Significant time would be saved and there would be more assurance that the information would be updated and preserved over time. However, there are additional benefits beyond the traditional functions, such as data collection, searching, capacity building, and knowledge management.

Data Collection and Coordination:

An NGO repository would facilitate the identification of public health problem areas, data collection, and problem solving for decision makers. In addition to making health information about these areas more

accessible to researchers and decision makers, use of the repositories could facilitate coordination among NGOs and others who want to provide assistance to these countries. A repository could be useful in identifying NGOs that have had experience in certain areas by preserving a record of the NGO's work. It would then be easier to discern where resources could best be used.

Building Health Capacity in Developing Countries:
Repositories could serve as a mechanism for building health capacity knowledge and diffusion in developing countries. For example, a repository could be the mechanism for introducing new perspectives, or technical expertise, and a way to capture a snapshot of what is happening with disease control, vaccinations, health education, etc. In a recent article on open access archiving, Leslie Chan pointed out that scientific progress is greatly hampered in developing countries by their inability to have access to essential medical literature.[6] A repository of NGO reports and documents could centralize access to global NGO health-related documents, particularly to those documents from other developing countries that are most relevant for public health, social, and technical situations of a developing country.

Knowledge Management Tool:
There are direct benefits to NGOs. NGOs that publish many reports and documents would benefit from a repository to support content and knowledge management activities. The management of information about research and projects already conducted can support the re-purposing of that information to enhance development, marketing, and outreach efforts, as well as the creation of future funding proposals. For example, several years ago IIa helped Conservation International, an international environmental NGO, identify ways it could better capture and manage the knowledge created by its individual projects and principal investigators in environmental hot-spots across the globe. Development of Conservation International's system continues to this day in the continued implementation of a content management system for creating, disseminating, locating, and re-purposing its web site content. Similar approaches would be reasonable for large public health NGOs.

A repository is a major component of an information asset management system that would manage and support every aspect of information creation and dissemination. Information asset management is the ability for people to get whatever information they need, anywhere, anytime, and in compliance with the organization's policy. As part of this function, a repository would enable the NGO to identify best practices, focus on key projects and their users, and look for partnering opportunities.

IV. Barriers/Challenges to Repository Development

Unfortunately, there are many obstacles to the development and use of such a repository or series of repositories, because of insufficient funds earmarked for health problems in developing countries, inefficient application of resources, and the lack of technology transfer.[7] For purposes of this publication, three barriers/challenges are highlighted – organizational structure and politics, funding, and collection development policies.

Organizational Structure and Politics:
A key challenge in establishing a repository for NGOs is their wide variation in organizational structure that includes confederations, federations, separate and independent organizations, and variations of these.[8] With all these structures, the challenge is to create a model that will facilitate the transfer/capture of documents from all of them, since some NGOs do not work together due to political or philosophical differences.

Funding:
The funding source impacts how and what information an NGO releases and distributes, as well as its fiscal ability to create reports for release. For example, a faith-based NGO may choose not to report on contraceptive needs or abortions, although they may have this information. Also, funding can determine which NGOs support what efforts in what countries. If several NGOs with a similar purpose, such as HIV/AIDS prevention, obtain funding from a single source, the probability of obtaining their documents for a repository is greater than if they were funded by a variety of sources, because it would perhaps eliminate some of the constraints on releasing material to the public.

NGOs may be funded by foundations, religious organizations, special interest groups, governments, international or national organizations, or any number of other methods. Their respective funding sources may impact the types and accessibility of reports or other information published. Insufficient funds, of course, may mean little or no publicly accessible information and/or the lack of a publications program. NGO funding sources can also impact the willingness to share information because of political or other reasons.

In 2003, the WHO examined the funding sources of NGOs with whom they had official relationships. The majority of NGO funding (41%) came from admission fees and member dues. The next largest funding

source was from unspecified grants (21%). The remainder of the funding came from other fund raising (12%); NGO grants (4%); company funding grants (3%); government and inter-governmental grants (4%); Congress and publication fees (9%); and government contracts and consultancy fees (6%).[9] It should be noted that there are more NGOs that have unofficial relationships with the WHO, and so are not reported in these statistics. As civil societies have continued to increase in number, funding has increasingly come from governments (approximately USD 1 billion) and other non-governmental agencies (about USD 1 billion) annually.[10]

Collection Development Policy:
Given the issues outlined above, it may be impossible and perhaps not even desirable to collect all NGO documents in a single collection. A policy for collection development would need to be agreed upon even among a small group of NGOs with similar interests, such as HIV/AIDS or women's health. Another consideration would be the variation in the types of documents published by NGOs. Not all NGOs publish annual reports. Would preliminary reports or field reports with raw data be included? What about surveys or training manuals? These questions would need to be balanced by the current need for health information in the country.

V. Relevant Web Sites in the Public Health Domain
Despite these challenges, there are several examples of web sites that begin or partially fill the need as a repository for grey literature in public health.

- United States Aid In Development (USAID) http://www.usaid.gov/
  - The USAID library focuses on sustainable development with the primary mission of serving the information needs of USAID staff.
  - USAID documents, reports, publications, and project summaries can be publicly accessed through the Development Experience System (DEXS), which has over 100,000 records with some 20,000 available for electronic download. Its purpose is primarily to strengthen USAID development projects, activities, and programs and make them publicly available. DEXS offers four major services: USAID contractors/grantees can (1) submit documents to the system, (2) search the DEXS database, (3) order documents (paper, electronic, CD), and (4) subscribe to free USAID reports via email.
  - The DEXS submittal process is described in documentation available on the web site. Documents for submittal should include those documents which describe the planning, design, implementation, evaluation, and results of development assistance activities which are generated during the life cycle of the program or activity.

- Human Info NGO http://humaninfo.org
  - Uses Greenstone software.
  - Has 35 to 40 Humanitarian CD Libraries on the Joint United Nations Programme on HIV/AIDS (UNAIDS), community development, food and nutrition, health library for disasters, Rural Hygiene in Africa, Africa Collection for Transition, as well as others.
  - About 5,000 copies of each library are distributed annually.

- World Health Organization (WHO) http://www.who.int/en
  - Site can be searched by country or health topic.
  - The WHO Library and Information Networks for Knowledge (LNK) provide access to WHO-produced recorded information and to worldwide health, medical, and development information resources. The Information Networks for Knowledge provides technical support to help improve the health-related information transfer structure in developing nations. The services are primarily for WHO headquarters, regions, and country offices; ministries of health and other government offices; health workers in Member States; other UN and international agencies; and diplomatic missions. The WHO library programs help regions and developing countries achieve self-sufficiency in providing information services to the health sector. The library has over 70,000 bibliographic records and 30,000 links to full text documents.
  - Blue trunk libraries concept was developed by the library for installation in district health centers in Africa to compensate for the lack of current medical and health information. The collection of more than 100 books on medicine and public health is shipped in blue trunks fitted with two shelves. It is not known if CDs are part of this shipment.
  - Unknown if there is a repository for NGO grey literature and/or the submittal process.

- Global Health Council http://www.globalhealth.org
  - World's largest membership alliance of healthcare personnel, NGOs, organizations, government agencies, and other public and private institutions. Mission is to ensure that information and resources are available to those who strive for improvement and equity in global health.

- Advocacy group who reports on world health problems to governments, public and private organizations, and the global health community.
- Publications section includes a variety of press releases, reports from NGOs and other agencies, notes from the field, annual reports of the Council, and other publications.
- Unknown if there is a repository and/or the submittal process, but it does have a member login/password.

- British Library of Development Studies (BLDS) http://blds.ids.ac.uk/blds
  - BLDS is Europe's largest library on international development at the Institute of Development Studies in Sussex.
  - Extensive collection of government publications, NGO publications, World Bank, United Nations, World Trade Organization, and research institutes worldwide. They also have over 200 development journals that are scanned and selected articles added to the BLDS catalog.
  - Online library catalogue can be searched at http://blds.ids.ac.uk/blds/search/search.html.
  - Document delivery is via interlibrary loan; some items free to download.
  - Not a repository, but a great prospect for finding NGO material.

Most of the web sites identified above are searchable by geographic area and have some project report summaries. Some sites are subject oriented, such as the Human Info NGO and the Global Health Council. The Human Info NGO has created repositories on CDs, by subject, for distribution to developing nations and other interested parties. USAID has a growing database of health information from its partners and a defined process for the submittal of documents from NGOs to DEXS. The BLDS collects material in many subject areas and provides, via email notification, updates to the collection. However, the documents are not always in an electronic format, free, or current, though the library makes every effort to get documents to people who request them in an efficient manner.

The WHO web site has vast resources and pointers to documents, but to our knowledge makes no effort to collect NGO material. The WHO library is primarily for WHO and its associated organizations. The WHO maintains relations with other international organizations and external partner NGOs. Formal relations with NGOs require that certain criteria be met. In December 2004, there were 189 NGOs that had official relations with the WHO.[11] The WHO also maintains informal working relations with other NGOs. Regional or national NGOs affiliated with international NGOs are usually charged with developing and implementing a program of collaboration with the regional and national levels of WHO to ensure implementation of health-for-all strategies at the country level.[12] Although WHO has the Library and Information Networks for Knowledge (LNK) that provide access to WHO-produced recorded information and to worldwide health, medical, and development information resources, it does not, to our knowledge, have a repository for their NGO documents or current initiatives underway for such a repository.

These examples, while scattered, could serve as the basis for more consistent repository development. However, a more community-wide effort is needed to achieve this goal.

VI. Repository Models

Assuming that the barriers could be overcome, there are several repository models that may be viable for the further development of these web sites or for the development of more formal repositories by NGOs. These include:

PubMed Central is a digital archive of life sciences and biomedical journal literature, developed and managed by the National Center for Biotechnology Information at the U.S. National Library of Medicine (NLM). This system features voluntary participation, flexible public release dates of materials deposited by contributors, and retention of copyright by the author or corporate sponsor.[13] In May 2005, the National Institutes of Health's (NIH's) new policy on enhancing public access to archived publications was implemented. Authors are now requested to submit an electronic version of the author's final manuscript to PubMed Central upon acceptance for publication. The Policy is intended to: (1) create a stable archive of peer-reviewed research publications resulting from NIH-funded research; (2) ensure the permanent preservation of these vital, published, research findings; (3) secure a searchable compendium of these peer-reviewed, research publications that NIH and its awardees can use to manage more efficiently and to better understand their research portfolios, to monitor scientific productivity, and ultimately, help set research priorities; and (4) make published results of NIH-funded research more readily accessible to the public, healthcare providers, educators, and scientists.[14] Such a model may work for NGOs, especially if they have partners or other organizations assisting them in their work.

DSpace at Massachusetts Institute of Technology (MIT) is a digital repository created to capture, distribute, and preserve the intellectual output of MIT. DSpace features access to content through the web. Similar to PubMed Central, DSpace at MIT (and other DSpace institutions) uses the submission model. Authors from among the faculty provide their final manuscripts to the DSpace system. Some

initial information is provided along with the manuscript, and then a "bibliographic record" or metadata file is finalized by library staff. The manuscripts are grouped into collections that represent particular communities of interest, academic colleges, or disciplines. It offers the advantage of digital distribution and long-term preservation for a variety of formats, including text, audio, video, images, datasets, etc., and the opportunity to provide access to all the research of the institution through one interface.[15]

For small NGOs, the approach of the **Association of Learned and Professional Society Publishers (ALPSP)**, http://www.alpsp.org/default.htm, may be of interest. These are "community" organizations that have been created to build the capacity of the small publishers using the digital environment.

Google Scholar is a search service that allows users to search for scholarly material across the web and view abstracts of the material found.[16] Special metadata is no longer necessary for all the pre-publication versions of papers which are deposited anywhere on the web[17]. Although many people still claim that they can not find information on Google or Google Scholar, this eliminates the need for an NGO to develop an elaborate search system for its own documents.

Open Access Archives (OAAs) is another model that would encompass the variety of types of information published by NGOs. Using self-archiving, the NGO would submit documents to an institutional or community archive of its choice. This model is favored in the article by Leslie Chan as a quick way to build research capacity in developing nations.[18]

Greenstone digital library software is open-source and multi-lingual and provides a way of organizing information and publishing it on the Internet or on CD-ROM. An NGO could use this software to build its own digital libraries. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato and developed and distributed in cooperation with the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the Human Info NGO. The Human Info NGO is based in Antwerp, Belgium, and works with United Nations agencies and other NGOs. They have established a worldwide reputation for digitizing documents in human development and making them widely available and free to developing nations and on a cost-recovery basis to others. A new development with Greenstone is the ability to build collections on a remote server while using a modified version of the Greenstone Librarian Interface, so there is no need to run Greenstone locally. Multiple users can collaborate on the same collection, although not simultaneously.[19]

The software for the basic development of a repository is available, and most of it is open source. Greenstone has the additional benefit of being multi-lingual and portable. However, the submission and/or harvesting approaches for capturing grey literature must be carefully considered, as would a collection development policy.

VII. Conclusions

As an information management and research company, IIa believes that grey literature is a vital component of public health information, particularly in developing countries. One or more repositories of grey literature from across NGOs in the public health community would be beneficial to researchers seeking to use this information. While there are many barriers to achieving such a repository, the benefits are great, and there are a variety of models that could be used. There are several existing web sites that begin to fill this need, but a more community-wide effort is needed to provide consistent, complete, and effective coverage of this grey literature. The benefits to the research community are obvious, but the ultimate benefit is to advance the use of public health research to improve the lives of people world-wide.

References

1 Directory of Development Organizations. 2005. http://www.devdir.org
2 World Health Organization. WHO and Civil Society Linking for Better Health. http://www.who.int/civilsociety/documents/en/CSICaseStudyE.pdf
3 Fundacion FEVI. http://www.fevi.org/
4 Partnership with NGOs and Civil Society. International Federation of Agricultural Development. http://www.ifad.org/ngo/index.htm
5 Strategic Alliance: The Role of Civil Society in Health. Civil Society Initiative. External Relations and Governing Bodies. World Health Organization. Discussion Paper No. 1 CSI/2001/DP1. http://www.who.int/civilsociety/documents/en/alliances_en.pdf
6 Chan, Leslie, B. Kirsop, and S. Arunachalam. Open Access Archiving: the Fast Track to Building Research Capacity in Developing Countries. http://www.scidev.net/ms/openaccess
7 Delisle, Helen, et al. The Role of NGOs in Global Health Research for Development. Health Research Policy and Systems. Vol. 3(3). 2005. http://www.health-policy-systems.com/content/3/1/3
8 NGOs and Organizational Structure: Challenges and Opportunities. October 2003. www.poli-sci.utah.edu/~pitima/1ec7.ppt
9 World Health Organization. Some Statistics on NGOs in Relations with WHO. http://www.who.int/civilsociety/csi_statistics/en/print.html
10 World Bank Funding for Civil Society. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/CSO/0,,contentMDK:20094251~menuPK:220439~pagePK:220503~piPK:220476~theSitePK:228717,00.html
11 World Health Organization. Relations with Other International Organizations and External Partners NGOs. http://w3.shosea.org/en/Section1257/Section1259_5127
12 World Health Organization. Relations with Other International Organizations and External Partners NGOs. http://w3.shosea.org/en/Section1257/Section1259_5127
13 Johns Hopkins University. Scholarly Communications Group. Publishing Models. http://openaccess.jhmi.edu/publishing.cfm
14 NIH Public Access Policy. http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html and http://www.nih.gov/about/publicaccess/index.htm
15 Johns Hopkins University. Scholarly Communications Group. Publishing Models. http://openaccess.jhmi.edu/publishing.cfm
16 Sullivan, Danny. Google Scholar Offers Access to Academic Information. Search Engine Watch. http://searchenginewatch.com/searchday/article.php/3437471
17 ALPSP. Preprint and postprint repositories and their impact on publishing. http://www.alpsp.org/events/2005/PPR/default.htm
18 Chan, Leslie, B. Kirsop, and S. Arunachalam. Open Access Archiving: the Fast Track to Building Research Capacity in Developing Countries. http://www.scidev.net/ms/openaccess
19 Greenstone Digital Library Software. http://www.greenstone.org/cgi-bin/library

A Survey of Open Access Barriers to Scientific Information: Providing an Appropriate Pattern for Scientific Communication in Iran

Mohammadreza Ghane
University of Tehran, Iran

Abstract
Serial pricing crisis and permission crisis restrict scholars to their scholarly findings. Scientists as a vital component of scholarly communication are losing their control on it. These crises along with library budget cutback interrupted the free flow of scientific information. This case study investigated academics' views of Shiraz University (Iran) on open access publishing and its four channels. Findings showed that in spite of their low familiarity with open access materials 92% of them had positive view on open access movement. 70% of respondents chose open access journal for their publishing model and the second vehicle is self-archiving (62.5%). Majority of academics knew pricing crisis and permission crisis as an obstacle to their scholarly information.

INTRODUCTION

With the advent of Internet and WWW, three components of scholarly communication, i.e. publishers, authors and libraries confronted with rapid changes in scholarly publishing. The new opportunity brought about by journal pricing crisis and new information technology. The Association of Research Libraries statistics (ARL, 2003) showed that the average annual increase of the serial unit cost was9% since 1986 and the consumer price index for this period increased 64%. Spiral pricing levels of scientific journals and library budget cutback restrict institutions in providing needed journals. Consequently, great numbers of scientists in the world, especially in developing countries, are unable to access the research findings they need. Basically, there is a gap between the large amount of the scholarly materials that libraries can provide and the literature that scientists need. According to ARL's statistics this gap has widened over 1986 to date.

Scholars are losing their control on a system that they created. Journal spiral prices on the one hand and library budget cutback on the other hand interrupted free flow of scholarly information. Scientists' findings are given away to commercial publishers. Publishers, then sell them back to their libraries at unjustifiable prices. Consequently, **A:** university scholars and their peers in different countries have access to less and less scientific materials published in their fields, **B:** scientists haven't professionally many incentives to work and **C:** at least**,** this affect science at national level**.** As a result of the problems described above many scholars and their institutions, in a global attempt, decided to make literature freely available (BOAI, 2002).

This study intends to survey academics of Iran universities on their attitudes toward open access publishing and providing an appropriate pattern for scholarly communication. There are some research studies which are close to this goal. A large-scale survey of journal authors' opinion was carried out on 4/000 senior researchers from 97 countries (Rowlands, Nicholas, and Huntingdon). This survey investigated authors' views on current journal system and open access publishing. Low awareness (82%) of research community of open access, in spite of their positive attitudes toward open access movement, revealed the urgent need to raise awareness of them to this issue. A survey of journal authors on behalf of the JISC and OSI has been carried out by Key Perspective Ltd. (2004). According to this survey authors' awareness of open access is high. The reason for publishing in open access models is the principle of free access to research findings (90%).

A series of studies (RoMEO Projects) funded by UK JISC investigated the intellectual property rights issues relating to open access movement. The aim of study 1(Gadd, Oppenheim and Probets, 2003) was to examine the attitudes of three parties (academic staff, universities and publishers) towards copyright ownership and the impact of copyright ownership on the open access movement. Findings showed that "self-archiving is not best supported by copyright transfer to publishers". Respondents from self-archivers and non-archivers are compared in RoMEO studies 2 (Gadd, Oppenheim and Probets, 2003). The main purpose of this study "was to ascertain how authors wanted to protect their self-archived research papers in order to develop the right metadata." Study 3 (Gadd, Oppenheim and Probets, 2003) wanted to ascertain how academics expect to use others' papers and also investigated if there was any significant difference between the attitudes of two groups (academics-as-authors and users) towards using and protecting research papers. 99% of academics-as-users expected to display and print open-access materials either freely or under limits or conditions. The findings showed that academics-as-users do not perform all the activities (excerpt, aggregate and annotate) with open-access research papers that academics-as-authors would allow. The study indicated highly significant differences between the two groups on all permissions except display, excerpt and save. In general, academics-as-authors are more liberal on using their works than academics-as-users.

OBJECTIVES

This study seeks to determine the following regarding open access movement:

- Academics' attitudes toward open access
- Academics' awareness of open access
- Academics' attitudes toward self- archiving
- Academics' attitudes toward institutional repository
- Academics' attitudes toward subject-based repository
- Academics' attitudes toward open access journal
- Academics' attitudes toward pricing crisis
- Academics' attitudes toward permission crisis
- Academics' reasons for not publishing via open access channels
- Academics' views on the usage of open access materials as-users and authors

METHODOLOGY

This is a case study that presents the findings of a survey of academics' views on the open access issues. The survey population consists of Shiraz University academics. According to statistics of Ministry of Science, Research and Technology of Iran entitled ***Faculty Members in Iranian Higher Education Institutes, Year 2003***, 377 doctorial faculties were employed in Shiraz University. The subject disciplines chosen for study include Humanities (116 faculties), Science (85 faculties), Engineering (98 faculties) and Agriculture & Veterinary (78 faculties). The sample size is calculated from a preliminary sample population variance, which is estimated 0.05 (s^2=0.05).

According to the formula $2 Z_{\alpha/2} S/\sqrt{n} = 0.129$ the sample size is 50 by using 95% confidence level and 12.9% confidence interval (The sample size can also be calculated using the software designed by Custominsight.com). The respondents were randomly chosen from the above source. The same number of respondents was chosen for each discipline. The survey instrument was a questionnaire, which was sent to the respondents' address and 41 returned. Gathered data which present in figures and tables analyzed through descriptive statistics (demographic) and inferential statistics (Chi-square test) to investigate academics' views and awareness of open access and its four channels (Self- Archiving, Institutional Repositories, Subject- based Repositories, Open Access Journals), their attitudes toward copyright law, and the reasons for not publishing through an open access outlet.

RESULTS

The demography of respondents was as follows:

- **Subject discipline**

Respondents were asked to name their subject discipline. The disciplines were divided into four categories: Engineering, Science, Humanities and Agriculture & Veterinary. Figure 1 below shows the breakdown.

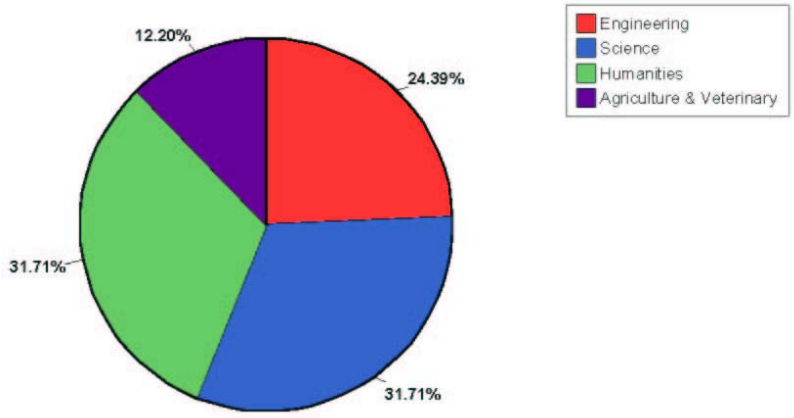


Figure 1: Subject disciplines of respondents

- **Respondents' academic status**

Q4 asked respondents to name their academic status, which is divided into three categories: full professor, associate professor and assistant professor. Figure 2 shows the breakdown.

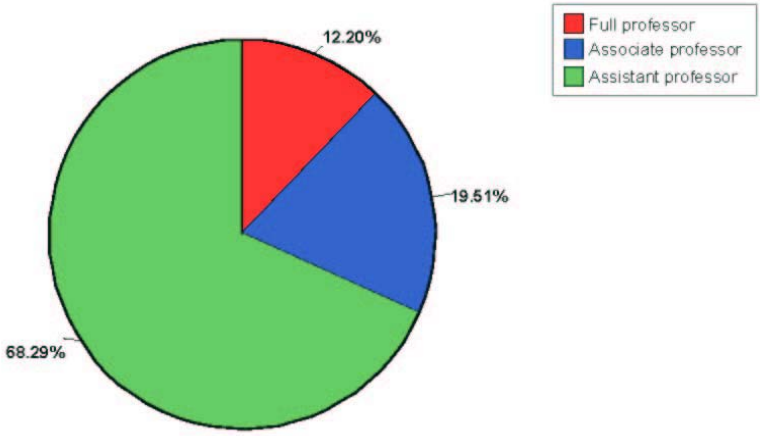


Figure 2:Academic status of respondents (Q4)

- **Length of time as academics**

Figure 3 shows academics' responds to Q6: How long had they worked in academia? Four options were given: 5 years or less, 6-10 years, 11-15 years and more than 15 years.

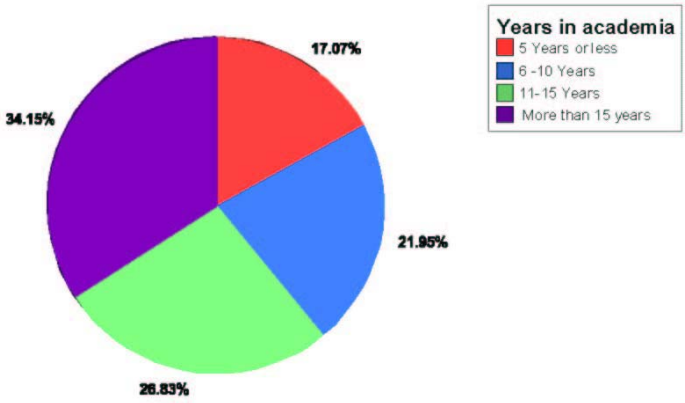


Figure 3: Survey respondents by length of academic service (Q6)

34 percent of respondents were long-standing members of the academic community, with over 15 year services.

- **Research papers published by academics**

Figure 4 indicates academics 'responds to Q7: How many research papers had you published in your career? Approximately half of the respondents (48.5%) had published ten or fewer papers during their careers; and 30% published between 11- 30 papers. 65 percent of respondents were member of academic community between 5- 15 years and they had published 48.5 percent of papers.

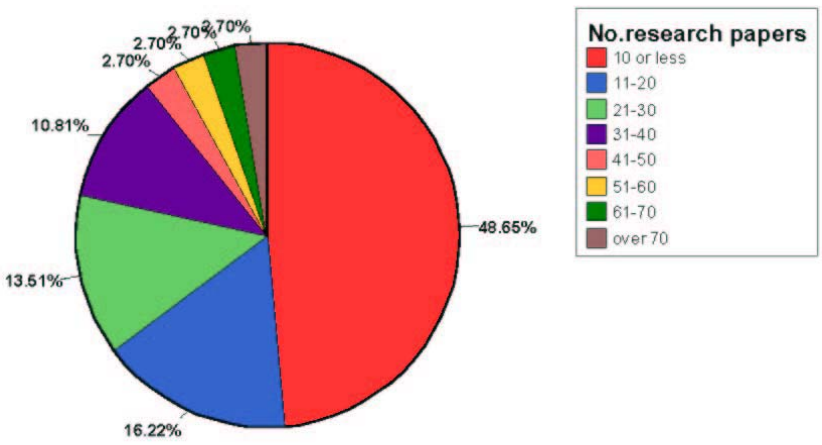


Figure 4:Research papers published by respondants (Q7)

- **Academics' attitudes to open access publishing**

Respondents were asked to indicate their attitudes toward open access publishing and its promotion. The responds were measured by Likert scale. 92% of academics supported to make their research papers available via open access publishing. Interestingly, there were no negative responds (disagree or strongly disagree). Nearly 7 percent of academics responded "No comment". Figure 5 shows the breakdown.

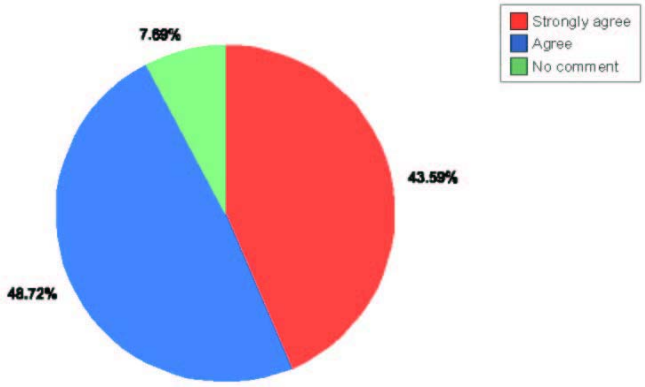


Figure 5: Academics' attitudes to open access publishing(Q 18)

- **Academic views on self- archiving**

62.5 percent of respondents are "agree" and "strongly agree" on publishing via self- archiving. Humanities are more likely than other disciplines to self- publish: 40% of humanities have been compared to 20%, 24% and 16% of Science, Engineering and Agriculture & Veterinary, respectively.

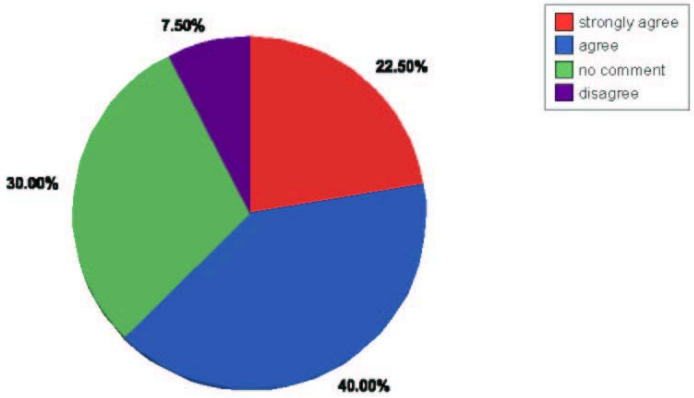


Figure 6 : Academics' views on publishing via self archiving (Q 20)

- **Academic views on OA journal**

Q22 asked respondents whether they are willing to make available their papers via oa journal. 30% and 40% of respondents strongly agreed and agreed to publish via oa journal ,respectively . High portion of Science (35.7%) has been compared to 32.1%, 17.9% and 14.3% of Humanities, Agriculture & Veterinary and Engineering , respectively showed that Sciences are more likely than other disciplines to publish via oa journal.

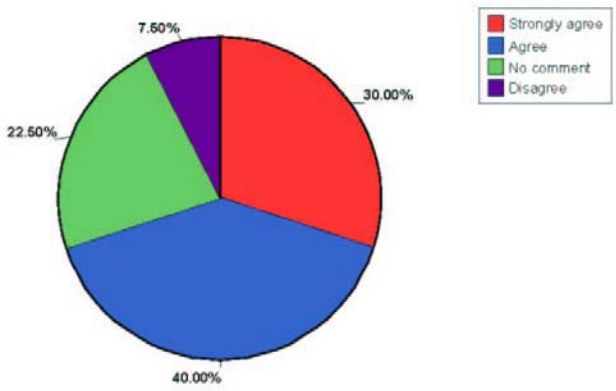


Figure 7:Academics' views on publishing via oa journal (Q22)

- **Academic views on institutional repository**

55% of respondents said that they would deposit their scholarly articles in an institutional repository, 42% had no comment. High portion of Science (38.1%) showed that academics of this discipline are more likely than Humanities, Agriculture & Veterinary and Engineering (33.3%, 13.4% and 13.4%, respectively) to deposit their scholarly materials in an institutional repository.

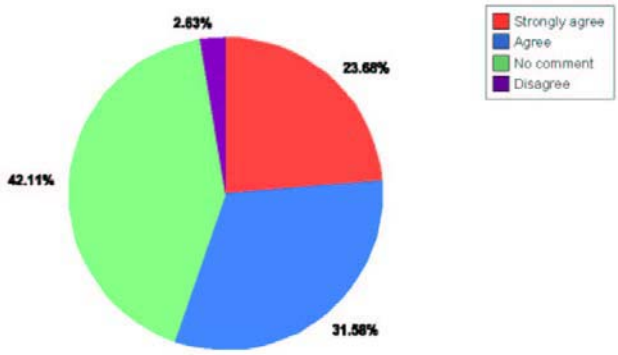


Figure 8 : Academics' views on publishing via institutional repository (Q 24)

- **Academic views on subject- based repository**

60% of respondents are "strongly agree" and "agree" on depositing scholarly materials in subject-based repository. 35% had no comment. Sciences are more likely than other disciplines to deposit their scientific work in subject- based repository: 37.5% of Science has been compared to 29.2%, 20.8% and 12.5% of Humanities, Engineering and Agriculture & Veterinary, respectively.

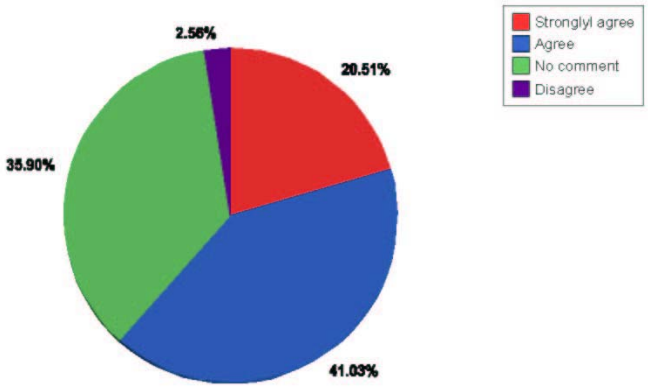


Figure 9 : Academics' views on publishing via subject repository (Q 26)

- **Familiarity with self- archiving**

45% of respondents have "very low" and "low" familiarity with self- archiving. In addition, 30% have "no familiarity" with self- archiving. Only 25% had either "very high" or "high" knowledge about it.

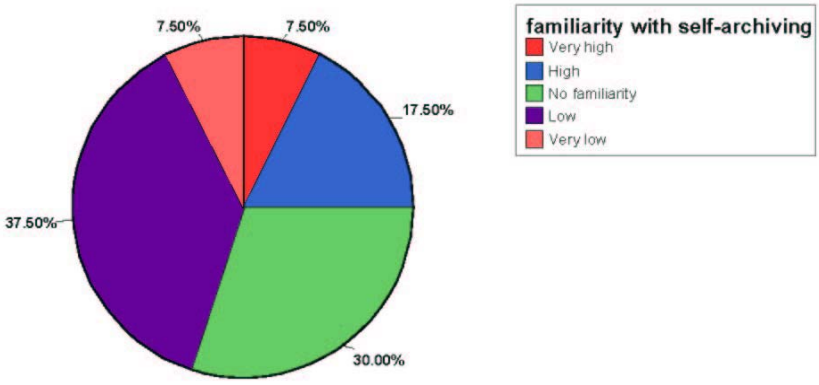


Figure 10 :Awareness of OA publishing models(Q 19)

- **Familiarity with oa journal**

39% of respondents had "very high" and "high" awareness of oa journal, but 29.7% had "no familiarity" and 26.83% had just "low" familiarity with oa journal.

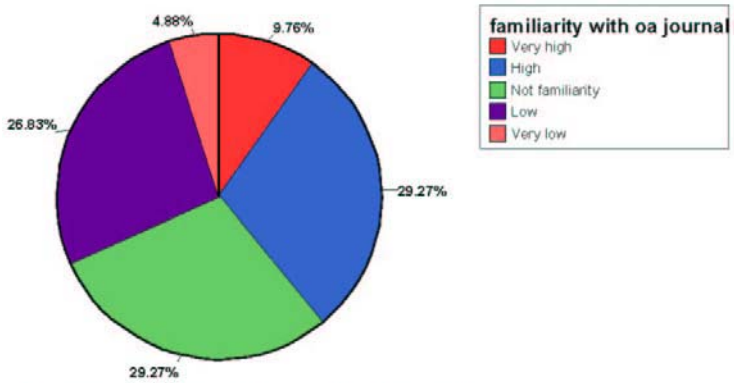


Figure 11 : Awareness of OA publishing models (Q 21)

- **Familiarity with institutional repository**

Respondents awareness of institutional repository is 26% but about 74% of them were "not familiar" with or had the least amount of knowledge about institutional repository.

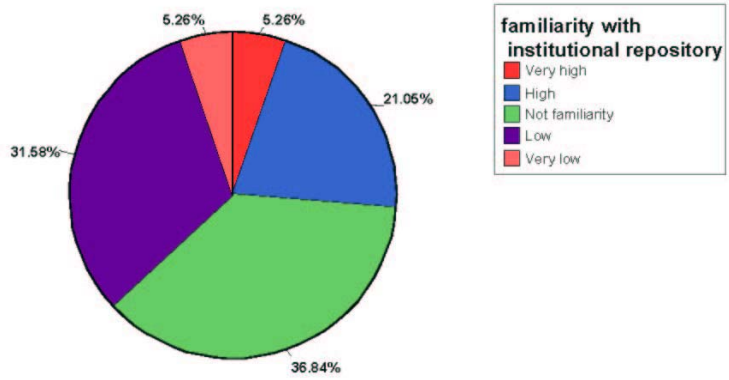


Figure 12 : Awareness of OA publishing model (Q 23)

- **Familiarity with subject- based repository**

33% of respondents had either "very high" or "high" awareness of subject-based repository. But 66% were either "not familiar" with or had low awareness of subject-based repository.

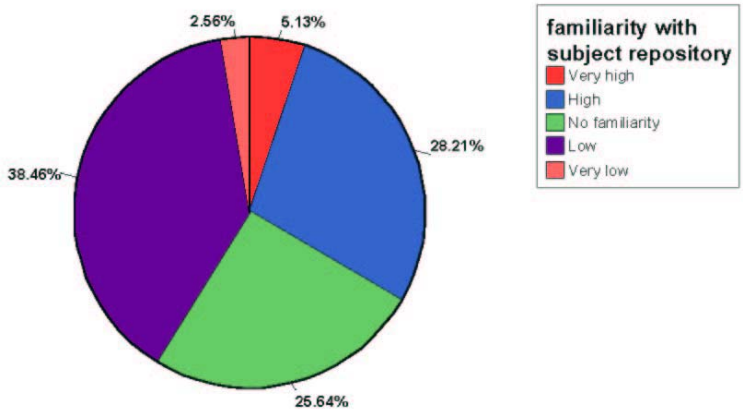


Figure 13:Awareness of OA publishing (Q 25)

- **Copyright law**

Respondents were asked to indicate their views of copyright protection of scholarly papers, 51.28% of them were not familiar with current copyright system. Current copyright system provides little

(28.21%) and very little (7.69%) expectations of respondents concerning copyright protection of scholarly papers. Only 12.82% of respondents were satisfied with current copyright system.

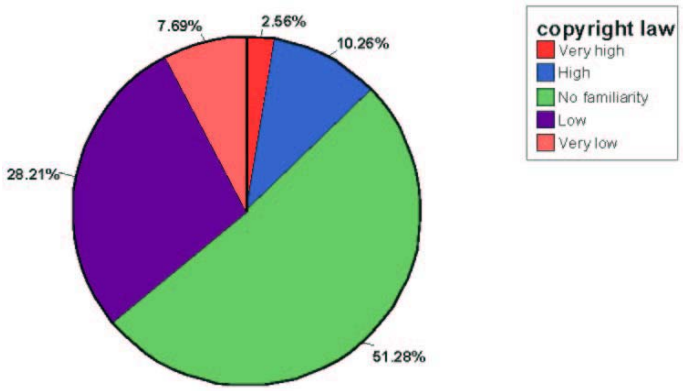


Figure 14 : Academics' awareness of copyright law (Q 40)

- **Views on copyright**

Respondents were asked to give their understandings of copyright law. Q41a: copyright law should protect large and open knowledge distribution. 54.06% of respondents were "strongly agree" and "agree" with this concept, 29.73% had "no comment" and only 16.21% were "disagree" and "strongly disagree". Q41b: copyright law means to restrict access to scholarly materials. Respondents were asked to indicate their attitudes, 44.12% of them disagreed, 41.18% had no comment, only 14.70% strongly agreed and agreed.

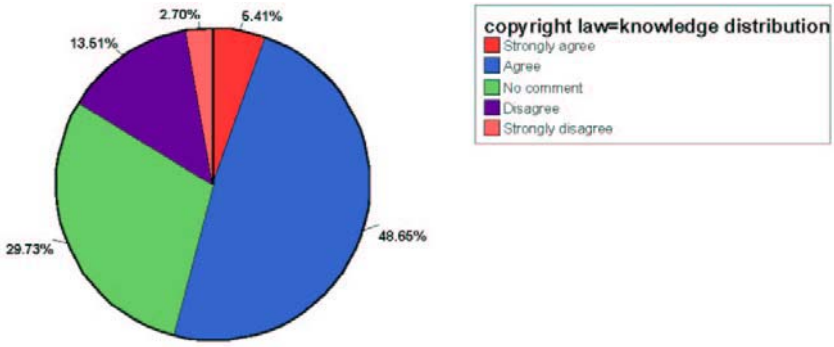


Figure 15 : Academics' attitude toward copyright law (Q 41a)

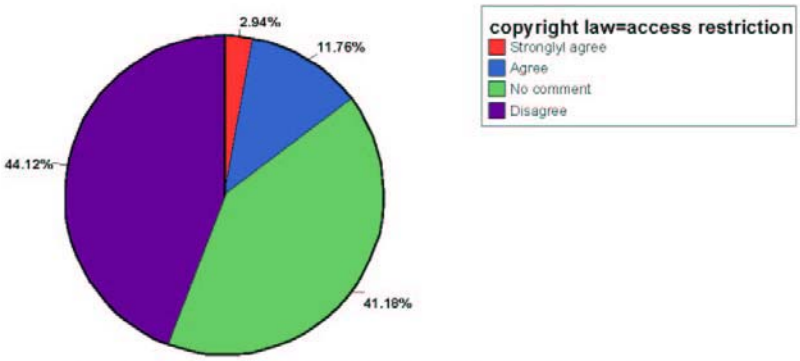


Figure 16 : Academics' attitude toward copyright law (Q 41b)

- **Copyright assignment to the publisher**

Q46 asked respondents to indicate their attitudes toward copyright assignment to the publisher. 53.84% of them were "strongly agree" and "agree" that copyright assignment to the publisher restricts access to scholarly articles, 33.33% had no comment and only 12.82% disagreed.

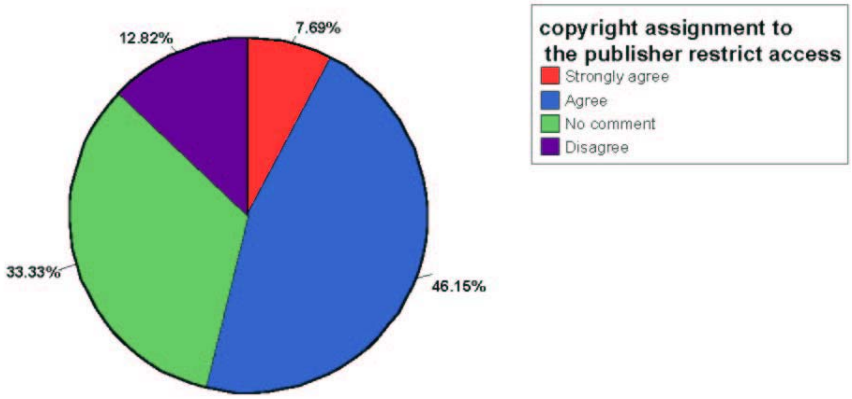


Figure 17: Academics' attitude toward copyright law (Q 46)

- **Open access materials usage comparison: academics-as-users and authors**

Q33 and Q36 compare respondents' views on how they expected to use others' scientific materials (i.e., academic-as-users) with how academics expected others to use their works (i.e., academics-as-authors). This was performed by chi-square test (Fisher exact test) on the data collected from each group at the 0.05 level. The below Table shows the breakdown.

Table1: open access materials usage comparison

Open access material usage	As- authors		As-users		
View on screen freely	Agree %	Disagree %	α=0.05 P=0.059	Agree %	Disagree %
	86.49	5.4		**92.5**	5
View on screen conditionally	34.29	**42.86**	α=0.05 P=0.001	33.33	**47.22**
Print freely	**83.33**	13.89	α=0.05 P=0.003	**89.47**	2.63
Print conditionally	39.39	**42.42**	α=0.05 P=0.002	39.39	**45.45**
Anybody-Any purpose	40	**42.5**	α=0.05 P=0.000	**43.59**	41.02
Educational & Research purpose	**51.38**	35.9	α=0.05 P=0.000	**55.26**	36.84

Based on the Fisher's Exact Test to investigate the independence of data (in the nominal level α=0.05) there is a very strong relationship between academics' attitudes on the two sides of the chain with respect to the above six options. The results show that 92.5 percent and 86.49 percent of academics-as-users and authors are fond of free availability of open access articles on-screen, respectively. Slightly more academics-as-users expected to perform this (6%) than academics-as-authors. The second option shows their views on reading articles on-screen with limits. 47.22% and 42.86% of respondents as users and authors disagreed respectively, but 33.33% and 34.29% of academics-as-users and authors agreed with limits, respectively. According to the third option in the table, respondents were asked to indicate their attitudes toward their articles allowed to be printed freely, 89.47% of academics (as users) and 83.33% (as authors) agreed. Academics-as-users and authors (45.45% and 42.42%, respectively) disagreed for implementing limitations to copies that may be printed out. The fifth option shows their views on using freely open access materials for any purposes by anybody. 43.59% of academics-as-users against 40% of academics-as-authors agreed with no restriction. But 42.5% and 41.02% of respondents as authors and users disagreed, respectively. The last option indicates respondents' views on usage of open access materials for research and educational use only. 55.26% and 51.38% of them as users and authors agreed and 36.84% as users, 35.9% as authors disagreed, respectively.

- **Pricing and permission crisis**

Suber (2003) in his study showed that there are two barriers that limit access to research findings:

1- **serial pricing crisis**, 2- **permission crisis**. The latter arises from copyright law and the former arises from journal subscription fees. Accordingly, respondents were asked to indicate their views on pay-per-view(PPV), site-license(SL), journal subscriptions and journal spiral prices. The majority of them believed that pricing and permission crisis restrict access to their articles. Table 2 shows the breakdown.

Table 2: percentage of respondents' views on journal pricing and permission crisis

	PPV %	**SL %**	**Subscription fees %**	**Journal price increase %**
Strongly agree	13.16	7.89	10.81	25
Agree	**55.26**	**55.26**	**54.03**	**60**
Disagree	10.53	13.16	10.81	5

- **Reasons for not publishing in open access models**

Q49 presented a list of possible reasons for academics not publishing in open access models. The results indicated that the most common reason for not publishing in open access vehicles is quality of works- **I believe that OA materials have poor peer review procedures**. The second reason concerning their familiarity with OA articles (36.1 %). Institutional promotion is in the third place that academics are concerning about. The fourth rank belongs to the factor that was: I perceive OA will reduce my scientific position. Less citations and small readership ranked 5 and 6, respectively. The breakdown is presented in Table 3.

Table 3: Academics' reasons for not publishing in open access models

Reasons	**Strongly agree%**	**Agree %**	**Strongly agree plus Agree%**	**No comment%**	**Disagree %**	**Strongly disagree %**	**Strongly disagree plus disagree%**
I am not familiar with OA scientific articles	13.9	22.2	**36.1**	30.6	27.8	5.6	33.4
I perceive OA will reduce my career advancement	-	23.1	23.1	30.8	38.5	7.7	**46.2**
I perceive OA will affect my academic promotion badly	5.3	26.3	31.6	28.9	31.6	7.9	**39.5**
I perceive the readership of OA to be smaller than traditional jouranls	-	5.1	5.1	25.6	59	10.3	**69.3**
I think OA articles may be less frequently cited	7.7	12.8	20.5	28.2	43.6	7.7	**51.3**
I believe that OA materials will have poor peer review procedures in place	15.8	23.7	**39.5**	23.7	28.9	7.9	36.8

DISCUSSION

The overall purpose of this study was to ascertain the views of academics of Iranian universities toward open access movement and finally provide an appropriate pattern for scholarly communication in

Iran. To do this a preliminary study was conducted on academics of Shiraz University which is one of the most famous universities in Iran.

- **Open access movement**

At the beginning the discussion is concentrated on academics' views and awareness on open access and its four channels: i.e., **self-archiving, institutional repository, subject-based repository and open access journal (**Bjork, 2004). 92% of respondents had positive view on open access publishing and its promotion. This brought about by their belief in the principle of open access (Swan & Brown, 2004). The subject discipline chosen for this study compiled of Engineering, Science, Humanities and Agriculture & Veterinary (A&V). The results presented that Sciences had the higher portion to open access publishing (33.3%), the Humanities has the second place (28.2%) , Engineering is in the third place (25.6%) and A&V the last place (12.8%).

Table 4: Academic views on four channels of open access publishing

Four channels / Subject disciplines	Self-archiving 62.5%		Institutional repository 55%		Subject-based repository 60%		Open access journal 70%	
	Agree %	Rank	Agree %	rank	Agree %	Rank	Agree %	Rank
Humanities	**40**	**1**	33.3	2	29.2	2	32.1	2
Science	20	3	**38.1**	**1**	**37.5**	**1**	**35.7**	**1**
Engineering	24	2	13.4	3	20.8	3	14.3	4
Agriculture & Veterinary	16	4	13.4	3	12.5	4	17.9	3

According to table 4 open access journal is a vehicle for publishing that had the most acceptance (70%) by the academics. The reason behind this is a cultural issue because it is more than 300 years that print-based journals have been the only vehicle for publishing and metrics for evaluating scholarly materials. In addition the results showed that academics are more familiar with open access journal than the three other channels. Self-archiving had the second rank (62.5%) and subject-based repository and institutional repository gained the 3rd and 4th ranks, respectively. Among the disciplines, academics of Humanities are more willing to publish via self-archiving (40%), but those of Sciences are more willing to publish via institutional repository (38.1%), subject-based repository (37.5%), and open access journal (35.7%).Respondents' familiarity with these four channels was not satisfactory. Table 5 shows the breakdown.

Table5: Academics familiarity with four open access channels

Four channels	Very high %	High %	NO familiarity %	Low %	Very low %
Self-archiving	7.4	**17.6**	30	**37.5**	7.5
Open access journal	9.75	**29.27**	29.27	**26.83**	4.88
Institutional repository	5.26	**21.06**	36.84	**31.58**	5.26
Subject-based repository	5.13	**28.21**	26.64	**38.46**	2.56

The results showed that there is a lack of awareness of open access movement among academics. In spite of this, the majority (92%) of them are willing to choose such vehicles for publishing their works. These findings point to an urgent need to raise academics' awareness of four open access channels.

- **What do academics think about copyright?**

The results indicated that 51.28% of respondents are not familiar with copyright law and 30% of them were not satisfied with current copyright regime. Academics' views on copyright showed their opinions were that copyright law should protect "large and open knowledge distribution" (54%) and act against access restriction to scholarly materials (44%). Finally they believed that copyright assignment to the publisher restricts access to research findings (54%). Their attitudes revealed that in spite of their ignorance (half of them) of copyright law they believed that current protection offered by copyright law exceeds the protection required by most academics (the same result showed in RoMEO studies; 2).

- **Open access materials usage comparison**

The majority of academics-as-users (92.5%, 89.47%) and authors (86.49%, 83.33%) expected to display and print open-access material freely. It is what has been shown in RoMEO studies (RoMEO studies 3). Table 3 showed that near half of academics-as-users and authors disagreed to display and print open-access works under limits. Two groups showed that they are slightly liberal about free display (3.16%, 3.03%) than free print. This indicates that academics are still a little bit concern about printing a work. In general academics-as-users are more liberal than academics-as authors. A chi-square test was performed to test the null hypothesis that whether there is any difference between academics' views in two groups on the six options in Table 3. The results showed that there was no significant difference between the attitudes of the two groups.

- **Serial pricing crisis and permission crisis**

Association of Research Libraries has indicated that journal pricing crisis put pressure on library budgets that led to more title cancellations and budget shifting from monographs to serial acquisition. Consequently, more title cancellations, less access to journal articles (serial pricing crisis). The survey examined academics views on serial pricing and permission crisis which is the result of legal and technological barriers (arise from copyright law and licensing agreements and digital rights management),(Suber, 2003). The results showed that majority of academics considered pay-per-view (68.42%), Site-license (63.15%), Subscription fees (64.84%) and journal price increase (85%) as access barriers to scientific works. Accordingly a new pattern of publishing except printed-base journal is needed.

- **Academics' reasons for not publishing in open access models**

Respondents' reasons for not publishing in open access channels presented in table 3 showed that quality control of open access materials was very important for them (39.5%). The second reason was that they were not familiar with open access scientific articles. Accordingly, academics' urgent awareness on open access is badly needed. Academics believed that the readership of open access articles will not be smaller than printed-based journal and the same as frequency of citations. Their first reason for not publishing in open access models is because of poor peer review procedures of open access materials. This view rooted in the fact that it is more than 300 years that printed-based journal is used so dominantly in measuring academics careers. So, there is a misconception that the only criterion is traditional journal and open access materials have lower standards of peer review. They should understand open access does not mean that peer review is bypassed.

CONCLUSION

The results of this study proved previous findings concerning open access movement. Academics of Shiraz University (Iran) in spite of their low awareness of open access movement significantly agreed to use open access vehicles for their publications. Their first choice in new pattern of publishing is **open access journal** and the second one is **self-archiving. Subject-based repository** and **institutional repository** are at the third and fourth place, respectively. It seems that universities should think more about the last two choices. The lack of awareness of copyright law is an important issue that should be considered in the near future by academics and their institutions. What is clear is that academics current view on open access is strongly rooted in the world of print, and we still have quite a long road ahead.

REFERENCES

- Association of Research Libraries (2003), Monographs and serial costs in ARL libraries, 1986-2003 Home Page http://www.arl.org/stats/arlstat/graphs /2003/ monser03.pdf
- Bjork, B.-C. (2004), Open access to scientific publications-an analysis of the barriers to change?-in: Information Research, 9(2) Jan 2004.-ISSN 1368-1613. Home Page Http://informatior.net/ir/9-2/paper 107.html
- Gadd , E., Oppenheim, Ch. , Probets, S. (2003), RoMEO studies 3: How academics expect to use open-access research papers.-in: Journal of Librarianship and Information science, 35(3) Sep 2003,pp. 171-187. – ISSN 0961-0006.Home Page http://www.lboro.ac.uk/departments/1s/disresearch/romeo/RoMEO%20studies% 203. pdf.
- Gadd, E., Oppenheim, Ch., Probets, S. (2003), RoMEO studies 2: How academics want to protect their open-access research papers .- in: Jornal of Information science,29(5) Sep 2003,pp.333-356.-ISSN 0165-5515.Home Page http://www. lboro. ac. uk/ departments/1s/disresearch/ romeo/.
- Gadd, E., Oppenhim, Ch., Probets, S. (2003), studies 1: The Impact of copyright ownership on academic author self-archiving.- In: Journal of Documentation, 59 (3) pp. 243-277.-ISSN:0022-0418. Home Page http://www.lboro.ac.uk/ departments/1s/disresearch/ romeo/ RoMEO %20studies %201. pdf. RoMEO
- Rowlands, I., Nicholas,D., Huntingdon, P. (2004), Scholarly communication in the digital environment: what do authors want? Findings of an international survey and author opinion: project report. ciber, city university, London,UK. Home Page http :// ciber. Soi. City. Ac. Uk/ ciber-pa-report. pdf .
- Suber, P. (2003), Removing the barriers to research: an introduction to open access for librarian.-in: College and Research Libraries News, 64(2) Feb 2003, pp.92 – 94, 113. Home Page http://www.Earlham.Edu/~peters/ writing /acrl.Htm
- Swan, A. P., Brown, Sh. N.(2004), JISC/OSI journal authors survey report. Key Perspectives Ltd. Truro, uk. Home Page http://www.jics.ac.uk/uploaded_ documents/JISC OA report 1. pdf.
- The Budapest Open Access Initiative (2002), Home Page www.soros.org/ openaccess/read.shtml

AUTHOR INFORMATION

Mohammad Reza Ghane (B.A., M.A., Ph.D.) is faculty member of Regional Library of Science and Technology (RLST) which is located in Shiraz, Iran .He was assistant for administrative and financial affairs from 1993 to 1998 and has been assistant for research affairs since 1999.He has taught courses at graduate and postgraduate levels He graduated with B.A. in English Language and Literature from Shiraz University, Iran in 1985. He completed an M.A. in Library and Information Science at Shiraz University in 1991. In 2001 he was awarded a scholarship to carry out a PhD in Library and Information Science, University of Tehran, Iran. He has great interest in Open Access Movement and published an article and had presentation on this area. His PhD thesis is carrying out on "Procedures for Providing an Appropriate Pattern for Iranian Universities Academics' Open Access to Scientific Information".
Email: mghane@ut.ac.ir

Patterns of Research Output Produced by Scholarly Communities in Korea

Hyekyong Hwang, Heeyoon Choi and Tae-Sul Seo
Korean Institute of Scientific and Technical Information, KISTI, South Korea
Soo-Sang Lee
Pusan National University, South Korea

ABSTRACT
Open Access is a new trend in the scholarly communication. The purposes of this study are to identify this new trend, to aid in the development of an open-access-based knowledge and information flow model organized by subject areas, and to further seek ways facilitate the communication process in scholarly communities in Korea. To this end, types of research output that are produced as a result of R&D activities are reported here. As well, research patterns and levels of understanding of information sharing are investigated. Data are collected through surveys. The survey population consists of registered KISTI customers. The major areas of analysis include the types of research output, publishing objectives, awareness regarding copyrights, information sharing, preservation of information, and the respondents' opinions of and experiences with trusted-digital archives. In summary, our study results show the following:

1) The most commonly reported research output is a technical report (65.2%), followed by presentation materials (63.2%).
2) The main objective of publishing research output is to satisfy the requirement of an author's affiliation (65.2%).
3) The awareness of copyrights regarding research output should be managed first by themselves (68.0%), and secondly by the organization where they are employed (24%).
4) Having Open Access while producing research output would be helpful and beneficial for improving the research process (75.2%).
5) The preferred method for sharing among Korea researchers is off-line (29.7%).
6) A majority of researchers in Korea prefer personnel computers as a preservation tool (48.3%).
7) In the opinion of TDR, the majority of researchers want an IR (Institutional Repository) primarily operated and managed by their organization (39.6%), or managed by a governmental organization (34%).

Keywords: Open Access, Scholarly Communication, Research Activities, Publication, Trusted Archive

I. Introduction

Background

Internet has provided people with a lot of benefits. New forms of digital information such as primary research data have become available for researchers with access to a personal computer. This aids them in doing their own data interpretation, and reduces their reliance on the published interpretations of others. As a result, research behavior is changing. Researchers are demanding immediate, direct access to information and increasingly expect to be able to access that information in its raw as well as its interpreted form (Brindley, 2005).

Meanwhile, with the ever-increasing journal subscription charges, many academic libraries cannot afford to buy all of the titles appropriate to the needs of the research staff in their organizations. As a result, users cannot often access papers that are pertinent to their research topics. Over the past two decades, the so-called 'serial crisis' has become more and more acute. Journal prices have risen faster than both the rate of inflation and the rate of increases in library budgets, exacerbating the problem (Oppenheim, 2005).

The concept 'Open Access' emerged in 1990's in an effort to overcome this situation. Open access involves the use of scholarly materials without a subscription charge. There are typically two kinds of open access movements: open access journals and electronic repositories (Falk 2004). In 1992, just five journals offered open access to the material they published. Today, that number has grown to approximately 1,900 journals. As well, the number of the repositories has increased to about 350.

The Situation in Korea

The situation regarding the Korean scholarly journal market is somewhat different from that of more developed countries. Although there are many scholarly journals published in Korea, researchers whose work is related to science and technology subscribe to more foreign journals than they do domestic ones. However, the need for an electronic repository may not differ from that of foreign countries. Accordingly,

it will be meaningful to investigate the attitudes of Korean scholars toward a repository without restrictions on access.

Problem definition

The trend toward open access has different characteristics according to the different academic fields or specific local fields. Some authors have emphasized the importance of considering these differences among fields when developing a scholarly communication model (Kling 2000; Joung 2004).

Purpose and method

The purpose of this paper is to identify this new trend, to aid in the development of an open-access-based knowledge and information flow model, and to further seek ways to facilitate the communication process within and among scholarly communities in Korea. To this end, types of research output that are produced as a result of R&D activities are reported here. In addition, research patterns and levels of understanding of information sharing are investigated. The data are collected through surveys. The survey population consists of registered KISTI customers, and the major areas of analysis include the types of research output, publishing objectives, awareness of copyrights, information sharing, preservation of information, and the respondents' opinions of and experiences with trusted-digital archives.

II. Related Works

Overseas

A project sponsored by JISC (Joint Information Systems Committee) is ongoing at Loughborough University in the UK. Termed the RoMEO (Right of Metadata for Open Access) project, it is focused on a copyright problem that could be raised when researchers deposit their research output into their institutional repositories. The respondents in this project were authors, publishers, data providers, and service providers around the world. 60% of the respondents agreed with the concept of free use of their products if protection of the moral rights of authors, and a non-commercial purpose by users can be guaranteed (Oppenheim 2003).

A research team at the University of Rochester in the USA undertook the IMLS (Institute of Museum and Library Services) Dspace project to investigate the research patterns of professors so as to observe how the institutional repositories should best be used by these and other professors. The project team concluded that how the professors perceive their work practices should be the most important consideration before the development of an institutional repository (University of Rochester 2004) is undertaken.

Key Perspectives Ltd. conducted a study of authors who had published their work in open access journals, and compared and contrasted those authors to authors who had not done this. The report showed that many authors want both self-archives as well as open access journals to be available (Swan & Brown 2004).

The University of Southampton in the UK maintains a sustainable multidisciplinary e-Prints archive, called TARDis (Targeting Academic Research for Dissemination and Disclosure). They performed an environmental assessment of research publication activity and related factors impacting the development of an institutional e-Print repository at the University of Southampton (Hey 2004).

Domestic

Joung (2003) performed a survey on copyright awareness and the publication motives of agricultural field researchers to explore the feasibility of an overseas information-sharing archive model. Based on the survey results, she proposed a scholarly communication model adopting an information sharing license that permits free use for academic and/or non-profit use.

III. Definition of Research Output

Types of research outputs

Scientific research is performed mainly in universities and research institutes by students, professors and researchers. In this paper, research output is defined as all types of information resources produced from these scientific research activities. These are divided into four phases: learning, proposing, performing and publishing. The research output produced in each of the phases is shown in Table 1. The learning and proposing phases do not include any research activities, but the output from those phases are used, and considered as very important materials in performing research.

The output of the performing phase comprises presentation materials, technical memos, drawings of apparatuses, data sets, meeting records, questionnaires, data sheets, and travel reports. These types of output are almost always included in the grey literature.
Researchers produce theses, papers, posters, articles, books, technical reports, patent specifications, as well as standard specifications, guide books of products and technologies, and program source codes in the publishing phase along with some forms of these output types in the performing phase.

Table 1 Research outputs by research phases

Research phases	**Learning**	**Proposing**	**Performing**	**Publishing**
Research output examples	Syllabus, text books, reference materials, assignments, term papers	RFP, Proposals, Presentation materials	Presentation materials, Technical memos, drawings of apparatus, data set, data sheets, meeting records, questionnaires, travel reports	Theses, papers, posters, presentation materials, articles, books, technical reports, patents specifications, standard specifications, guide books of products/technologies, and program source codes

IV. Survey Results and Data Analysis

Overview of survey

To collect data, a questionnaire was constructed, which included questions related to the types of research output, publishing objectives, awareness regarding copyrights, information sharing, preservation of information, as well as the respondents' opinions of and experiences with trusted-digital archives. It consisted of 14 items, grouped into 6 sectors, in addition to 8 separate population statistical items. 4,792 customers were selected from among the total pool of registered KISTI customers as a candidate survey population. They were all believed to be active researchers. The questionnaire was sent to them primarily using e-mail; some were sent by regular mail. The total number of answered questionnaires after one month was 250, representing a 5% response rate among the recipients. The data was analyzed using SPSS, a software program specializing in statistical analysis. Even though a comparably small number of answer sheets were collected, some meaningful results were obtained.

Basic results

Table 2 shows the composition of the respondents with respect to job, and major of their final degree. Chiefly, 53.6% of the respondents were academic researchers who work at universities or in governmental institutes. The incidence of researchers working in industry was 34.8%. For the major of the final degree of these researchers, engineering was most commonly reported among the respondents. The rate for this was approximately 60.0%. The second most common reply to this question was science, at 23.6%.

Table 2 Composition of the respondents

Classification	**Types**	**Occurrences**	**Rates (%)**
Job	Academics	159	63.6
	Industry	87	34.8
	Others	4	1.6
Major of final degree	Engineering	148	59.2
	Science	59	23.6
	Medicine	9	3.6
	Humanities/Social	23	9.2
	Others	11	4.4

Number of production by types of research output

Table 3 shows the numbers of respondents who answered 'yes' to the question asking whether they produce at least 1 output in recent three years. If a yes answer was received, these were then organized by type of output. The most common output type was a technical report, and the next most commonly reported was presentation material. A peer-reviewed paper, a conference paper, a proposal, a patent, and a conference poster follow successively. This means that many Korean researchers are involved in projects sponsored by the government or other funding organizations requiring technical reports and presentations.

Table 3 Number of Research Output by Type

Research Phases	Types of Research Outputs	Producing more than 1 in 3 yrs.	Rates (%)
Publishing	Technical reports	163	65.2
Publishing/Performing/Proposing	Presentation materials	158	63.2
Publishing	Peer reviewed papers	141	56.4
Publishing	Conference papers	129	51.6
Proposing	Proposals	125	50.0
Publishing	Patents	108	43.2
Publishing	Conference posters	104	41.6
Performing	Experiment Material	100	40
Performing	Questionnaire	67	26.8
Publishing	Book	66	26.4
Publishing	Thesis	60	24.0
Performing	Drawings of apparatus	59	23.6
Performing	Statistics Data	54	21.6
Performing	Experiment Equipment and System Explanation	50	20
Performing	Data sheets	45	18.0
Publishing	Program source codes	45	18.0
Performing	Software	36	14.4
Performing	Multimedia	36	14.4
Performing	Data set	34	13.6

Comparison of output patterns of engineering and science fields

Table 4 represents a comparison of output patterns for two research fields: engineering (mechanical and metal) versus science (physics and chemistry). This set of data shows that researchers in engineering fields produce more than those in science fields in the case of peer-reviewed papers, conference papers, patents, technical reports, and proposals; while researchers in science-related fields are more productive in terms of conference posters, presentation materials, and experimental materials. It was unexpected that the total rate of peer-reviewed papers and conference papers from those in engineering fields was greater than the rate from those in science-related fields. In terms of more than 10 articles produced, science researchers were dominant with respect to the number of output per researcher.

Table 4 Comparison of Output Patterns of Research Fields

Type of outputs	fields	Producing in 3 yrs.					Rate (%)
		1-3	4-6	7-9	10>	Total/no. respondents	
Peer reviewed papers	Eng.	16	6	4	3	29/42	69.00
	Sci.	1	5	5	6	17/29	58.60
Conference papers	Eng.	11	8	2	4	25/42	59.50
	Sci.	3	3	6	3	15/29	51.70
Conference posters	Eng.	8	3	1	2	14/42	33.70
	Sci.	1	7	2	4	14/29	48.30
Patents/Utility Model	Eng.	10	4	3	4	21/42	50.00
	Sci.	6	4	1	1	12/29	41.40
Technical reports	Eng.	15	6	3	2	26/42	61.90
	Sci.	7	3	5	2	17/29	58.60
Presentation Materials	Eng.	8	2	10	2	22/42	52.40
	Sci.	4	2	10	2	18/29	62.10
Proposals	Eng.	10	3	4	3	20/42	47.60
	Sci.	3	1	3	2	9/29	31.00
Experiment Materials	Eng.	7	3	4	4	18/42	42.90
	Sci.	3	2	7	3	15/29	51.70

Motive for Publishing

It was identified that Korean researchers have the same motivations as those in other countries where similar surveys had been performed *(Gibbon 2004).* According to the current survey, Korean researchers

publish their research output mostly in order to meet the requirements of their affiliations (65.2%), secondly to bolster their academic reputation (61.2%), and thirdly to share their work (52.0%) with others.

Copyright holder

Korean researchers think that the primary copyright holders of their research output should be themselves (68.0%) according to the survey results. According to them, the next most important holders were, successively, the organizations who employ them (24.0%), sponsors providing research funds (9.2%), and the publisher disseminating their output (2.0%). The primary copyright holder rate of the researcher added to his or her affiliation together becomes 82.0%. This strongly suggests the need for institutional repositories exist.

Table 5 Awareness of Researcher on Copyright Holder

Entity	**1st**	**2nd**	**3rd**	**4th**	**others**
Researcher	170(68.0%)	50 (20.0%)	20(8.0%)	4(1.6%)	6(2.4%)
Organization	60(24.0%)	116(46.4%)	53(21.2%)	9(3.6%)	12(1.2%)
Sponsor	23(9.2%)	67(26.8%)	126(50.4%)	9(3.6%)	25(1.0%)
Publisher	5(2.0%)	8(3.2%)	23(9.2%)	154(61.6%)	60(24.0%)

Intention to Open-Use of Research Output

88.4% of respondents are reportedly willing to provide open access to their research output. 91.0% from among these are willing to allow non-profit use. 63.3% would even allow to their outputted works to be modified. These findings are similar to the findings of the RoMEO project (Oppenheim 2003).
75.2% of Korean researchers think that sharing their research output with colleagues before publication is helpful. The method of sharing that Korean researchers prefer is off-line (29.7%). E-mail (17.2%) ranked second. Online sharing, including personal homepages (5.7%), communities on the Internet (13.5%), the affiliate's homepages (11.7%), and in-company intranets (9.9%) were suggested as possibly widely used; however, these rates represent very low levels compared to those of a similar MIT survey (*Barton & Walker 2002).*

Preserving Research Output

Korean researchers preserve their research output mainly in the hard-disc memory of their personal computers (48.3%). The next most commonly reported methods were, successively, a CD-ROM (21.2%), followed by USB memory devices (13.9%). The rate of an intranet, which is a similar concept to an institutional repository, takes a very low portion at 6.5%. This means that Korean researchers are reluctant to practice their intention of opening their research output to outside view on an intranet. We are in the early-stages of sharing research output through the institutional repository. Concerning a trusted-digital archive, it is shown that 23.2% are willing to submit all of their research output to this type of archive; 56.8% would submit all except for what they deem very important output. Taken together, this suggests that 80.0% of Korean researchers feel a need for institutional repositories for reserving at least some of their research output.
Finally, Korean researchers prefer themselves or governmental organizations (39.6% and 34.0%, respectively) as having authority over a trusted-digital archive.

Table 6 Candidate Authority of Trusted-Digital Archive

Candidate Authorities	Occurrences	Rate (%)
Organization that employs researcher	99	39.6
Governmental organization	82	34.0
Researcher	35	14.0
Community researcher participate	27	10.8
Others	2	0.8
No Response	2	0.8

V. Conclusion Remarks and Directions for Future Studies

Conclusion

The purpose of this study was to identify the intention of researchers and suggest some strategies for an open access-based scholarly information service in Korea. To this end, a survey was given to a group of researchers registered as KISTI customers. The major areas of analysis included the types of research

output, publishing objectives, awareness of copyrights, information sharing, preservation of information, and the respondents' opinions of and experiences with trusted-digital archives.
In summary, the most common output type among Korean researchers was found to be technical reports, and the next most common was found to be presentation material. Researchers in engineering fields produce more peer-reviewed papers, conference papers, patents, technical reports, and proposals than do those in science fields, while science researchers produce more conference posters, presentation materials, and experiment materials. However, in terms of both peer-reviewed and conference papers (more than 10 articles), science researchers were dominant with respect to the number of output per researcher.
Korean researchers' main reason for publishing their research output was found to be that it was a requirement of their affiliations. They think that the primary copyright holders of their research output should be themselves or the organizations that employ them.
Most Korean researchers think that sharing their research output with colleagues before publication is helpful. The preferred method of sharing for Korean researchers is off-line. Diverse online sharing tools are also used, but the rates are at very low levels compared to the rates of other countries.
Korean researchers preserve their research output mainly on the hard-disc memory in their personal computers (48.3 %). The low rate of the use of an intranet means that Korean researchers are reluctant to put into practice their intention of openly sharing their research output. Concerning trusted-digital archives, many Korean researchers want institutional repositories for reserving some of their research output. Korean researchers prefer their own, or governmental organizations over other entities as an authority of a trusted-digital archive.

Directions for Future Study

This is primarily an initial study to tap into the possibility of developing an open access archive in the Korean scholarly environment, especially in the science and engineering fields. A more specific study can follow with an aim to discover these same possibilities in more specific academic fields such as nano-technology, IT technology, bio-technology, as well as others.

References

- Swan A. and Brown S. (2005), Open Access Self-Archiving: An Author Study, Key Perspectives, UK.
- ALPSP 1999, What Authors Want: The ALPSP Research Study on the Motivations and Concerns of Contributors to Learned Journals. West Sussex, U.K.: Association of Learned and Professional Society Publishers. Quoted in: Gibbons, Susan (2004), Establishing an Institutional Repository. *Library Technology Reports*, 40(4).
- Barton, M.R. & Walker J. H.. 2002. MIT Libraries' DSpace Business Plan Project: Final Report to the Andrew W. Mellon Foundation. [cited 2005. 5.10].http://libraries.mit.edu/dspace-mit/mit/mellon.pdf.
- Oppenheim C. (2005), "Open Access and the UK Science and Technology Select Committee Report Free for All?", *Journal of Librarianship and Information Science*, Vol. 37 No. 1, pp. 3-6.
- Gibbons, S. (2004), Establishing an Institutional Repository. *Library Technology Reports*, 40(4).
- Hey, Jessie M.N. (2004), "An environmental assessment of research publication activity and related factors impacting the development of an Institutional e-Print Repository at the University of Southampton". [cited: 2004.11.1]. <http://eprints.soton.ac.uk/archive/00006218/>.
- F. Howard (2004), "Open Access Gains Momentum", *The Electronic Library*, Vol. 22 No. 6, pp. 527-530.
- Joung K.H. (2002), A Study on the Open Access Model for Scholarly Communication, *Journal of the Korean Society for Information Management*. Vol. 19 No. 4, pp. 383-399.
- Brindley L. (2005), "The British Library: its origins, development and future", *Interlending & Document Supply*, Vol. 33 No. 2, pp. 76-80.
- Oppenheim, C. (2003), Project RoMEO Final Report. [cited 2004. 10.2]. <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/RoMEO-Final-Report.doc>.
- University of Rochester IMLS Grant (2004), [cited 2004.10.20]. <http://docushare.lib.rochester.edu/docushare/dsweb/Get/Document-13766/IMLSGrantFAQs.html>.

Towards a Continuum of Scholarship: The Eventual Collapse of the Distinction Between Grey and non-Grey Literature?

Marcus A. Banks
New York University School of Medicine, NYU, USA

Introduction

The open access publishing movement has begun to enhance the availability of research published in peer-reviewed journals. Valuable grey literature also deserves extensive exposure, but in too many cases it languishes unnoticed. A common means of increasing access to grey literature is the development of discipline-specific portals to this literature, most of which remain fairly obscure. A more comprehensive solution would be to integrate access to grey literature within the databases that scholars regularly consult.

These databases already identify peer-reviewed articles. The open access movement is a political endeavor to enhance retrieval of those materials that are easy to identify but not to obtain. With a focus on the grey literature of public health, this paper will suggest political steps that could enhance access to materials that would often be easy to obtain if they could be identified.

Shifting from a political to technological analysis, I will propose that institutional repositories represent an exciting possibility for both the preservation and retrieval of grey literature. The development and curation of robust institutional repositories should obviate the political struggle necessary to challenge the current hierarchy between non-grey and grey literature.

I. The Waning Primacy of Peer Review

Publishing a peer reviewed article in a prestigious journal remains the highest validation for a work of scholarship. Peer review has served scholars well for centuries. The concept of reporting and validating research findings began in 1665, with the founding of the *Philosophical Transactions of the Royal Society of London* [1]. Peer review has enhanced the rigor and relevance of many scientific breakthroughs.

In spite of its many benefits, peer review is flawed. Scholars tend to publish only statistically significant findings rather than all of their findings, which inflates the perceived importance of reported results [2]. More nefariously, business and political interests have become adept at using "research" teams to lend a patina of scientific rigor to their product development plans or policy goals [3]. These practices pervert the principle of objective scrutiny, which should be a foundation of scholarly research.

Personal computers did not exist in 1665, and for more than 300 years afterwards. The only way to disseminate research results was to print them, and this acted as a natural constraint on what scholars could produce. Within the confines of this system, the peer-reviewed article emerged as the vector of choice for scholarship.

Personal computers are ubiquitous at academic institutions in the developed world today, and this has begun to alter conceptions of what constitutes scholarship. Digital media has substantially displaced print media, particularly in the scientific, technical, and medical (STM) domains [4]. There is no longer a pragmatic reason that the peer-reviewed journal article must stand atop the hierarchy of evidence. It is now possible to envision a seamless continuum of evidence, which is housed in well-organized and maintained repositories [5]. This continuum would still contain formal articles, which would ratify rather than validate scholarly work. Content at other points along the continuum of evidence could come to be as highly regarded as journal articles. Much of this content—such as technical reports and public policy overviews—would be what we now consider grey literature.

In this long-term vision, grey literature will become indistinguishable from non-grey literature. For the foreseeable future, this distinction will remain. Raising awareness of the value of grey literature is a political endeavor, and should be understood as such.

The open access movement, which primarily seeks to increase access to peer-reviewed articles, is explicitly political. In less than a decade, open access has blossomed from a fringe idea into a perspective that merits serious consideration [6]. Although the future of open access is unclear, it seems inevitable that the scholarly communication landscape will contain some version of open access. Advocates for grey literature can learn valuable lessons from the political history of the open access movement.

II. Political History of the Open Access Movement

One of the most comprehensive timelines of the open access movement places its beginning in 1966, when the United States Department of Education made the Education Resources Information Center (ERIC) freely available. Open access took on political overtones in 1989, when Dr. Eddy van der Maarel resigned as Editor of the journal *Vegatatio* to protest the policies of the journal's publisher, and convinced most of the editorial board to resign as well [7]. In a statement, der Maarel explained that the intellectual labor behind the journal was free or poorly compensated, even though the journal was one of the most profitable for the Wolters Kluwer publishing company [8]. The idea that scholars should control the publishing process, rather than relying upon publishing companies, is an underpinning of the open access movement.

Throughout the 1990s, enterprising scholars in various disciplines began to publish journals that were available online at no charge. During the same period, established publishing companies invested heavily in online delivery systems. Journal prices rose sharply over this period, particularly for STM materials [9]. By the end of the decade, librarians were exasperated by the cost increases. In 1998, the Association of Research Libraries launched the Scholarly Publishing and Academic Resources Coalition (SPARC) as a means of providing cost-effective and high-quality alternatives to journals produced by major publishers [10].

In 1999, Dr. Harold Varmus proposed the development of e-Biomed, a freely accessible digital archive that would contain un-reviewed pre-prints and peer-reviewed final drafts of articles [11]. Varmus was the director of the US National Institutes of Health (NIH) at the time, and his proposal generated a great deal of criticism that the federal government was attempting to control scholarly publishing. Within a year the proposal had morphed into the politically palatable alternative of PubMed Central (PMC), a digital archive managed by the US National Library of Medicine [12]. At its inception PMC only concerned final drafts rather than pre-prints, and publishers could choose whether or not to participate.

Varmus's actions quickened the pace of calls for reform in the scholarly communication system. Although STM materials remained a flash point, scholars in all disciplines began to realize that the Internet provided a new means of developing and distributing scholarly materials. In 2000, the Association of American Universities and Association of Research Libraries released the "Principles for Emerging Systems of Scholarly Publishing." The principles state that the cost of published research should be contained, and that electronic capabilities have the potential to provide wide access to scholarship [13]. The term "open access" does not appear in the Principles, but the desire to exploit electronic capabilities is an antecedent of the formal open access movement.

The phrase "open access" gained currency in 2001, when the Open Society Institute convened the Budapest Open Archives Initiative. The Initiative states that, "The literature that should be freely available online is that which scholars give to the world with no expectation of payment." In this spirit, it states that "open access" provides scholarly articles on the "public internet," which people can access without any "financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself." Furthermore, people are free to re-use these materials in any way they wish, provided that they acknowledge the source and "give authors control over the integrity of their work" [14].

The Budapest Initiative directed attention to the concept of open access, but did not provide a precise definition of what this means. This was the accomplishment of the Bethesda Statement on Open Access Publishing, published in 2003. The Bethesda Statement defines an open access publication as meeting two conditions: it is freely available for re-use subject to proper acknowledgement of the source, *and* is deposited in a long-term digital archive for the purpose of preservation. PMC is such a repository for the biomedical sciences [15].

Less than a week after the Bethesda Statement appeared, US Representative Martin Sabo (D-MN) introduced the "Public Access to Science Act" in the US House of Representatives. The Act sought to exclude from copyright protection any research that was "substantially funded by the Federal Government" [16]. This proposal was more politically volatile than Varmus's e-Biomed proposal. Although it did not succeed, Sabo's bill re-opened discussion about how the US federal government should facilitate access to the results of the research it funds.

In July 2004, approximately one year after Sabo introduced the Public Access to Act, the House Appropriations Committee directed NIH to *require* deposit of articles that resulted from NIH-funded research in PMC. Grantees and publishing companies would have up to six (6) months after publication to fulfill this request. When NIH released its draft implementation policy for this plan, the "requirement" had become a "request" that grantees deposit their articles. After much consultation with various stakeholders, NIH released the official "Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research" (NIH Public Access Policy) in February 2005. The policy remains a

request rather than a mandate, and grantees and publishing companies have twelve (12) months to decide whether to heed the request. The policy went into effect in May 2005 [17].

Open access advocates have expressed dismay about the softening of the initial mandate. It is probably the best that can be expected in the United States, given a heavy presumption against governmental involvement in the publishing sector. To date the NIH has not received many submissions to PMC as a result of the policy [18]. This would almost certainly be different if a mandate were in effect. In October 2005 the Wellcome Trust, one of the United Kingdom's largest funders of medical research, began to mandate that its grantees deposit papers in an open access repository [19]. It will be important to compare the relative success of the US and UK approaches over time.

Many organizations lobbied for the enactment of the NIH Public Access Policy. Grey literature advocates would benefit from studying the political approach of one of these organizations in particular, the Alliance for Taxpayer Access (ATA) [20].

The ATA has put a human face on the consequences of restricted access to government-funded research. Its core membership includes people grappling with the conditions that NIH seeks to address. They ask a pointed question, which has galvanized the political discussion: If the government can spend my money to study my disease, shouldn't I be able to see the results?

For the most part, the NIH funds clinical research that results in peer-reviewed articles listed in publicly available databases. Although access barriers can preclude reading the full article, at least people know that the articles exist. The NIH Public Access Policy is an attempt to increase the proportion of this material that is available to everyone.

Public health relies significantly upon grey literature, which is difficult to identify because it is not consistently indexed [21]. The New York Academy of Medicine's Grey Literature Report is an attempt to address this challenge, but the challenge is vast [22]. As described at GL6 in 2004, the Center for Natural Language Processing at Syracuse University is developing tools for natural language searching of grey literature about public health interventions. The Robert Wood Johnson Foundation has funded this work, which is another attempt to reduce the difficulty of retrieving this material [23].

In stark contrast to the clinical research funded by NIH, there is no compelling beneficiary group to advocate for improved access to public health grey literature. Because public health benefits everyone, it appears to benefit no-one in particular. Although this perception would make it difficult, public health advocates could form an "Alliance for Public Health Promotion" in the spirit of the ATA. Through this Alliance the representatives of leading public health organizations would highlight the effects of a fragmented public health system on the lives and well being of everyday Americans. With the US Congress spearheading the effort, public health might become a greater priority for the federal government. If that occurred, comprehensive retrieval mechanisms for public health grey literature—integrated into databases that are already widely searched—could be developed.

Given an American political dynamic that favors limited investment in public health, this scenario is very unrealistic at the present time. However, if public health were to become a larger priority in the United States, increased access to grey literature would be one of many consequences.

The goal of improving public health is not simply to facilitate access to grey literature. But it is fair to say that the ease of access to information is one measure of how much societies value that information. One reason for the gains of the open access movement is because many people desire to obtain peer-reviewed clinical research without restrictions. Given the high status of formally peer-reviewed articles, political advocacy for the complementary importance of grey literature is essential for the near-term. Over the long-term, institutional repositories could eliminate the need for such advocacy.

III. The Potential of Institutional Repositories to Increase Access to Grey Literature

Institutional repositories are a relatively new approach to the challenge of digital preservation of an institution's intellectual output. Disciplinary archives such as arXiv, which collects pre-prints in the fields of physics, mathematics, nonlinear sciences, computer science, and quantitative biology, are an antecedent of institutional repositories [24]. In the United States, the Library of Congress is leading a federal digital preservation effort, the National Digital Information Infrastructure and Preservation Program [25]. Self-archiving, the practice of archiving scholarly material on personal web sites, is the complete opposite of the federal approach to digital preservation [26]. Although the approaches vary widely, they all point to the urgent importance of preserving digital scholarly materials.

The self-archiving movement began in 1994, with the publication of "The Subversive Proposal" by Dr. Stevan Harnad [27]. Despite much effort by Harnad, this movement has not flourished. Researchers in

many disciplines are not accustomed to taking responsibility for the preservation of their work, and have not changed their habits even after software to streamline self-archiving became widely available.

Institutional repositories shift the onus of preserving digital materials from individual faculty members to their institutions. The concept of institutional repositories became prominent in 2002, as a result of the DSpace initiative between Hewlett Packard and the Massachusetts Institute of Technology [28]. As of November 2005, over 100 institutions utilize open source DSpace software to manage their institutional repositories [29].

According to a seminal article by Clifford A. Lynch, "The development of institutional repositories [has] emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication." Lynch defines an institutional repository as a "set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members." Repositories can store a variety of content, including "*experimental and observational data captured by members of the institution that support their scholarly activities*" (emphasis mine) [30]. These types of data are a form of grey literature.

Faculty must be convinced of the value of contributing to institutional repositories, even if they do not have to manage them. Investigators at the University of Rochester have documented widespread faculty confusion regarding the benefits of institutional repositories [31]. This confusion resulted in smaller repository growth than anticipated for the first year after the launch of DSpace. However, there was enough activity in this period to hearten advocates of grey literature. 58% of the contents in institutional repositories during this first year were "other documents," including a healthy proportion of technical reports and working papers [32].

Interest in institutional repositories continues to increase globally. In mid 2005 repositories around the world typically contained digital versions of print-based materials, with a mixture of traditional scholarship and grey literature. Repositories in Australia and the United States conform to this pattern, but with a larger palette of content types than in most countries [33].

In the United States, institutions may begin to require that faculty members deposit their formal papers in institutional repositories. For other materials, including grey literature, deposit is likely to remain optional [34]. This is evidence of a continuing hierarchy between peer-reviewed and grey literature, but of a softer variety than before. Grey literature is not consciously excluded from institutional repositories, in the same way that it is omitted from many traditional databases.

Institutional repositories did not emerge only to facilitate access to grey literature. They are containers for capturing the complete digital output of an institution. But accessing grey literature is an important use of repositories, both now and in the future. Any organization could benefit from hosting an institutional repository. For example, non-profit research centers often produce reports that are difficult to retrieve after a short period of time. Institutional repositories could remedy this problem, and the software is available at no cost.

Content analyses of institutional repositories are critical to understanding their growth, particularly during the initial years of their development. It will soon become important to understand how researchers utilize the contents of repositories. Will people continue to gravitate to peer-reviewed articles? Or will they seek out novel content, including grey literature, because it is easier to obtain than ever before? My guess is that the latter possibility will prevail, which will lead over time to a much enriched understanding of the value of grey literature. The peer-reviewed article is an artifact of a print-based system, and not an inherently superior way to depict research results. Even if the current hierarchy between peer-reviewed materials and grey literature remains, it will be much more flat.

Acknowledgments
Many thanks to Kathel Dunn, New York University's Ehrman Medical Library; James G. Neal, Columbia University Libraries; and MacKenzie Smith, Massachusetts Institute of Technology Libraries, for their helpful comments about this paper.

References

1. Royal Society: Brief History of the Society. http://www.royalsoc.ac.uk/page.asp?id=2176 [Accessed 21 November 2005].
2. Conn VS, Valentine JC, Cooper HM, and Rantz MJ. (2003), Gray literature in meta-analyses. In: Nursing Research, 52 (4) July-August 2003, pp. 256–61. – ISSN 0029-6562
3. Center for Science in the Public Interest: Integrity in Science Conference. http://www.cspinet.org/integrity/conflictedscience_conf.html [Accessed 21 November 2005].
4. Wilson B. (2003), Improving access to STM literature: the need for continuing dialogue. In: D-Lib Magazine, 9 (6) June 2003. – ISSN 1082-9873. Available online at: http://www.dlib.org/dlib/june03/06editorial.html [Accessed 21 November 2005].
5. Lynch C.A. (2003), Institutional repositories: essential infrastructure for scholarship in the digital age. In: ARL Bimonthly Report, 226, February 2003, pp. 1-7. Available online at: http://www.arl.org/newsltr/226/ir.html [Accessed 21 November 2005].
6. SPARC Open Access Newsletter: Open Access in 2003. http://www.earlham.edu/~peters/fos/newsletter/01-02-04.htm#2003 [Accessed 21 November 2005].
7. Timeline of the Open Access Movement. http://www.earlham.edu/~peters/fos/timeline.htm [Accessed 21 November 2005].
8. FOS Forum: Eddy van der Maarel Statement. http://lists.topica.com/lists/fos-forum/read/message.html?mid=902159566&sort=d&start=44 [Accessed 21 November 2005].
9. The impact of serial costs on library collections. In: ARL Bimonthly Report, 218, October 2001, p. 9. Available online at: http://www.arl.org/newsltr/218/costimpact.html [Accessed 21 November 2005].
10. Scholarly Publishing & Academic Resources Coalition. http://www.arl.org/sparc/ [Accesed 21 November 2005].
11. Original Proposal for E-biomed (Draft and Addendum). http://www.nih.gov/about/director/pubmedcentral/ebiomedarch.htm [Accessed 21 November 2005].
12. Kling R, Fortuna J, King A. (2002), The Remarkable Transformation of E-Biomed into PubMed Central. Available online at: http://rkcsi.indiana.edu/archive/CSI/WP/wp01-03B.html [Accessed 21 November 2005].
13. Principles for Emerging Systems of Scholarly Publishing. http://www.arl.org/scomm/tempe.html [Accessed 21 November 2005].
14. Budapest Open Access Initiative. http://www.soros.org/openaccess/ [Accessed 21 November 2005].
15. Bethesda Statement on Open Access Publishing. http://www.earlham.edu/~peters/fos/bethesda.htm [Accessed 21 November 2005].
16. Public Access to Science Act. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_bills&docid=f:h2613ih.txt.pdf [Accessed 21 November 2005].
17. Public Access Home Page. http://publicaccess.nih.gov/ [Accessed 21 November 2005].
18. NIH Public Access: September 2005 Submission Statistics. http://nihms.nih.gov/stats/2005-09.html [Accessed 21 November 2005].
19. Open Access a Must for Wellcome Trust Researchers. http://www.wellcome.ac.uk/doc_WTX026822.html [Accessed 21 November 2005].
20. ATA: The Alliance for Taxpayer Access. http://www.taxpayeraccess.org/ [Accessed 21 November 2005].
21. Gray BH. (1998), Sources used in health policy research and implications for information retrieval systems. In: Journal of Urban Health, 75 (4), December 1998, pp. 842-852. – ISSN 1099-3460.
22. New York Academy of Medicine Library: Grey Literature Report. http://www.nyam.org/library/greyreport.shtml [Accessed 21 November 2005].
23. CNLP: Research Projects. http://www.cnlp.org/research/project.asp?recid=34 [Accessed 21 November 2005].
24. arXiv.org e-Print Archive. http://arxiv.org/ [Accessed 21 November 2005].
25. Welcome to the NDIIPP. http://digitalpreservation.gov/ [Accessed 21 November 2005].
26. Self-Archiving FAQ. http://www.eprints.org/openaccess/self-faq/ [Accessed 21 November 2005].
27. Overture: The Subversive Proposal. http://www.arl.org/scomm/subversive/sub01.html [Accessed 21 November 2005].
28. Dspace Federation. http://dspace.org/ [Accessed 21 November 2005].
29. Dspace Instances: Dspace Wiki. http://wiki.dspace.org/DspaceInstances [Accessed 21 November 2005].
30. Lynch C.A. (2003), Institutional repositories: essential infrastructure for scholarship in the digital age. In: ARL Bimonthly Report, 226, February 2003, pp. 1-7. Available online at: http://www.arl.org/newsltr/226/ir.html [Accessed 21 November 2005].

31. Foster N.F., Gibbons, S. (2005), Understanding faculty to improve content recruitment for institutional repositories. In: D-Lib Magazine, 11 (1) January 2005. – ISSN 1082-9873. Available online at: http://www.dlib.org/dlib/january05/foster/01foster.html [Accessed 21 November 2005].
32. Ware M. (2004), Institutional repositories and scholarly publishing. In: Learned Publishing, 17 (2) April 2004 – ISSN 0953-1513.
33. van Westrienen G, Lynch CA. (2005), Academic institutional repositories: deployment status in 13 nations as of mid 2005. In: D-Lib Magazine, 11 (9) September 2005. – ISSN 1082-9873. Available online at: http://www.dlib.org/dlib/september05/westrienen/09westrienen.html [Accessed 21 November 2005].
34. Lynch CA, Lippincott JK. Institutional repository deployment in the United States as of early 2005. In: D-Lib Magazine, 11 (9) September 2005. – ISSN 1082-9873. Available online at: http://www.dlib.org/dlib/september05/lynch/09lynch.html [Accessed 21 November 2005].

Author Information

Marcus A. Banks was an Associate Fellow of the US National Library of Medicine (NLM) from 2002-2003. NLM is a component of the US National Institutes of Health. The fellowship provides a detailed introduction to the important role for NLM in the preservation and dissemination of medical information. During his year at NLM Marcus became interested in the opportunities available for improving scholarly communication, as well as the challenge of preserving content that is born digital. As a corollary interest, Marcus grew interested in the challenge of improving access to the grey literature that is scattered across the Web. Since leaving NLM Marcus has published and presented about how the open access publishing movement can serve as a model for improved access to grey literature.
Email: banksm@mskcc.org

Challenges for Collections in New Collaborative Teaching and Learning Environments: Does Grey Literature Fill a Void?

Julia Gelfand
University of California, Irvine, UCI, USA

Abstract
Academic institutions around the globe are introducing and launching bold new teaching initiatives that utilize technologies in the classroom and for remote teaching. The emphasis on distance learning is but one example of this. However, the focus of this paper is to explore the collaborative nature of teaching, learning and studying where faculty and students have to be aware of access to information and the new course management and bibliographic software options to share information while promoting collaborative and group work. The disciplines of business/management and engineering were at the forefront of such work. But, the increasing use of evidence-based management, project design, and case studies common in medicine and other applied fields suggests that more changes in classroom teaching styles are forthcoming. This paper suggests that information needs may be more complex in the future. Grey literature and content found in nontraditional collections such as in repositories, in databases and on courseware platforms will be more prevalent. Influencing commercial products in this direction is common but organizing library collections to lend to such a variety of teaching practices will be an ongoing challenge in a more technology focused learning environment where teaching and learning is done in small groups.

Two primary case studies will be used to demonstrate how faculty are relying differently on content in classroom teaching. One, a medical informatics course, will be described where evidence-based principles are the foundation and how post-graduate medical students use information resources to manage a patient-care case; and the second example will be an undergraduate engineering project design course, where students have to research, plan, design and construct a product. In both of these examples collaborative group work is the teaching method.

The conclusions of this paper will demonstrate that technology is what lends to a collaborative course having a theory, research, writing and evaluation component and where students must form teams or groups. The information needs that are now required include the ability to integrate technology to research, capture, evaluate, document, present, and archive the process of the course and to create the product or final assignment. Utilizing resources such as grey literature contribute to the successes of these case studies.

From the preface of *Everything and Nothing: Deconstructing Evidence-Based Practice*, the late critic and philosopher, Jacques Derrida, writes, "Here is what I wrote, then read, and what I am writing that you are going to read. After which you will again be able to take possession of this preface which in sum you have not yet begun to read, even though, once having read it, you will already have anticipated everything that follows and thus you might just as well dispense with reading the rest."[1] This sets my stage.

Introduction:

A recent article in the *Chronicle of Higher Education* by market researcher Daniel Yankelovich entitled "Ferment and Change: Higher Education in 2015" offers his speculation about what higher education will look like a decade from now. He identifies five trends, which he qualifies "will radically transform higher education in the coming years." Those trends, converging with one another are pressure points for institutions of higher learning. He concludes that if taken seriously and together they pose an enormous challenge that if neglected, will mean serious trouble for higher education and the United States."[2] These trends I surmise have great global relevance and consequences not only for the US but generations of students worldwide. They are:

1. Changing life cycles as our nation's population ages
2. America's growing vulnerability in science and technology
3. The need to understand other cultures and languages
4. Increasing challenges to higher education's commitment to social mobility
5. Public support for other ways of knowing[3]

Compounding these challenges is the mystery of what strategy will be employed to remedy the landscape and suggest some productive resolutions and whether we are equipped to assess when progress has been made in each of these areas. There is not a university or academic institution with which I am familiar that is not in the throes of self-examination - how to make a college education more relevant and prepare a generation of leaders in a wide range of fields in this complex era of extremes - the haves and have nots, the digital divide, the social, political, religious and economic pandemics facing the world.

Richard Hersh, Senior Fellow at the Council for Aid to Education, Co-Director of the Collegiate Learning Assessment Project, and a former president of several liberal arts colleges in the U.S., writing in last month's *Atlantic Monthly* tries to answer the question, "What Does College Teach?" and suggests how "higher education is the only industry in which competitors are rated on the caliber of their customers rather than on their product… because we don't have any other way to measure collegiate quality."[4] Rankings and surveys hardly get to the real question of impact of one's education, but those that ask students to assess their experiences appear more promising according to Hersh. Assessment remains the buzzword among those that study different aspects of higher education and the emphasis that is universally accepted includes how students should master the following skills:

- Weigh, organize and synthesize evidence from different sources
- Distinguish rational from emotional arguments; facts from opinion
- Analyze data
- Deal with inadequate, ambiguous or conflicting information
- Spot deception and holes in the arguments of others
- Recognize what information is or is not relevant to the task at hand
- Identify additional information that might help to resolve issues[5]

This reinforces how strong sentiments are that articulation, presentation, computation, analytical and research skills are critical in defining a well educated individual. "If higher education is to take the lead in innovation, and to define, seek and demand excellence from its students, today's academic should be satisfied with nothing less"[6] concludes Hersh.

Thus, the motivation of this paper, to explore more effective ways to teach using new techniques and resources while engaging in a new collaboration with and between students. In the academic library environment information literacy is the umbrella of such goals. This strategy encourages librarians to participate and partner with faculty to incorporate informatics in the core curriculum. How to find, manipulate, evaluate and apply relevant information is the essence. The Association of College and Research Libraries (ACRL) has defined Information Literacy as

The set of skills needed to find, retrieve, analyze, and use information. Information literacy is more closely tied to course-integrated instruction but it extends far beyond coordination between the reference librarian and the individual faculty member. Even a cursory review of the Information Literacy Competency Standards (link) will show that there is much more to information literacy competence than library-related research. Students must demonstrate competencies in formulating research questions and in their ability to use information as well as an understanding of ethical and legal issues surrounding information. This requires a campus culture of collaboration and focus on student learning.[7]

Addressing the problem

Librarians are engaged in directing users to appropriate resources, and teaching how to use them, but they are also involved in the content management of that intellectual property, by explaining best practices of scholarly communication, distinguishing copying from interpreting, and how to cite, organize and recall information, using new software such as bibliographic management software (EndNote, RefWorks are such examples) and the utility of course management platforms used not only in distance education but for group projects, collaboration, team teaching, integrated learning, multimedia, etc (WebCT, Blackboard and hosts of other products and local systems are in this marketplace).

Evidence-Based Practice

Still the effort remains most fuzzy and distorted when we examine pedagogy, teaching styles and learning behaviors and outcomes. Recently, medical education at my institution and throughout the world has explored how evidenced-based principles (EBP) can advance the teaching process in clinical fields like nursing, public health/epidemiology, mental health, medicine and now social services and welfare and we are also seeing it applied in business and management curriculum, criminal justice and law enforcement. This author is trying to pioneer some applications in engineering design and project management curriculum where she sees enormous utility of EBP. It also has application in the daily practice of consumerism and how we evaluate products and services we need, acquire or lease. Once called, research-enhanced clinical care, Evidence-based medicine is the coined phrase known as EBM and it builds on a tool-kit dependency of core resources and how to know when it is appropriate to use them.

We can't forget the famous quotes from the late Carl Sagan, "If we did not respect the evidence, we would have very little leverage in our quest for truth" and "Absence of evidence is not evidence of absence." In EBP, learning objectives are defined and influenced by best evidence available in the decision-making process using the expertise, values and preferences of the individual or patient in medical settings, their families and the communities who are served and thus culled from available resources. Evidence-based care involves the integration of best research evidence with clinical expertise and patient values. This "holistic" approach of bringing together all the component parts is the centrality of evidence-based practice and involves attitudes, skills and behaviors and suggests how organic and fluid it is as new evidence presents itself. An example of why this method exists can be illustrated by those professionals who think that a patient who consults for or seeks services and deserves it - "they have an ethical and often a legal right to know the answers to such questions as 'what evidence is there

that this treatment will help me?" What is the cause of my illness or situation,' and only a transparent process that equips patients and practitioners with powerful tools to address important questions of therapy, prognosis, diagnosis and etiology."[8] Again, it builds on being able to argue the case, explain it cogently to everyone who needs to know, summarize it for the medical file and demands use of resources to support specific values.

Many proponents of EBP suggest that the practice allows for better management of the proliferating literature in nearly every field, while the human side of medicine is also emphasized, such as the compassion, understanding, personal choice factors and gains more attention in the overall delivery of service. Consistent rationale for why EBP is valuable is well summarized as:

- Ethically - what will help me feel better
- Allows decision process to become more transparent
- Offers useful tools to clinicians and patients
- Helps to identify a lack of resources
- Highlights gaps in knowledge
- Enables all parties to get the best from a huge amount of information
- Identifies the variability in health care[9]

There are also some weaknesses in EBM and they may include a cultural bias, or focus on traditional drugs versus more alternative forms of treatment, therapy or medicine, and publication bias can be rather strong where previous results are not fully reflective of true samples and experiences. Without dwelling on how it may be dysfunctional as a practice, the evolution of EBP has been described as having five essential steps in its practice:

- Convert information needs into answerable questions
- Track down, with maximum efficiency, the best evidence with which to answer them
- Critically appraise that evidence performance for its validity and usefulness
- Apply results of the appraisal in clinical practice
- Evaluate performance[10]

What stands out as a unique feature in EBP is that it requires that a systematic literature search be conducted and thus enhances the accountability to stakeholders. The key elements that resurface in nearly all scholarship in EBP are intensive and defined as methods for:

- Finding Research Resources
- Accessing Research
- Appraising Research
- Applying Research

The background questions which set the stage for the interactions are answered by using "background" resources such as textbooks, web resources and narrative reviews. In addition they can be divided into two types of clinical questions, the Background and Foreground. The latter are more specific, focused and complex and tend to be answered by consulting primary and secondary sources and contains the PICO elements for structure:

P = Patient/Population/Problem/Situation
I = Intervention (treatment of method)
C = Comparison (alternatives)
O = Outcome interest

<u>The Collection: Introducing Grey Literature</u>

By refining the literature search and interview questions with patient subject one achieves a hierarchy of types and levels of evidence to reflect:

Meta-Analysis
Systematic Review
Randomized Controlled Trial
Cohort Studies
Case Control Studies
Case Series/ Case Reports
Basic Laboratory and Animal Research

By completing an exercise that converts the patient/provider interview into PICO format and identifies search concepts, one can then evaluate the online resources with traditional criteria:

- Who operates the site?
- Who supports or pays for the site?
- What is the purpose of the site?
- Where does the information come from?
- What is the basis of the information?
- How and on what basis is the information selected?
- How current is the information?
- How does the site choose links to other sites?
- What information about the user does the site collect and why?
- How does the site manage interactions with visitors?

The EBP competencies conclude that:

- Sufficient research must have been published on the specific topic
- Provider must have sufficient skill in access and critically analyzing research

Changes in treatment plan are made due to evidence-based experience.

The teaching skills that are critical include achieving a productive interview, the balance between listening and talking has to be right. Getting sufficient information to complete the PICO is the objective. Getting it right, efficiently is the aim and is usually achieved by strong relationship building skills, establishing a high level of trust and perfecting questioning skills and mastering the critical thinking process which has the following components:

- Problem identification and analysis
- Clarification of meaning
- Gathering the evidence
- Assessing the evidence
- Inferring conclusions
- Other considerations
- Overall judgment[11]

Depending on the subject matter and the traditions of the corresponding literature, one can sense the value of evidence-based practices. For example, in a subject where the literature leads one to a benefit that is not so individually-based or case management focused such as in the provider/patient relationship but may be more community-intensive, such as in public health, epidemiology, problem-solving, corrections, or engineering, dealing with solutions to problems in aerodynamics, new product development, seismic ratings, etc. The interventions may include:

- Policies of governments and non-government organizations
- Laws and regulations
- Organizational development initiatives
- Community development initiatives
- Education of individuals and communities
- Engineering and technical developments for re-engineering
- Service development and delivery
- Communication, including social marketing[12]

The medical education component, obviously more mature has a corresponding set of resources to accompany parties to conduct information seeking, appraisal and application. These tools form the tool-kit in medical libraries that is now increasingly extensive and reflects the expensive compilation and packaging of a range of products such as:

- reference source-books (MedlinePlus),
- textbooks
- empirical studies, journal literature databases (PubMed, etc)
- systematic review literature (Cochrane Library)
- drug and clinical studies (PubMed Clinical Queries, UpToDate, Natural Medicines Comprehensive Database, eMedicine Database, TRIP Plus - Turning Research into Practice), InfoPOEMS/InfoRetriever, GIDEON
- specific resources from different specialties like nursing, etc), and research (Joanna Briggs Institute Reports, etc)
- grey literature - the more nontraditional, alternative sources although they are increasingly mainstream today in most treatment plans
- new products just being rolled out, like EBSCO's Evidence-Based Complementary Medicine (EBCM) database to be launched in May 2006.

In engineering and design disciplines, finding the best evidence available is more difficult because the resources are not yet produced with this strategy in mind. Engineers are very logical and rely heavily on their corpus of content that looks at data and may include, maps, field sample data in civil engineering, patents, handbooks, standards, dissertations, ergonomics and safety issues, and other government focused collections of innovation and intellectual capital, a range of materials - polymers, concrete, steel, ceramics, glass, etc where the physical sciences of chemistry, physics and mathematics influence relevance. Engineers are often on the prowl for unpublished information and will look at materials on the web, created by or contributed to by competitors often before they seek out traditional library-held resources, even though they rely upon key reference sources, formularies, handbook series, etc., all not surprisingly available online today. Increasingly, scholarly communication practices has influenced this as engineers have followed high-energy physicists in creating pre and post print archives, using the conference proceeding as a primary source and showing commitments to institutional repositories by depositing a range of materials. Still, there remains the hesitation to do so before exploring possibilities for intellectual transfer and potential for patents or new submissions and revisions of standards.

Collaboration in engineering has also been the norm far longer, as laboratories are larger requiring more people and the commitment to interdisciplinarity has never been more evident. Teams of different kinds of engineers work together in both academic and industry research centers and now the computer science /IT nucleus is even more integrated in the research process as technology leads the way. Also, both medical/clinical and engineering professionals are relying upon the business and economics literature to chart progress of new products, innovation, marketshare penetration, and other distribution channels to position themselves for post-discovery success based on demographic patterns. New forms of publishing including interactive media, blogs and wikis suggest the need for more intense evaluation in assessing quality and finding best evidence. Engineers also like current awareness services such as customizing personal "@mylibrary" pages and subscribing to RSS feeds.

In the engineering sector, more of an emphasis is on re-engineering, or refining products. The burgeoning field of biomedical engineering promotes that as newer technologies influence new procedures, products and innovation. The role of the laser is a good example, and there are hosts of others - making surgeries less invasive, products and processes more comfortable like for dialysis and transplants, and the computer storage capacity has become more affordable for large genetic and biological computing needs to handle DNA banks, visualization, etc.

Students in engineering project design come prepared to test, push limits and see their creative spirit demonstrate potential for entrpreneurship. However, learning and teaching evidence-based technique has not always been easy. Advanced and nonplussed by technology, young engineers are ambitious but not always focused on the mechanics of conducting a literature review. The "will try, can fix and will do" mentality more accurately describes the priorities over posing questions and finding sufficient evidence. However, when done, the package, product and research process appears more complete and provides a high degree of satisfaction and a record of the process. It proves that a lot of consultation took place and interaction with peers, specialists and advisors. For example, feedback is given by faculty and peers to a business plan. The strength of evidence as tested in clinical settings has more relevance today in the applied sciences and technology than one can imagine. This can be summarized by the Five Strengths of Evidence:

I	Strong evidence from at least one systematic review of multiple well-designed randomized controlled trials
II	Strong evidence from at least one properly designed randomized controlled trial of appropriate size
III	Evidence from well-designed trials without randomization, single group pre-post, cohort, time series or matched case-control studies
IV	Evidence from well-designed non-experimental studies from more than one center or research group
V	Opinions of respected authorities, based on clinical evidence, descriptive studies or reports of expert committees.[13]

Back to Instruction - Conclusions about Grey Literature

Technology drives the current teaching environment. Evidence-based practices in whatever discipline offer a theoretical framework that reaffirms a strategy or process to follow. It includes a strong research and evaluation component and due to the proliferation of information in all formats new information products that are created with evidence-based principles as their structure are available to help make the process easier. Finding the best evidence and evaluating information with a higher degree of confidence suggests how important evidence-based collaboration can be in the instruction and research environments in clinical and applied disciplines.

Grey Literature does fill a void when one examines the range of sources consulted to achieve successful evidence-based practices - one needs to utilize the nontraditional resources in many cases to substantiate best practices. The sheer number of resources one finds by entering a search of evidenced-based + grey literature supports this beyond observation. One of the most comprehensive websites containing the range of grey literature within an evidence-based context can be examined at the University of Calgary Health Sciences Library (http://library.ucalgary.ca/brancehes/hsl-greyliterature/index.php?) and the work done in software engineering and evidence-based applications by Barbara Kitchenham at the University of Keele (http://ease.cs.keele.ac.uk/keys2005.html) shows an impressive launch in this subject area. Bibliographies and web searching supports many more creative ways of introducing evidence-based practices.

Evidence based practice encourages being conversant with information architecture. It promotes how engineers with this familiarity can influence and redesign e-commerce, continue to actively play in the world of personalized media and create sophisticated new retail and living environments. I imagine that teaching tools in EBP are becoming more available and distance education is currently contributing to the need for them.

Author Information

Julia Gelfand has been a librarian with the University of California, Irvine Libraries since 1981. She has been tracking the grey literature movement since the late 1980s and has participated in all of the previous GL conferences and has published and presented widely on different topics in grey literature. Her particular interests are in scholarly communications, electronic publishing, collection development, bibliography of science and technology, and she thinks that with more emphasis on networking and digital libraries, Grey Literature has a very interesting future. She is currently the chair of the IFLA Science & Technology Section and vice-chair/chair-elect of the ALA ACRL Science & Technology Section. Email: jgelfand@uci.edu

References:

[1] Jaques Derrida in Preface to *Everything and Nothing: Deconstructing Evidence-Based Practice* by Dawn Freshwater and Gary Rolfe. Abingdon, Oxon: Routledge, 2004.
[2] Daniel Yankelovich, "Ferment and Change: Higher Education in 2015." *Chronicle of Higher Education*, November 23, 2005: B6.
[3] *Ibid*: B7-8.
[4] Richard Hersh, "What Does College Teach? It's time to put an end to "faith-based" acceptance of higher education's quality: College 2005." *Atlantic Monthly*, 296.4, November, 2005:140.
[5] *Ibid*: 144.
[6] *Ibid*: 144.
[7] Association of College & Research Libraries website (http://www.ala.org/ala/acrl/acrlissues/acrlinfolit/infolitoverview/infolitglossary/infolitglossary.htm)
[8] Simon Hatcher, et al, *Evidence-Based Mental Health Care.* Edinburgh, Elsevier Churchill Livingstone, 2005: 1.
[9] I*bid*, 3.
[10] D.L. Sackett and W.M.C. Rosenberg, "The need for evidence-based medicine," *Journal of the Royal Society of Medicine*, 88, 1995: 620.
[11] Milos Jenicek and David L. Hitchcock, *Evidence-Based Practice: Logic and Critical Thinking in Medicine*. Chicago, IL: American Medical Assoc. Press, 2005: 110-111.
[12] Michael Frommer and Lucie Rychetnik, "From Evidence-Based Medicine to Evidence-Based Public Health," in Vivian Lin, and Brendan Gibson, eds., *Evidence-based Health Policy: Problems and Possibilities*. Oxford: Oxford University Press, 2003: 59.
[13] J.A. Muir Gray, Evidence-Based Health Care. Oxford: Churchill Livingstone, 1997.|||
The author acknowledges the work of her colleagues, Linda Murphy and Steve Clancy in Evidence-Based Practice and Nurse Practitioners.

J-STAGE: System for Publishing and Linking Electronic Journals

Mitsutoshi Wada
Japan Science and Technology Agency, JST, Japan

Abstract

Since 1999 Japan Science and Technology Agency (JST) has been operating J-STAGE (Japan Science and Technology Information Aggregator, Electronic), a system helping academic societies in Japan to publish their electronic journals. J-STAGE supports the entire process of online publishing: submission / reception of manuscripts, peer review, judgment, editing, and loading articles on the web site. Users can search and retrieve articles by keywords or author names, and they can view the abstracts freely. The full texts can be downloaded in PDF format.

As of the end of October 2005, 245 journals and 89 conference proceedings are currently published and about 163,000 articles are loaded on J-STAGE. Total number of accesses to every pages of J-STAGE in a month is over 1,000,000, and more than 300,000 PDFs are downloaded every month.

About half of the journals on J-STAGE are written in English, and more than 80% of the articles have English abstracts. J-STAGE is accessed from over 120 countries in the world, and about 70% of the accesses to the articles on J-STAGE come from foreign countries outside Japan.

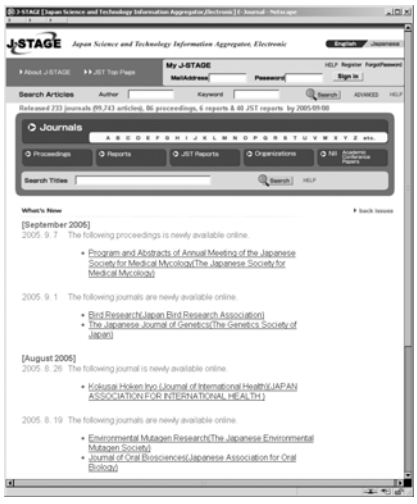
Using JST Link Center, launched in 2002, articles on J-STAGE are linked to other articles on various worldwide electronic journal sites and bibliographic databases via CrossRef, PubMed LinkOut, ChemPort, and JOISLink. More than half of the accesses to the articles on J-STAGE come through the link from other sites, especially bibliographic databases like PubMed.

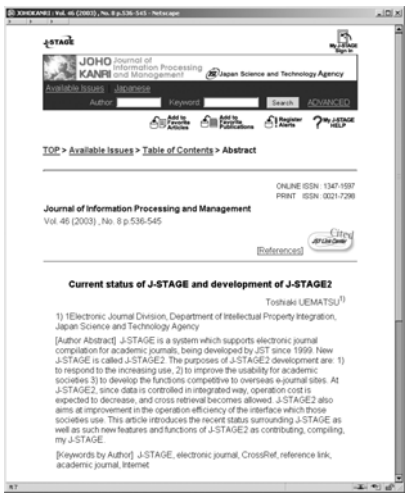
In 2005, JST has begun the Back Issue Digital Archiving Initiative. JST will produce back file data back to the first issues of major academic journals in Japan.

1. Introduction

Electronic journals have many benefits compared with printed journals. Once they are loaded on web sites, they can be accessed at any time from anywhere in the world through the Internet. Digitized articles can be searched and retrieved by keywords. Also they can be linked to / from articles on other sites.

J-STAGE (Japan Science and Technology Information Aggregator, Electronic) is a free platform of electronic journals for academic societies in Japan. It was developed and has been operated by Japan Science and Technology Agency (JST).





J-STAGE Top Page Abstract Page of an Article

Fig. 1 Page Images of J-STAGE

2. J-STAGE System

J-STAGE was launched in October 1999. In order to promote digitization and publication of scientific and technical information in Japan, J-STAGE helps Japanese academic societies to create and publish their electronic journals. The system is available 24 hours a day, 7days a week, and free of charge for academic societies to use. The system can be customized for each academic society. The copyrights of the articles on J-STAGE are still held by their academic societies. J-STAGE encourages academic community in Japan to disseminate outstanding research results to the world instantaneously.

For J-STAGE, JST gets the Japanese government budget about 1.2 billion Japanese Yen for each fiscal year. The total costs (including the systems development, maintenance, and operational costs) since 1998 is about 9.5 billion Japanese Yen.

J-STAGE supports the entire process of online publishing: submission / reception of manuscripts, peer

review, judgment, editing, and publication. Full text PDF files and bibliographic data of the articles are loaded on J-STAGE. Citation data for linking to other articles and full text data for searching are also loaded. Errata and electronic supplements such as video or audio data can be added to the articles.

Without any registrations, users can search and retrieve the articles with their bibliographical data or keywords, and they can view the abstracts freely. They can also download the full texts in PDF format. Access policy of each journal is decided by its publisher society. Most of the journals (about 80%) on J-STAGE are now freely accessible, but some journals require authentication to read full text. J-STAGE provides authentication by subscriber numbers or user-ids, and authentication by IP addresses of subscribing organizations. Pay-per-view system for non-subscribers to buy specified articles by credit cards is also available.

3. Journals on J-STAGE

As of the end of October 2005, 327 Japanese academic societies use J-STAGE to publish their journals or conference proceedings. J-STAGE currently hosts 245 journals, 89 conference proceedings, 6 reports, and 40 JST reports. About 163,000 articles are available online now. Fig. 2 shows growth of the number of titles on J-STAGE.

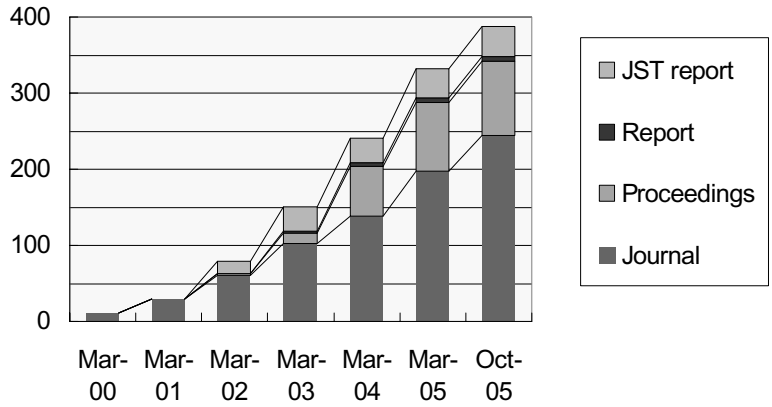


Fig. 2 Growth of the Number of Titles on J-STAGE

As shown in Fig. 3, about half of journals (49%) on J-STAGE are written in English language, 31% of them contain both English and Japanese articles, and the rest (20%) are written in Japanese. Because most of the articles written in Japanese also have English titles and English abstracts, more than 80% of articles' abstracts on J-STAGE can be read from foreign countries.

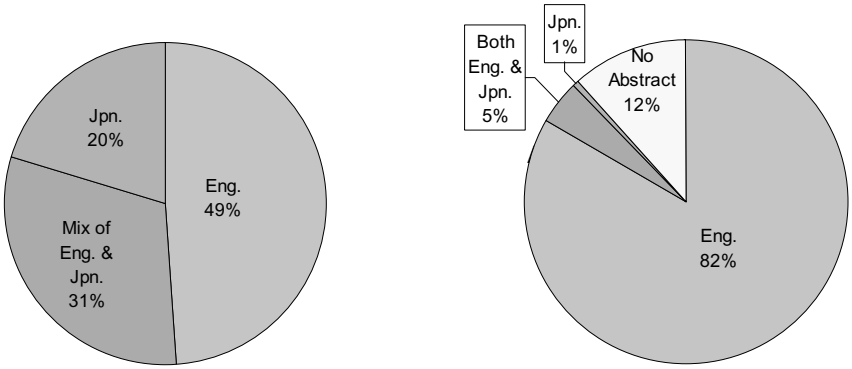


Fig. 3 Language Distribution of Journals on J-STAGE

4. Accesses to J-STAGE

Total number of accesses to every pages of J-STAGE in a month is over 1,000,000. More than 300,000 full-text PDFs are downloaded per month. Fig. 4 shows the growth of number of articles on J-STAGE and number of accesses to J-STAGE. Both are increasing steadily year by year.

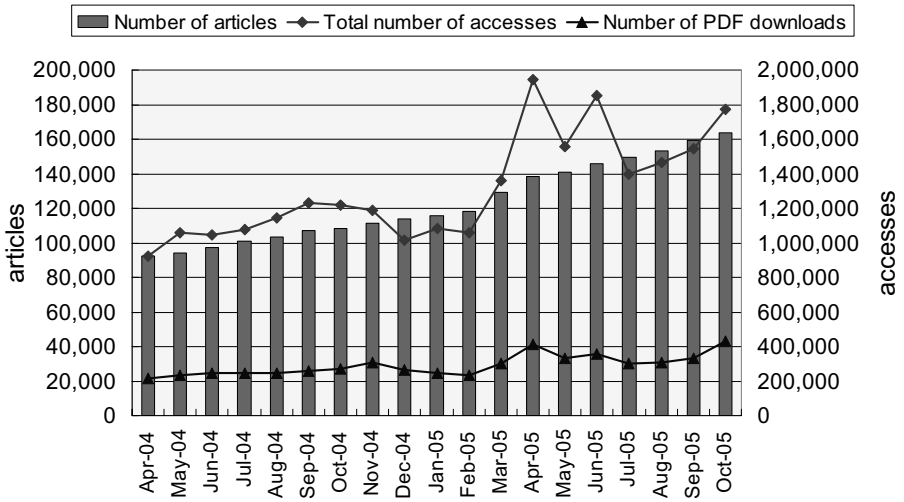


Fig. 4 Number of Articles on J-STAGE and Number of Accesses to J-STAGE

During a year and a half period from April 2004 to September 2005, J-STAGE was accessed from 122 countries in the world, and about 70% of the accesses to the articles on J-STAGE came from foreign countries outside Japan.

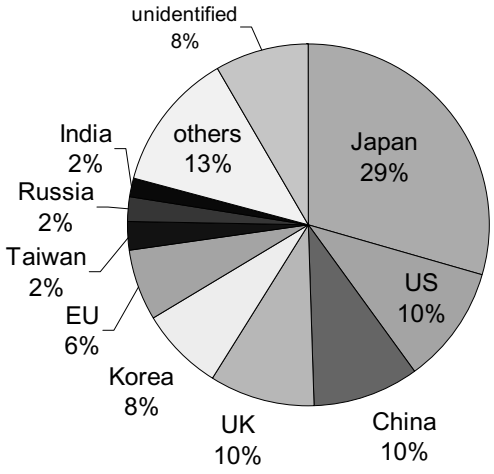


Fig 5. User Distribution among Countries

5. Links from / to J-STAGE

In 2002, JST launched JST Link Center, which is a subsystem of J-STAGE to create links between electronic articles. Using JST Link Center, articles on J-STAGE are linked to other articles on various worldwide electronic journal sites and bibliographic databases via CrossRef, PubMed LinkOut, ChemPort, and JOISLink. (Fig. 6) [1]

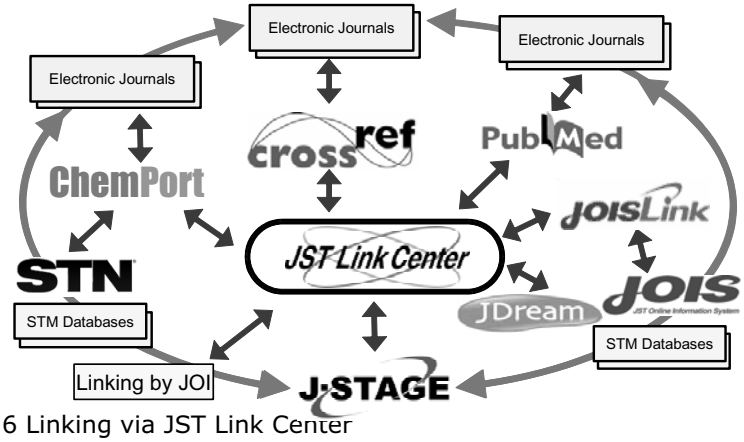


Fig. 6 Linking via JST Link Center

Fig. 7 shows that more than half of the accesses to the articles on J-STAGE come through the links from other sites, especially bibliographic databases like PubMed.

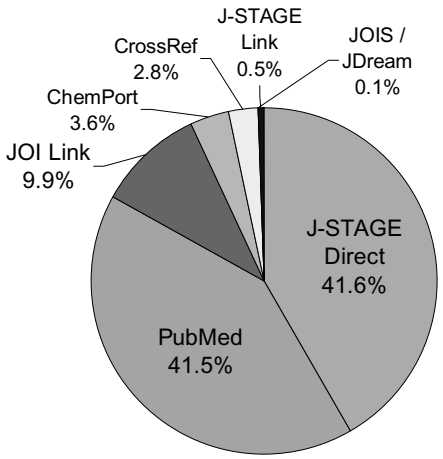


Fig 7. User Distribution among Access Channels

6. New Functions

JST developed the new version of the system and released it in January 2004. The new system becomes easier to used and easier to be customized for each academic society. It supports new services: cross search function among all articles on J-STAGE, virtual journal service (publishing new journals by combining articles already published in other journals), My J-STAGE service (to allow users signed-in by their registered e-mail addresses to register their favorite journals, favorite search conditions, and e-mail alert for newly arrived articles of specified journals), advance publication service (publishing new articles just after accepted before assignment of issue numbers and start pages), and HTML full text publication. [2]

Links to the citing articles from the original cited articles are now available among the articles on J-STAGE and the participating publishers in CrossRef's Forward Linking. [3]

A systems for electronic submission, reviewing, and tracking of manuscripts is also available. About 30 academic societies are using or going to use this system. [4]

7. Future Plans

JST plans to encourage more Japanese academic societies to use J-STAGE to publish their journals. The immediate target is more than 500 titles by the end of fiscal 2006.

JST also plans to extend links between J-STAGE and other sites; cooperation with Google to lead more users to J-STAGE, participation in CrossRef Search, and provision of OpenURL.

JST plans more new functions like alert mail when an article is cited, and provision of usage reports in COUNTER (Counting Online Usage of NeTworked Electronic Resource) format.

In 2005, JST has begun the Back Issue Digital Archiving Initiative. JST will produce back file data back to the first issues of major academic journals in Japan. They will be loaded onto J-STAGE and will be available through the Internet. [5]

References

[1] Kubota, S., Uematsu, T., Yamazaki, T., Kondo, Y., Tokizane, S., Omi, A. (2005), Linking of electronic journals via JST Link Center. - In: JOHO KANRI (Journal of Information Processing and Management), 48 (3), pp. 149-155. - ISSN 0021-7298 [in Japanese]

[2] Uematsu, T. (2003), Current status of J-STAGE and development of J-STAGE2. - In: JOHO KANRI (Journal of Information Processing and Management), 46 (8), pp. 536-545. - ISSN 0021-7298 [in Japanese]

[3] Kubota, S., Arakawa, N., Wada, M., Kondo, Y., Kokubo, H., Takumi Yamazaki, T. (2005), New services with JST Link Center - Crawling J-STAGE by Google and citing & cited articles on electronic journals -. - In: Preprints of the 2nd Information Professional Symposium, pp. 101-104. - ISSN 1349-5941 [in Japanese]

[4] Arakawa, N., Wada, M., Tokizane, S. (2005), J-STAGE Manuscript Submission and Peer-Reviewing System - Development, Implementation and Future Plan -. - In: Preprints of the 2nd Information Professional Symposium, pp. 95-99. - ISSN 1349-5941 [in Japanese]

[5] Wada, M, Kubota, S., Omi, A. (2005), Current Status and Usage Trend of J-STAGE. - In: Preprints of the 2nd Information Professional Symposium, pp. 87-90. - ISSN 1349-5941 [in Japanese]

PsyDok: Electronic Full text Archive for Psychological Documents

Ulrich Herb
Saarland University and State Library, Germany

Abstract
The Saarland University and State Library (Saarlaendische Universitaets- und Landesbibliothek SULB) runs the Special Subject Collection Psychology which is part of an information system for the supra-regional literature supply in Germany. With the increasing opportunities for electronic information the following question emerged for the Special Subject Collection: how can a contribution to the availability and preservation of electronic documents particularly with regards to grey literature be made? Giving an answer SULB established the repository PsyDok in autumn 2002. PsyDok is one of the few subject-specific digital repositories in Germany. It is a registered data provider of the Open Archives Initiative (OAI) and certified by the German Initiative for Networkinformation (DINI), which developed a variety of standards digital repositories have to comply with to achieve this mark of quality. PsyDok offers many interfaces to disciplinary and multidisciplinary retrieval and information systems. The allocation of Uniform Resource Names (URN) guarantees persistent availability and citeability of the documents. Several services and mechanisms guarantee not only a maximum of visibility and a rapid and global scientific exchange and flow of information stored on PsyDok but also its long term availability and its scientific utilization. PsyDok fits the principles of "Open Access": Many different document types may be published, but PsyDok focuses particularly on diploma theses, dissertations, professorial dissertations and grey literature. Regarding aspects of visibility, information exchange and scientific utilization especially producers of grey literature benefit from PsyDok.

1 Introduction: Scientific information supply in Germany: Special subject collections and virtual libraries

In Germany scientific information supply is coined by a distributed, supra-regional system of special subject collections[1] (in German: Sondersammelgebiet, SSG) which are tied to scientific libraries. There are 121 special subject collections which can be found in more than 40 libraries all over Germany. These special subject collections are funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).[2] The main function of a SSG is to build a collection of subject related scientific literature which is as complete as possible; this includes domestic and foreign research literature (including electronic media, microfilms and multimedia objects) and especially scientific journals. Within this system Saarland University and State Library (Saarlaendische Universitaets- und Landesbibliothek SULB)[3] runs the Special Subject Collection Psychology since 1966. In 1998 the German Research Foundation adopted a memorandum focusing on the future trends of scientific information supply. This memorandum asked for the enhancement of information supply with regard to the means and opportunities of up to date information technology. Accordingly the German Research Foundation started funding so called Virtual Libraries. In 1999 the Virtual Library of Psychology at Saarland University and State Library went online. Virtual Libraries claimed to be central entry-points to scientific information by bunching subject-specific information on the WWW and making it accessible. These virtual libraries did not try to preserve or publish information but only to link to it. Therefore most virtual libraries are using a subject gateway[4] and a subject specific meta search engine. They usually did not make any contribution to the availability of scientific information on the www.

2 The availability of psychological information

In autumn 2002 SULB and the Institute for Psychology Information (Zentrum fuer Psychologische Information und Dokumentation ZPID, Germany)[5] launched a joint project named „Digital Psychology Information (DPI)", which was funded by the German Research Society until spring 2005. The aim of DPI was to improve the availability and accessibility of electronic psychological information. One of the work packages defined within DPI was the implementation of the subject specific repository PsyDok[6] run by SULB; its home page lives at http://psydok.sulb.uni-saarland.de. By establishing a service like PsyDok SULB did not only react to the ephemeral nature of most online information but made also a contribution to the availability and preservation of electronic documents particularly with regards to grey literature.

3 PsyDok: A subject specific scientific repository

3.1 Why PsyDok?

PsyDok is a nationwide and central publication platform for scientific psychological documents. Both publishing and using PsyDok-documents is free of charge, which also means the PsyDok fulfils the demands of Open Access. Unlike scientific documents published on conventional institutional repositories PsyDok-documents reach a much higher, maximized visibility and dissemination. This is based upon its centralized function within the discipline psychology and its manifold integration in both disciplinary and multidisciplinary retrieval systems. This systematic integration offers overwhelming advantages especially compared with the still usual but deficient practice of publishing documents on web servers of academic institutions or faculties. Such information can not be searched (or found) systematically, if they are found at all this happens accidentally by click-by-click-retrieval. In addition publications like these are of a very limited value within scientific discourse: both span of life and availability are short (or are at least unpredictable), permanent access, their immutability and citeability are doubtful. All in all PsyDok-publications do not fall victim to the informational fragmentation and do not vanish in the WWW's noise. By offering a rapid and global flow of scientific information PsyDok applies to the motto "maximizing research impact by maximizing research access".

3.2 What offers PsyDok?

In accordance to the special subject collection principle PsyDok acts as a supraregional, subject specific repository for digital scientific documents. It is the central point for free of charge psychological fulltext publishing. Many different document types may be published and found on PsyDok, but it focuses particularly on diploma and master theses, dissertations, professorial dissertations, preprints, teaching material, psychological scales and grey literature. Monographs, book chapters, journals and articles from journals can also be published as PsyDok is a postprint-server as well. PsyDok gives scientists an infrastructure for self-archiving and facilitates their walk on the green road to Open Access. Most of the documents published on PsyDok can be assigned to the genre grey literature. This category is of great value for the scientists as a study by the German Research Foundation proved[7] (DFG 2005, p. 22). PsyDok is not a genuine publication platform, all documents published were produced in another context or for another purpose and afterwards deposited on the repository – none of them was originally produced to be published on PsyDok.

3.3 What about technical matters?

PsyDok is based upon the software OPUS,[8] which was developed at the University of Stuttgart (Germany). Till now 50 German universities are using OPUS to run their digital repositories. It requires UNIX/Linux with a MySQL- or MSQL-Database and an APACHE/PHP-environment. OPUS is also mentioned in the Open Society Institute's guide to institutional repository software (Open Society Institute, 2004). The preferred file formats to be published on PsyDok are PDF-, PostScript- or HTML-files, the ideal format to date is PDF/A.[9]

3.4 What about time, digital archaeology and 404 errors?

Saarland University and State Library ensures permanent and long-term access, long-term usability and citeability for all documents published. In order to supply a persistent and stable electronic location of its documents, PsyDok offers a persistent identifier in form of an Uniform Resource Name[10] (URN) for each document. Each URN is registered by the German National Library[11] (Die Deutsche Bibliothek, DDB). URNs are comparable to the Digital Object Identifier (DOI)[12] or the less known Persistent URL (PURL).[13] With the allocation of URNs the persistent availability and citeability of the documents is guaranteed. For each document with a persistent identifier (no matter if in the form of an URN, DOI or PURL) permanent access, permanent reference and citeability can be acknowledged.[14] SULB is not going to develop techniques of long-term usability and access, nevertheless these advantages can be ensured. Technologies like these need a lot of research and development which can only be done at a subordinate level, but not by a single university library. In Germany several institutions (as the DDB) and projects (as nestor[15] or kopal[16]) are focusing on the aspects of long-term access and availability. Technologies and practices developed and evaluated within these projects will be applied to PsyDok. On the long run long-term usability seems to need the design and implementation of emulation and migration technologies.

3.5 What about quality and standardisations?

As PsyDok is not a genuine publication platform Saarland University and State Library does not have any influence on the genesis of the documents which are published on PsyDok and hence has little chance to apply any kind of quality control.[17] Nevertheless several related principles are taking effect before documents are deposited on PsyDok, so quality control can be derived from the type of document to be published.

- With institutional series or working papers series an editorial review should be assumed
- With dissertations and professorial dissertations a sufficient quality is at hand due to the type of document
- Diploma and master theses will only be published if the referee explicitly states a recommendation
- With preprints the authors revisable reputation can be derived from former publications or the quality of the document can be derived by the later formal publication
- Postprints are controlled by the publishing houses expert stuff, mostly by a peer review
- Each psychological faculty or institute is encouraged to use PsyDok's internal peer review module

Within the Open Archives Community (OAI)[18] there are vitally important discussions going on how an extended OAI Protocol for Metadata Harvesting (PMH) could offer a peer review functionality for records on OAI-Servers. This technique is essentially based on digital social networking models and focuses on preprints, but may be adapted to other document types as well (Rodriguez et al. 2005). This means that in the future digital repositories as PsyDok might offer a genuine quality control. Regarding standardisations, technological and organizational quality control of digital publication services in Germany the German Initiative for Networkinformation (German: Deutsche Initiative für Netzwerkinformation e.V., DINI)[19] has to be mentioned. DINI developed a variety of standards digital repositories have to comply with in order to become a certified DINI-Server. DINI intends to make electronic publishing with digital repositories more serious and reliable; the DINI-Certificate is a mark of quality. The criteria checked by DINI-experts concern organizational matters (as workflow optimization), questions of data integrity and authenticity, guaranteed archival storage periods, detailed description of the service, data export into long-term archival systems, questions of content indexing and question of metadata management and export. PsyDok met all these criteria and is a certified DINI-Server.[20] Another and even more important sign of standardization is the compliance with the OAI PMH.

3.6 What about dissemination and visibility?

Discussing the advantages of electronic publishing compared to conventional print publishing two of the main arguments in favour of the e-model are both the accelerated circulation and availability and the maximized visibility of digital publications. Notably relating to visibility PsyDok offers lots of benefits - especially comparing it with other institutional repositories. PsyDok is exceptional proper for the integration into disciplinary and multidisciplinary retrieval systems, since PsyDok is - unlike an institutional repository - no general dealer but a scientific specialist dealer. Subject specific search engines may index PsyDok as a whole without picking single domain related documents while multidisciplinary catalogues may integrate PsyDok data easily into their browsing scheme because all documents only focus on one subject. PsyDok offers many interfaces to disciplinary and multidisciplinary retrieval and information systems in order to facilitate the integration. First of all PsyDok is a registered OAI-data provider so OAI-based search engines (OAI service providers) can add its Metadata to their index. Other multidisciplinary retrieval systems are the *Karlsruhe Virtual Catalogue* (KVK)[21], a meta search interface which queries library and book trade catalogues, *OASE*[22] (Open Access to Scientific Literature) which allows to search the bibliographical data of the most important document servers in Germany and abroad. PsyDok documents can also be found via the OPUS meta search[23], a search interfaces which offers a simultaneous search in all OPUS-based digital repositories. PsyDok publications can be found via Scirus[24] (Elsevier's scientific search engine) and the Google spin-off for scientific documents, Google Scholar[25]. Usually most scientists pay much attention on disciplinary information systems, so PsyDok is well integrated into retrieval systems focusing strongly on psychology. First to be mentioned is the subject gateway PsychLinker[26]. PsyDok offers also an interface to ZPID's Psyndex (the most comprehensive abstract database of psychological literature, audiovisual media, and tests from the German-speaking countries).[27] All diploma and master theses are additionally listed in the diploma theses database of ZPID. Each PsyDok document is indexed by the robot-based psychology search engine PsychSpider[28] of ZPID. Another central access point for psychological information is the OPAC[29] of SULB with its special subject collection, which registers numerous queries for psychological information a day and which of course lists PsyDok documents as well. Even information systems which mainly focus on other subjects than psychology integrate PsyDok data. For instance selected documents can be found via the Collection of Computer Science Bibliographies[30] and CiteSeer[31]. Last but not least PsyDok-documents are also listed in WWW search engines where they produce very high rankings. Naturally Saarland University and State Library is eager to expand the number of retrieval systems hinting to PsyDok. PsyDok will be added to the index of the Social Science Search Engine run by the Social Science

Information Gateway (SOSIG)[32]. The same applies to the search engine PsychCrawler[33] of the American Psychological Association (APA)[34] and to Scirus Repository Search. The integration into the new APA database for grey literature, PsycEXTRA[35] is considered, it will depend on legal matters. Dissemination and visibility of PsyDok-documents are a multiple of what can be reached through conventional library catalogues, www search engines, traditional literature databases or subject mixed repositories. This can be proved by a simple example: During October 2005 the diploma theses "Gesichtsform und Attraktivität" by Markus Becker was viewed online 212 times. The print version of the same diploma theses is held by only one library (and can not be borrowed there), so this document would be nearly invisible if it was not published on PsyDok. This difference is of course very distinct with diploma theses, since these are mostly only held at the library of local psychological institutes. But the usage of electronic vs. print versions is also amazingly different with documents that can be borrowed. For example the report "Professioneller Wandel in der Medizin als Herausforderung für die psychologische Eignungsdiagnostik" by Hans Uwe Hohner is available at three university libraries in Germany. One of them is the Saarland University and State Library with its Special Subject Collection Psychology, which registers most of the queries, interlibrary loans and requests for document delivery for psychological literature in Germany. The report mentioned is held by SULB since 1998 and was not borrowed one single time since February 2004[36]; but during October 2005 it was viewed online 166 times. At the end of 2004 the burden on the server became so heavy that PsyDok had to be transferred to another server – another evidence for visibility and dissemination.

4 What's to come?

Beside the integration into more retrieval systems further enhancements are under development. First to mention are the implementations of a print-on-demand service[37] and a Creative Commons[38] licence management. Furthermore the APA's PsycINFO Classification Categories and Codes will be used as a browsing structure. The new multilingual layout developed by SULB will bring another improvement. Internationalisation will also be enhanced by the implementation of a cross-lingual search option that is based on the system CANAL.

To offer additional information about the quality of a document (which is often derived from its usage) PsyDok already offers anonymized statistics giving evidence of document related downloads without showing any visitor's IP-address or displaying the reader's identity. Admittedly these statistics are still quite noisy, but they give prove to the fluctuation and dissemination of digital publications and confirm the benefits of electronic publishing. Discussing the pros and cons of e-publishing there are mainly two arguments used to describe e-publishing as mediocre compared to conventional print-publishing with traditional publishing houses:

- a lack of citation counts and citation linking
- a lack of reliable, public statistics providing information of a document's usage

Obviously both arguments aim at issues of quality by inherently denying its existence on digital repositories or at least by denying any procedure of quality measurement. There are already existing models and technical answers to the first argument, especially the already mentioned CiteSeer, Citebase[39], OpenCit[40] and Google Scholar with its citation linking. These technical solutions are based on a distributed network approach: they can not be developed by single operators or administrators of digital repositories. Anonymized download statistics alleviate the second problem. Statistics like these shed light on the significant higher reading rate and the significant higher and faster dissemination of electronic documents compared to print documents - especially if these documents are systematically exposed to scientific and to lay-oriented retrieval systems. Besides: in fact it is proved that higher download rates correlate with higher citations.[41] Of course many technical problems have to be considered and to be solved, until digital repositories will reach a level of standardisation that allows and offers really reliable access statistics per document comparable or equivalent to the Counter Standard.[42] One of the most promising initiatives in this field is Interoperable Repository Statistics (IRS).[43] IRS plans to develop tools, which remedy the deficiencies of incompatible data that is hard to interpret, compare and aggregate (IRS Homepage 2005). Another startling approach is the analysis of usage statistics in order to compare journal usage and citation impact factor (Bollen 2005).[44] After all current endeavours tend to mould quality measures by examining correlations between citation counting of electronic documents and usage statistics and thus combining both strategies.

Literature

- Bollen, J., Van de Sompel, H., Smith, J., Luce, J. (2005), Toward alternative metrics of journal impact: a comparison of download and citation data. Information Processing and Management, 41(6):1419-1440

- Deutsche Forschungsgemeinschaft (2005), Publikationsstrategien im Wandel? Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung von Open Access, Bonn, Germany.

- Open Society Institute (2003), A guide to institutional repository software, 3rd edition. http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf

- Rodriguez M. A., Bollen J., Van de Sompel H. (2005), The Convergence of Digital-Libraries and the Peer-Review Process. A preprint forthcoming from the Journal of Information Science. http://arxiv.org/abs/cs.DL/0504084

- Interoperable Repository Statistics Homepage, http://irs.eprints.org/

Author Information

Ulrich Herb studied Sociology at Saarland University in Germany. He is employed at Saarland University and State Library, which is not only the local University and State Library for Saarland but also the special subject collection library for the discipline Psychology in Germany.
He headed different projects dealing with free of charge electronic scientific information in the field of Psychology like the virtual psychology library and Digital Psychology Information DPI. In addition to the above mentioned projects he engaged in a number of initiatives dealing with the aspects of nationwide information supply both disciplinary and interdisciplinary. Today his major tasks are digital publishing in general, development and management of electronic publishing services (regional at Saarland University and supra-regional for the psychologist's community in Germany), alternative publishing models and Open Access. Email: u.herb@sulb.uni-saarland.de

References

[1] For more information about these special subject collections, see http://webis.sub.uni-hamburg.de/
[2] For more information about DFG, see http://www.dfg.de
[3] For more information about Saarland University and State Library, see http://www.sulb.uni-saarland.de
[4] In a nutshell subject gateways are human controlled and quality controlled (not robot based) search engines or web catalogues containing links and metadata to scientific information resources in the www. Due to the ephemerality of most scientific websites the maintenance of subject gateways turned out to be quite time-consuming.
[5] ZPID creates electronic databases on literature, tests, audiovisual media, and Web resources. For more information, see http://www.zpid.de
[6] For more information about PsyDok, see http://psydok.sulb.uni-saarland.de
[7] The study examined both publication and reception preferences of German scientists from all disciplines and it pointed out that especially members of the social science and humanities community rely on grey literature.
[8] For more information about OPUS, see http://opus.uni-stuttgart.de
[9] For more information on PDF/A, see http://www.aiim.org/pdf_a/
[10] For more information about the URN, see http://www.persistent-identifier.de/?lang=en
[11] For more information about DDB, see http://www.ddb.de
[12] For more information about DOI, see http://www.doi.org/
[13] For more information about PURL, see http://www.purl.org
[14] An URN resolving mechanism passes the URN on the actual URL of a specific registered document. An abstract ID – for instance an URN as urn:nbn:de:bsz:291-psydok-5541 – is assigned to a concrete object (the conference paper "Human 'sense of space', moving images and architecture" by Rainer Schönhammer).
[15] For more information about nestor, see http://www.langzeitarchivierung.de/index.php?newlang=eng
[16] For more information about kopal, see http://kopal.langzeitarchivierung.de/
[17] Therefore SULB reserves the right to decide which documents will be published and which will not be.
[18] For more information about the Open Archives Initiative, see http://www.openarchives.org
[19] For more information about DINI, see http://www.dini.de
[20] Currently there are thirteen servers certified by DINI, a list of them can be found at http://www.dini.de/dini/zertifikat/zertifiziert.php. All in all there are about one hundred institutional repositories in Germany: http://www.dini.de/dini/wisspub/dokuserver.php
[21] For more information about KVK, see http://www.ubka.uni-karlsruhe.de/hylib/en/kvk.html
[22] For more information about OASE, see http://www.ubka.uni-karlsruhe.de/kvk/kvvk/kvvk_en.html
[23] This search interface can be found at http://opus.uni-stuttgart.de/opus/gemeinsame_suche.php
[24] For more information about SCIRUS, see http://www.scirus.com/srsapp/
[25] For more information about Google Scholar, see http://scholar.google.com
[26] PsychLinker is a scientific catalogue to psychological online information which is jointly run by ZPID and SULB, broad categories within PsychLinker are advised by members of the psychological community. For more information, see http://www.psychlinker.de
[27] Psyndex is run by ZPID, so ZPID members select documents to be added Psyndex.
[28] For more information about PsychSpider, see http://www.psychspider.de
[29] This online catalogue can be found at http://opac.sulb.uni-saarland.de
[30] For more information about the Collection of Computer Science Bibliographies, see http://liinwww.ira.uka.de/bibliography
[31] For more information about CiteSeer, see http://citeseer.ist.psu.edu
[32] For more information about SOSIG, see http://www.sosig.ac.uk
[33] For more information about PsychCrawler, see http://www.psychcrawler.com
[34] For more information about APA, see http://www.apa.org
[35] For more information about PsycEXTRA, see http://www.apa.org/psycextra/
[36] Statistics of former years are not available
[37] For more information about the service, see http://www.proprint-service.de/
[38] For more information about Creative Commons, see http://www.creativecommons.org
[39] For more information about Citebase, see http://citebase.eprints.org/cgi-bin/search
[40] For more information about OpenCit, see http://opcit.eprints.org
[41] You may read more about it at the OpenCit web site: http://opcit.eprints.org/oacitation-biblio.html
[42] For more information about Project COUNTER, see http://www.projectcounter.org
[43] IRS is funded by JISC since June 2005. For more information, see http://irs.eprints.org/
[44] This model uses data mining techniques in various ways: besides offering quality measures, tools like a usage-based recommender system could be offered. More details can be found within the slides of Johan Bollen's lecture "A framework for assessing impact of units of scholarly communication based on OAI-PMH harvesting of usage information" at the OAI4 workshop, http://oai4.web.cern.ch/OAI4/

LARA - Open access to scientific and technical reports

Christiane Stock and Emmanuelle Rocklin
Institut de l'Information Scientifique et Technique, INIST
Centre National de Recherche Scientifique, CNRS France

Aurélie Cordier, LORIA-SITE and Nancy-2 University, France

Abstract
INIST has a longstanding practise of collecting and referencing grey reports. Its collection of 60 000 government funded reports or institutional research papers is made available through traditional document supply.
Today reports are a minor issue in French repository projects. That's why INIST decided to create a national repository for grey reports: LARA - Libre Accès aux RApports scientifiques et techniques. Its aim is to allow direct and open access to the documents through the internet and to increase their visibility for the scientific community.
The project includes the identification of the report producers, negotiations and finally written agreements. It covers retrospective deposit of native electronic reports, the current production, but may also include the digitization of older print documents.
INIST will use the DSpace platform as basis for the repository. This choice implies the use of qualified Dublin Core metadata, harvesting facilities according to the OAI-MH protocol, and monitoring of formats (in the context of perennial access).
Besides contacts with the organizations, other points currently addressed in the project are technical issues like the customization of the platform and the creation of appropriate workflows. Legal aspects (contracts, licenses, etc.) concern not only the authors, but also their institutions and the end-user. French law and practice holds some particularities in this field.
LARA must be seen in the framework of the engagement of CNRS in the open access movement (CNRS signed the Berlin declaration in 2003). Together with OpenSIGLE, the free access project for the SIGLE database, it will be integrated in a new European project based on the use of meta-search-engines.

The French context:
As the national representative for the former SIGLE database, INIST-CNRS has a long-standing tradition in acquiring grey literature in general and reports in particular. Our current report collection holds more than 60 000 grey documents in paper copies. Some of the final reports belong to the dark grey category: they exist only in 2 or 3 copies. INIST receives a copy through agreements with the funding ministries or as ex-participants in our projects. Another form of dissemination used by organizations consists in the distribution of copies to a limited list of organizations has been more and more replaced by electronic versions accessible on the Internet of the producer. Reports are made available to users/customers through traditional document supply channels.

LIGRIA (for Administrative Grey Literature) is the name of a national experimentation in the nineties whose aim was to increase collecting of funded reports hidden in offices of the ministries and to make them available to the public. This experimentation ended in 1998. As an outcome of this action, la Documentation Française created a repository called Library of public reports or Bibliothèque des Rapports publics. It contains reports issued or funded by governmental bodies and dealing with administrative topics (Thiolon, 1997). In the beginning, only the records and abstracts were freely accessible and the download of the full text had to be paid for. Soon afterwards, all documents were proposed in full text and free of charge to the user. A similar offer for scientific and technical subjects was missing up to now.

Another project involving major research organizations in the nineties called Griseli (for Grey Literature) was more ambitious, but less successful. Reports were to be created in electronic form, then collected, indexed and disseminated by Griseli centres representing each a specific domain of research or technology. The standard SGML ISO 12083 was to be used for the bibliographic records and later on for the documents themselves (Comberousse, 1996). Only a small number of the centres went through with all aspects of the initial project, while others implemented parts of it.

Among traditional grey document types doctoral theses become more and more frequently published on the Internet, thanks to academic and governmental initiatives (e.g. Pastel, TEL, Cyther, Cyberthèses) (Paillassard, 2005). Other archives with grey literature may be discipline oriented like Mathdoc in mathematics. (MathDoc, 2005)

Today standards like the protocol OAI-PMH or technical tools to manage the deposit and dissemination of electronic documents (e-prints, DSpace, CDSWare) greatly facilitate a better access to electronic grey literature.
On the political level CNRS (Centre National de la Recherche Scientifique = French National Research Organisation) supports the creation of institutional archives and other measures in favour of open access. They were among the first French organizations to sign the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" (http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html) in 2003.

A favourable political and technological context together with a lack of visibility and accessibility of reports lead us to join the group of repositories by creating LARA, an interdisciplinary and inter-institutional repository for reports, allowing different search methods, free access to the full text and being OAI-PMH compliant as well.

Reports - hidden in the depths of the web
It is true that reports are available in full text on the institutional websites, but more often than not they lack visibility. Like the paper copies in the past, they may be widely disseminated and promoted for a short time, and disappear soon afterwards. Electronic reports will appear in the "News" category on the homepage of a given institution, and be well visible for all for a limited period of time, then be banished to the depths of the web. We observed that number of reports can be found available as full text, but classed in the archives, in the electronic resources accessible through the library, on personal pages etc. On the website of a research organization they may appear in the list of publications, sometimes presented by year, by research theme or by document type. These lists of publication may be global or presented separately for each research unit.
Frequently these documents are only enumerated in a list with a link to the full text. If they are searchable at all, it is a one case search on the institutional website or the corresponding webpage.
A search through Google or Yahoo may allow access to the full text through library catalogues or through Newsletter publications because the Newsletter or the library catalogue is indexed.

Reports - the unknown document type
Reports as a particular document type are often difficult to define, even for information professionals, especially if the research results in their fields are mainly published through articles or books. For some disciplines, research reports or technical reports are a long-standing form of publication and dissemination of their results, even if occasionally documents declared as reports are hidden "preprints". Among the communities, which are familiar with reports, we can cite computer sciences, physics, engineering, etc.
Governmental bodies or ministries are very familiar with two types of reports; final reports presenting the results of funded research, and "study reports" written by a single person at the express request of a ministry on a precise topic.
Moreover, many organizations, commercial institutions as well as research laboratories, write their annual reports or activity reports.
However, when we present LARA to a larger community, we are often asked the question: what do you understand by report? Is it the same as a preprint?*

A part of the "visibility" of reports is related to naming them explicitly as a particular document type. Let's have a look at the websites and existing archives: if documents are presented by document type at all, you seldom find reports as a distinct category. They may be declared as preprints, articles, or other, as we observed recently in an otherwise well structured open archive. Although this practice isn't harmful for the scientific contents or the accessibility, it doesn't increase the visibility of this particular type of document either. Less effort would be necessary to explain what reports are if they were found more often on the Internet as such.

Frequently, reports are just cited as "other publications" in the research environment. Several sites in computer sciences list internal reports, research reports and pre-prints in one category. A check on documents issued 5 years ago shows that only a minor part has been submitted to or accepted by journals. Reports as a distinct category have a better recognition on the websites of funding organizations, ministries, local government bodies and in the applied technology field….(cf. BRP, 2005)

In the context of scientific publishing and evaluation of researchers –and probably for some scientific domains only- reports seem to be the "parent pauvre" (poor relative) of the publications. They are useful to obtain funding for the laboratory, but they are not taken into consideration for the evaluation of the

* The same question is asked when we try to obtain very deep grey conference proceedings: « Proceedings? What is this? Do you mean the papers?"

searchers, nor are they peer-reviewed. So some units are reluctant to include reports in their publication lists on the Internet or even to make them publicly available. They may be mentioned though in the annual report of the institute. Legal aspects (in particular author's rights and copyright issues) are another excuse for not making an effort for the publication of a final report: we wrote the document, but we aren't authorized to disseminate it. Please contact the funding organization.

Of course these problems linked to the scientific evaluation process don't apply to organizations, which are mainly interested in the solutions to a specific problem proposed in the report and of course the funding organizations.

A need to increase visibility by added values and functionalities
Besides the terminological issue, ergonomic aspects of accessing reports on the Internet are in disfavour of their visibility. As mentioned before, numerous sites present their publications, or even reports as a list of titles and authors, sorted by year or classed by subject. Sometimes a fuller bibliographic record may be available, and even the full text through a link.

Search features are often limited to one field search, looking for the words in the record, if it exists. Browsing facilities are less frequent, the possibility to limit search even less. The most frequent criteria are by date of publication, name of the author or the title in alphabetical order.
The stability of the web addresses (URL's) for the documents has improved over the years, but there is still a danger when the institutional website is redesigned.

There's another aspect reports can have in common with conference proceedings, when they don't have a specific home on the website: they may appear in a prominent place on the institutional homepage for a short time (News, actualités), and disappear into the depths of the web after some days or weeks. We observed this practice on the websites of some ministries, for example. Other report producers or publishers propose extended research facilities for the current year production, whereas the preceding years are classed in so-called archives with simplified research facilities or none at all. ADIT proposes a research interface for the two most recent years (ADIT, 2005). CERTU puts forward the study reports of the current year on a specific webpage with records and links to the document. The older ones are grouped with other document types under title or subject headings (CERTU, 2005). The same observation is often made for conference proceedings.

So there is still a lot of work to be done to increase the visibility and the value respectability reconnaissance of reports, not the least of it a campaign to convince the producers of the value of their documents. Worldwide visibility of reports through the Internet might even be considered as a disadvantage to some producers, for not belonging to peer-reviewed categories. Governmental bodies may be conscious of the loss of information and money if reports are not disseminated, but the creation or maintenance of appropriate structures can be difficult for them.

A citation analysis presented at GL6 revealed: "Reports, which held second place at the first conference (GL'93) dropped to third place by the second conference (GL'95) and to fourth place by the third conference (GL'97). These had traditionally been the mainstay of GL, but this research shows that conference papers have taken and hold second place." (Schöpfel et al., 2005)

LARA on DSpace
In 2004 several platforms for archives/repositories were tested and evaluated at INIST-CNRS and decided to adopt Dspace for three applications: irevues (http://irevues.inist.fr/), Mémoire de l'INIST and LARA.

The Massachusetts Institute of Technology (MIT) and Hewlett Packard developed Dspace for institutional repositories. It is an Open source software system still in development.

Structure of the archive:
While many repositories adopt a structure similar to the organisation or the university which sets them up, INIST opted for a discipline-defined structure for LARA.
As a matter of fact, DSpace proposes the organization of the documents by communities, sub-communities and collections. Since LARA is a multi-institutional and multidisciplinary repository, subject categories that are already used for other INIST-CNRS portals like BiblioSHS (http://biblioshs.inist.fr/) or BiblioVie (http://bibliovie.inist.fr/) were adopted.

Three basic divisions of sciences (Exact and Applied Sciences, Humanities and Society Sciences, Life and Health Sciences) are followed by more detailed subdivisions, which contain precise scientific domains. For example, in the Exact and Applied Sciences community you will find the sub-community Engineering

Sciences that includes the Computer Sciences collection where all reports about this domain are available:

Sciences de l'Homme et de la Société
Sciences de la Vie et Santé
Sciences Exactes et Sciences Appliquées
Sciences chimiques
Sciences physiques et mathématiques
→ ***Sciences de l'ingénieur***
Automatique, recherche opérationnelle
Aéronautique, astronautique
→ ***Informatique***
…

Functionalities :
There are three web-user interfaces in LARA: one for the searching and consultation of reports, another one for the deposit of documents and the last one for the system administration.
The first web-user interface allows you to search and browse either on a general subject level or on a specific one. You can browse by date, subject, author or scientific domain, while the searching is operating on titles, authors, sponsors, subjects, handles and abstracts.
It is possible to authorize every end user to search, display and download documents, or to allow restricted access to a given collection or item of a list of registered users only. Since LARA is in open access, and confidential reports will not enter it, these possibilities were not fully explored. Therefore, you do not have to be registered to retrieve a report from LARA: item metadata and associated bitstreams are all available to our end-users.
Meanwhile, being registered will offer you some interesting functionalities, as for examples the Alerts that advise you by e-mail for each new item integrated in your favourite collections.
However, it is mandatory to be registered to submit an item or to intervene in the workflow process.
If you want to deposit reports in LARA, the administrators will grant you necessary rights to fill-in the submission form and to download the appropriate bitstreams.

Metadata:
DSpace uses a qualified Dublin Core metadata set based on the DC-Library Application Profile with 66 possible fields (Smith 2003). While the current metadata-set proposed by DSpace is sufficient for current documents, it is not so easy to add fields for specific purposes (e.g. dissertations). And of course, it's far less detailed than a MARC format or a detailed XML scheme. While other repositories working with Dspace have added a number of fields (eg. IRIS, the successor of Grisemine in Lille: see Claerebout, 2004; for ERA, see Jones, 2004) we decided for LARA to stay close to the proposed dataset, in order to take full advantage of the interoperability with other systems. As we mentioned before, DSpace is OAI-PMH compliant.
Only three fields are required for the report to be integrated: title, language, and submission date. Of course, we strongly recommend to create a bibliographic notice as complete as possible, in order to obtain the optimum visibility for the document. Note that each item (e.g. report) can be composed of one or multiple files.
Once the submission is complete, it will be given a persistent identifier working as an URI. To this end, DSpace has chosen CNRI handles (see Smith, 2003).

Workflow:
At this time, the item is still not integrated in LARA: the administrator of the collection must review, maybe edit and finally approve it before it will be available to all end-users. An additional functionality is that, while a specific collection is chosen for the deposit, the administrator can link the item to different other collections if it recovers several scientific domains.

In the system administrator interface, DSpace allows the administrator, the librarian, to specify the access rights not only for collections, but even on a lower level. This can be done directly without any help from computer specialists. The same interface allows the administrator to specify the steps of the workflow and to assign roles: at which stage emails are sent out for notification, which e-people give the final validation of the document for the repository. It is easy to build a network with distributed administrators for different subjects or assign control at the report producer's level.
For the beginning of LARA, the same workflow will be applied for all the archive. Submitters will not have the rights to administrate a collection or community, because different institutions working on a same scientific domain will submit in the same collections. It would be delicate for a laboratory to validate the reports of another one.

Import:
The submission through the web user interface is the regular way to deposit a document in LARA. The author, or a local administrator, librarian or other e-people can do it.

However, INIST intends to add documents and metadata from important report producers through import routines.
Since the use of import routines implies the necessity to write conversion programmes to conform to the DSpace metadata scheme, and other scripts, several conditions should be fulfilled to justify the initial investment. The number of documents to be uploaded into DSpace from the same source should be more than 50-100 items; metadata should be consistent through all records, well structured and rich in details. Although the initial preparation for conversion /import can be time consuming, it allows us to enhance the metadata by inserting default values which are not explicitly mentioned in the source record, for example add 'research reports' as a document type, or the name of the organization.

A new look for LARA:
Another part of the project was to customize the user interface.
LARA on DSpace is our first experience with a software development; the project group had the chance to have the participation of a student in information science who worked for 3 months on the user interface and other texts. We had to take two basic points into consideration. On the one hand, LARA is in the first place an archive dedicated to French reports and facing French-speaking users. On the other hand, it is a CNRS website.

The fact that LARA is an archive dedicated to French reports lead us to translate the interface into French. It concerns the administrator's and user's interface, including the submission interface and specific help screens. We also wrote an online guide into French. Although the current DSpace version we use (1.2.2) is stable and operational, surprises can occur. A new version was loaded when we worked on the translation of the user interface, and we lost some data.
Otherwise, the CNRS imposes a graphic charter mandatory for all CNRS websites. This means that we were limited in the choice of fonts and colours as well as for the screen-width for example.

DSpace archives and repositories can be easily identified by their look, if the default software has been used. Several universities did some work to give it an individual look or to provide the user with additional features. Some remarkable sites are

- Papadocs: http://papadocs.dsi.uminho.pt:8080/index.jsp
- TSpace Toronto, https://tspace.library.utoronto.ca/
- Papyrus https://papyrus.bib.umontreal.ca/dspace/

Drawing our inspiration from these models and considering the CNRS graphic chart, we tried – with the help of our in-house web manager – to make LARA welcoming and functional for our end-users.

Copyright and licence issues
The DSpace submission interface includes a copyright licence agreement that must be accepted by the author. Otherwise it is impossible to upload the document. The default text proposed by DSpace was created by MIT. This text will be customized for the individual deposit through the Web User Interface for LARA. Many other sites have replaced the MIT version by their own licenses. DSpace now offers links to Creative Commons options.

Our priority is to conclude agreements with the institutions, not with individual authors (except for particular cases). For important report producers, we intend to cover all legal aspects in a contract. The supplying organization should assure that the authors have agreed to the dissemination of their report through LARA. The contract form allows us to add other services in the future like the digitization of the paper version of reports or to detail the scope of the cooperation such as subject or temporal coverage.

Our project benefits greatly from institutions, which already have implemented/prepared policies with regards to repositories or archives, either by making reports available through the Internet or by adding clauses in the funding contracts. There are far more hesitations in organisations which discover the world of open access. For research units with only a small number of documents we have prepared 2 texts for agreements: one to be approved by the author and one by the holder of the document. This array of legal texts will be completed by a text for the general use of the archive/repository. All these texts have been prepared with our in-house legal specialist.

Digitization
So far we have dealt with native electronic reports. The LARA project also includes an aspect concerning the retrospective digitization of our paper collection. Different questions arise in this context. In concert with the issuing organizations or the funding ministries we have to decide which subjects should be

chosen for the digitization and how far we will go back in time. These questions cannot be answered in a global way, since some research subjects may be of very short-lived interest. We will discuss them on an individual basis with each organization.

On the technical side it is important to find the ideal result between the quality of the image and the file size. Reports can be quite important in number of pages. Calculations from our existing collection show an average number of 70 pages per report, and the higher the resolution, the more important the file size. Even with growing performances of PC's and Internet connections, recommended sizes for the exchange of documents through e-mail are 4 megabytes. Several tests with in-house machines showed us good results with a digitization in black and white and a 300 dpi resolution.
It may be necessary to include some colour images if our partner insists, but then the file size increases quickly.
OCR can be an option, but the majority of documents will be in image format only.
With DSpace it is possible to cut a bulky document into several parts in order to maintain a decent file size for each part. This is also practical for annexes, if their volume is important.

Perspectives: preservation and outlook
At present LARA is still a prototype, but we already know that it will change further – as do electronic documents.
Almost from the beginning DSpace has taken the digital preservation of documents into consideration for the software (Smith et al., 2003).
Minimal technical metadata to allow a digital bit preservation were included (file format, MD5 checksum and creation date). During the deposit workflow a list informs the authors which file formats are supported by the archive (i.e. which of them will be migrated to newer versions so that they will always be accessible). DSpace also works on the METS* scheme.

The choice of handles as unique identifiers for the items is another towards access preservation. Handles remain valid even if the platform moves to a new hardware environment. According to a study at Rochester University, the preservation issue can be a valuable argument to convince researchers and interest them in online repositories/libraries (Foster, 2005).

In the context of LARA some other developments announced by DSpace are of particular interest: a greater flexibility for the management of metadata and a multilingual interface which will allow us to propose a French and an English version of the user interface.

LARA will contribute to improve the visibility and accessibility of French reports and grey literature in general. The repository will hopefully join other French and European archives and repositories in a European project based on the use of meta-search-engines: MetaGrey Europe (Schöpfel, 2005)

Author Information

Christiane Stock graduated from the University of Freiburg in 1984. She joined INIST-CNRS the French Institute of Scientific and Technical Information in 1989. Member of the Technical Committee for the SIGLE database since 1993, she also set up the national agency for the ISRN (International Standard Report Number). Today she is the head of the monographs and grey literature section at INIST.
Email: christiane.stock@inist.fr

* METS : Metadata Encoding & Transmission Standard

References

1. ADIT Agence pour la Diffusion de l'Information Technologique – homepage (2005): http://www.adit.fr
2. Bibliothèque des rapports publics BRP (2005), [Webpage presenting the service] http://www.ladocumentationfrancaise.fr/rapports/presentation.shtml (accessed November 2005)
3. Claerebout, M.-F. (2004), Grisemine, a digital library of grey university literature. –In: GL5 Conference Proceedings: The Fifth International Conference on Grey Literature: "Grey Matters in the World of Networked Information", 4-5 December 2003. – Amsterdam : GreyNet/TextRelease, 2004, pp.27-31. - (GL-conference series, ISSN 1385-2316 ; No. 6). ISBN 90-77484-01-9
4. Comberousse, M. (1996), Grey Literature in France : the GRISELI Program. - In: GL'95 Conference Proceedings: The Second International Conference on Grey Literature "Grey Exploitations in the 21st Century" held in Washington D.C., 2-3 November 1995. – Amsterdam : GreyNet/TransAtlantic, 1996, pp. 126-132. - (GL-conference series, ISSN 1385-2316; No. 2). - ISBN 90-74854-08-7
5. CERTU Centre d'Etudes sur les réseaux, les transports, l'urbanisme et les constructions publiques (2005), [catalogue of publications webpage] - http://www1.certu.fr/publications/ (accessed November 2005)
6. DSpace, a digital repository system, [homepage] (2005): http://www.dspace.org/ (accessed November 2005)
7. Foster, N. F. and S. Gibbons (2005), Understanding Faculty to Improve Content Recruitment for Institutional Repositories. - In: D-Lib Magazine, v.11 :no.1, January 2005. - ISSN 1082-9873 - http://www.dlib.org/dlib/january05/foster/01foster.html (accessed November 2005)
8. Jones, R. (2004), The Tapir: Adding E-Theses Functionality to DSpace. – In: Ariadne. Issue 41, October 2004. - http://www.ariadne.ac.uk/issue41/jones/intro.html (accessed November 2005)
9. I-Revues [homepage]: http://irevues.inist.fr/ (accessed in November 2005)
10. MathDoc: French network for documentation in mathematics run by the University of Grenoble-1 and the CNRS: http://math-doc.ujf-grenoble.fr
11. Paillassard, P., Schöpfel, J. and C. Stock (2005) How to get a French doctoral thesis, especially when you aren't French. - In: GL6 Conference Proceedings: The Sixth International Conference on Grey Literature "Work on Grey in Progress" 6-7 December 2004. – Amsterdam : GreyNet/TextRelease, 2005, pp. 143-151. - (GL-conference series, ISSN 1385-2316; No. 6). ISBN 90-77484-04-3
12. Papyrus : Dépôt institutionnel numérique de l'Université de Montréal, [homepage] (2005), https://papyrus.bib.umontreal.ca/dspace/ (accessed November 2005)
13. RepositóriUM - repositório institucional da Universidade do Minho (2005) [homepage, in Portugese]: https://repositorium.sdum.uminho.pt/index.jsp
14. Schöpfel, J., Stock, C. Farace, D.J. and J. Frantzen (2005) Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit. - In: GL6 Conference Proceedings: The Sixth International Conference on Grey Literature "Work on Grey in Progress" 6-7 December 2004. – Amsterdam : GreyNet/TextRelease, 2005, pp. 55-63. - (GL-conference series, ISSN 1385-2316; No. 6). ISBN 90-77484-04-3
15. Schöpfel, J. (2006) MetaGrey Europe, a Proposal in the Aftermath of EAGLE-SIGLE. - In: GL7 Conference Proceedings : The Seventh International Conference on Grey Literature "Open Access to Grey Resources", Nancy, 5-6 December 2005
16. Smith, M., Bass, M., McClellan, G., Tansley, R. (et al) (2003), DSpace : An Open Source Dynamic Digital Repository. – In: D-Lib Magazine, v.9 :no.1, January 2003. - ISSN 1082-9873 http://www.dlib.org/dlib/january03/smith/01smith.html (accessed November 2005)
17. Thiolon, C. (1997) L'accès à la littérature grise d'origine administrative en France via Internet; rapports inédits pour non-initiés. – In : IDT'97 : salon de l'information électronique, Paris , 3-5 juin 1997. – Paris : ADBS, 1997, pp. 63-67

Find the piece that fits your puzzle ...

The Grey Literature Report

from The New York Academy of Medicine

Focused on health services research and selected public health topics, the Report delivers content from over 250 non-commercial publishers on a bi-monthly basis.

Report resources are selected and indexed by information professionals, and are searchable through the Academy Library's online catalog.

Let us help you put it all together; subscribe to the Grey Literature Report today!

For more information please visit our website:

www.greylit.org

Or contact us at:

GreyLitHelp@nyam.org

The New York Academy of Medicine

www.nyam.org

The accessibility to doctoral theses in Spain: A political change and a reconsidering of its nature

Isabel de Torres Ramírez
University of Granada, Spain

Mª. Dolores Ayuso-García
University of Murcia, Spain

INTRODUCTION

The term Grey Literature was cained in the XIX century in USA, thanks to the literature reports [Ramos de Carvalho, 2001]. In 1978, after the attempt to define Grey Literature (from now GL) identifying it with the non conventional one, the attending ones to the York Seminary –not without reticence- were forced to make an index of the documents that could be integrated in it. However, nowadays we have to take into account that communication and information technologies, (CIT), have contributed to the inclusion of a new typology of Grey Documents, that emerge basically with the development of Internet [Ayuso García, 2000] and are enriching and transforming the already consolidated typology of Grey Documents [Ayuso García, M.D. y Martínez Navarro, V., 2004a: 50-51).

In this context, we find the Doctoral Theses and the original researching works about this topic related with the scientific, technical or artistic field from the doctorate programme that the PhD student make, which are composed to obtain the maximum university degree. They are inserted between the publications in the scientific and academic world, besides to the preprints, congress certificates, reviews, offprint, discussion forum, notebook works, translations, curricula, etc. [MacKenzie Owen, J.S.]. A Doctoral Thesis can also be considered as a post print. We are referring to the fact that it's a finished editorial product, positively evaluated and previously printed, with interest to the scientific community, but generally without a big commercial value[1].

The selection of this GL typology for making the analysis for this paper justified itself because these documents, that have a big interest for the researching community, present big access difficulties. The reason for this difficulty is found in the fact that Doctoral Theses have not always been published and, when they appear printed in multiple copies, they do it often in restricted circulation and edition circuits, due to their specialised researching nature. That means that they are not distributed in the usual ways, and that they cannot be consulted easily, according to the grey documents conditions of the mentioned York meeting.

Without never arriving to be equal to the one of the reports, the attention that the scientific and documental community was giving to these academical works has always been high and its justified for it high scientific, informative –sometimes unique- methodological and bibliographical value. A consequence of this high interest is the repertories, which have been published by university, institutional or commercial initiatives in some countries very soon[2] and in others later in the XX century, like Spain.

This paper also shows the evolution that has taken place in the Spanish university policies regarding the conservation, diffusion and access to the defended theses in all the Spanish public or private universities and the multiple ongoing projects orientated to allow an easy access to this important scientific information resource. It also analyzes the positive aspects of the ongoing open access projects in our country that are similar to other experiences in other countries.

1. Considerations about the interest of GL in Spain.

Before we go deeper into the topic, it's convenient to revise the attention that GL has had in general in Spain. Seven years after the holding of York's Seminar, Alfredo Lara Guitard has written in the Revista Española de Documentación Científica: "un tema tan importante hoy día como el de la llamada 'Literatura gris' no ha sido aún objeto de la debida atención en las páginas de nuestra revista". Several years later, we can maintain this statement "continúa todavía vigente y no sólo en las páginas de la Revista mencionada, sino en el resto de las ya numerosas publicaciones que en

[1] We mention these points very often in our doctorate subject in the Universidad de Murcia, "Las nuevas formas de comunicación científica en la Sociedad del Conocimiento".
[2] In 1884 it appears in France the most important doctoral thesis catalogue that arrive to us in the data base *Telethèse: le Catalogue des thèses de doctorat soutenues devant les universités françaises*; the American doctoral thesis data bases keep information of the defended thesis in USA and Canada from 1881 (Cfr. El *Comprensive Dissertation Index* del ISI de Philadelphia).

nuestro país se ocupan de cuestiones relacionadas con nuestra área de conocimiento; hasta tal punto es esto así que, tras una búsqueda exhaustiva por fuentes nacionales y extranjeras no hemos encontrado más que dos o tres artículos escritos por autores españoles que traten más o menos directamente la mencionada cuestión" (Torres, 1994: 19).

This question has not really moved forward in the theory field, but a little bit in the practical one, because "el interés que, en general, existe hoy en el mundo entero por la LG, está igualmente presente, también de forma general, en España entre los miembros de la comunidad científica que se ocupan y preocupan por cuestiones relacionadas con la información y la documentación. Este interés dio sus frutos sobre todo en dos campos concretos: el de las Tesis Doctorales y el de la cooperación con EAGLE, como participantes en el proyecto SIGLE" (Torres, 1998: 350).

In spite of all this and in the field that we have just shown, it's not strange to see that the first national repertory of doctoral theses has not appeared until the 70´s, and that in 1994, in the VIII Jornadas Bibliotecarias of Andalucía, celebred in Huelva, a paper titled "Fuentes nacionales para el conocimiento de tesis españolas" was presented. This paper was a short reflection whose objective was to motivate people not to overlook the importance of letting people know and have free access to the excepcional information contained in the doctoral theses. (Cfr. Torres, 1994: 204). Luckily nowadays nearly every university have their theses repertories printed or/and electronically recorded, although this is an enormous task[3].

2. The situation of Doctoral Theses in Spain before the digitalisation.

Due to the intrinsic interest of doctoral theses, in Spain lately, the efforts of the academic authorities to make an exhaustive control of the ones that have already been defended in the public and private universities in our country have been multiplied. And also promoted by the users, the people in charge of the University Libraries have arrived to the conclusion that Doctoral Theses are an essential material that can not lack in their collections. It seems to be unavoidable to resolve access difficulties and solve problems related to bibliographie control.

Bibliographic control implies the store of the document, its description and inclusion in retrieval recuperation instruments: catalogues, repertories, databases. The access means its inclusion inside a stock where it can be looked up withoud being plagiarized, avoiding in this way the violation of the author rights. The digitalisation has solved many of the problems experienced by people who needed to consult previous defended theses.

2.1. The archive of Doctoral Theses in the Universities.

In Spain it doesn't exist a unitary legislation about the Theses archive. In the recent normative that regulates the official university postgraduate studies, in general and for all universities, it is only said that once the Doctoral Thesis is approved, the University will take care of its archive and send the appropriate and established index card to the Ministerio de Educación y Ciencia and to the Consejo de Coordinación Universitaria.

As it can be noticed, from the very moment of the defence of the Thesis, it's clear the worry for conserving and taking it into control, but it's not said nor how where they have to be archived, leaving this point to the universities. The most common thing is to give one copy to each member of the tribunal that has to evaluate the thesis, three copies to the doctorate Commission, another to the Faculty or High Technical School and another to the Department where the Thesis is going to be defended, where it's exposed to guarantee that the Doctors that are interested on it can read it, analyze it and, when this is the case, critic it and veto it before its public defence. One copy is stored in the Secretaría General and later it is sent to the University Library of the Faculty, where it's allowed to be consulted always with care, because a direct authorization is needed from the author to consult it outside the Library. At the moment, nearly all Spanish Universities establish that from the three copies that the doctoral candidate has to give to the University, one of them has to be in electronic format and also has to fill in the application form from the legal deposit and electronic ISBN. In this way, the author rights are protected since the very first moment and the diffusion of the electronic version of the Doctoral Theses is easier. One of the tasks of the Secretaría del Consejo de Universidades is to establish a card index of Doctoral Theses and publish a report of the ones that have a positive evaluation (Reglamento del Consejo de Coordinación Universitaria: Real Decreto 1504/2003, BOE nº 292 de 6 de Diciembre de 2003). The

[3] The universities of Murcia, Córdoba, Navarra and many more picked up the references of presented and approved doctoral thesis in a printed form, and nowadays in digital format, in a repertory way.

TESEO data base from the above mentioned Consejo collects and allows to retrieve information of Doctoral Theses considered suitable in the Spanish Universities since 1976.

2.2. Information retrieval tools.

In the Spanish State, the projects that include Doctoral Theses information retrieval tools have been slower than in other countries and have been started up in its origin thanks to the initiative of some universities that have tried to make available the catalogues or repertories of their own theses; in time these lists, thanks to the technologies application, have become Data Bases. Later on, it was seen the necessity of elaborated Doctoral Theses national bibliographies or collective catalogues, first printed and after automated in the whole country.

Leaving apart particular contributions[4] of each university, we can stste that in the Spanish State it exists a national information resource about Doctoral Theses defended in its public or private universities: the TESEO DataBase. This DataBase has been preceded by an attempt of published printed repertoire, promoted by the Ministerio de Educación y Ciencia[5]. TESEO was born in the 80s, with information since 1976, which is updated periodically and fed from the data sent by the Doctoral Commission of each university, by the already mentioned doctoral index card, which that is nothing but an official formulary with all the data related to the study done, including a big summary, the author must write explaining the methodology used and the principal conclusions of his/her Thesis. Without any doubt, TESEO is the most complete DataBase available in Spain. It is a strictly referential DataBase[6], having into consideration its character of general national repertoire; it's a multidisciplinary DataBase that uses the International Nomenclature for the fields of Science and Technology from UNESCO for the information organisation and systematisation. The data, just as it appears in the work that we use as reference to write this paper[7], offered in every registration are: title, author, academic year, University, reading place, execution centre, director, tribunal, descriptors and summary.

2.3. About availability and access questions.

It is not enough with the referential information, even if this is an analytic repertoire. To have a plenty access to the content of a document it is necessary that this document is available in a library and, if possible, with enough copies, which was something impossible until a recently. If we talk about Doctoral Theses that only "se podían consultar a partir de ediciones impresas o microfilmadas, lo cual implicaba toda una serie de obstáculos para su propagación y para que los resultados obtenidos tuviesen rápida incidencia en los ámbitos científicos, hoy todos estos problemas se han resuelto gracias a la variedad de formatos que permiten digitalizar un documento y a los distintos sistemas de acceso telemático mediante los que se pueden conseguir dichos documentos" [Sorli y Merlo, 2002: 95]. In any case we think there is still much to do, and many problems to solve.

Also in this case, the process followed in the Spanish Universities has been sinilar to the rest in the university world: putting at the disposal of the users the unique copies in paper and microfiches –with a difficult and limited consultation that make complex and nearly impossible the access of the information that these copies had- to try an availability without limits by means of digitalized. Which are the reasons that nowadays move the universities to get their theses digitalization? We are going to take as a starting point the reasons given by the Universidad of Granada to integrate itself in a theses digitalization process (Cfr. Universidad de Granada, 2005).

- Make available, its own or from other places, the scientific production contained in Doctoral Theses to postgraduate students and university teachers.
- Make profitable the economic, scientific and human investment that demand the execution of this kind of works and make known the researching results of departments, institutes and researching groups to everybody that get closer to the researching work in whatever knowledge field.
- Make public the scientific advance contribution of each university community and of the country itself.

[4] That is the case of the Universidad de Murcia since 1998 and also others. It can be seen as an example, the thesis repertory lists elaborated by different Spanish universities as shown in the work of Isabel de Torres (Torres, 1995: 305-307).
[5] Appeared from 1977 to 1981. Data base TESEO. Available in: http://www.mcu.es/TESEO/teseo.html. Consulted: 09,07,05.
[6] There is no comparison with international projects like Networked Digital Library of thesis and Dissertations, URL Thesis.org which promote the access to the complete texts of the doctoral thesis without commercial restrictions and that also have sometimes degree dissertations. El Dissertation.com.
[7] As a sub product from this national data base, the Secretaría del Consejo de Universidades published a printed repertory in three volumes with the information from the courses 1976-1977 and 1988-1989, that has been actualized in several occasions with supplements that have the defended thesis in the Spanish universities during the academicals years after 1989-19990. The title is: Tesis Doctorales 1970-1977/1988-1989. Madrid. MEC, 1990.

- Open a new prospective of diffusion, use, quotation and pursuit of the different existent researching lines that, due to be spread, will be useful to be a starting point to the new researching works of the postgraduate students that begin their Theses.
- Open a way to make that the author of the work and the University establish an agreement that being visible and available doesn't violate the author rights.

In this moment, as it can be seen, the tendency in the Spanish Universities is not only to put at the disposal of every network user resources that allow knowing the existence of their Doctoral Theses but to make them have the access to the complete text. That means an spectacular advance in less than a quarter of a century, as a result of university policies that have been success to in carrying out some important collective projects which we will comment now.

3. Doctoral theses digitalization projects in the Spanish universities.

Approximately, we see that a meaningful step forward is happening in our country with reference to the Digital Theses free access through the university strengthening of their theses digitalization. That's it, the scientific researching make good use of the facility and immediacy that the network offers to published researching works [Ayuso-García y Martínez Navarro, 2004b: 50] and, in general, to provide and generates eprint. However, it hasn't been an institutional initiative, as the TESEO DataBase, that promotes it and justifies itself due to the University autonomy. The truth is that this process has been carried out in different ways that we can synthesize in two groups: Doctoral Theses digitalization projects through cooperatives projects and projects from the Spanish Universities, and the case of the Biblioteca Miguel de Cervantes.

3.1. Cooperative Projects.

The most important cooperative projects that have taken place are:

The server "**Tesis Doctorales en Red**" (TDR). One of the first initiatives of doctoral theses digitalization that has taken place in our country by the Centro de Supercomputación de Cataluña (CESCA)[8], consortium created in 1991 and revised in their statutes in 2003. It is integrated by the Generalitat de Catalunya, the Fundació Catalana a la Recerca i la Innovació (ICRI), for several universities of Cataluña, as Universidad de Barcelona, Universidad Autónoma de Barcelona, the Universidad Politécnica of Catalunya. From the beginning the Consejo Superior de Investigaciones Científicas (CSIC) collaborates in this consortium. Its main objective is the administration of a great calculation and communication system that allows the university research and that is competitive in three concrete actions: the super computation, the communications and the technologies. The first one allows the calculation of high services in fields such as Chemistry, the creation of new materials, the Bioinformatics, the Astronomy, the weather forecast, etc. It also supplies the access to the information agents through Internet.

The functions that contribute to the communication field are also remarkable. It manages the"Anella Científica", the scientific ring of Cataluña, one of the first European high speed networks appeared in Europe in 1993, that connects researching technology centres and universities. This experience has been exported to several Comunidades Autónomas, although not always managed by an autonomous consortium, but sometimes by a regional administration centre. It has to do with the Internet infrastructure and specifically with the Catalan Iris network management. In the field of Information Technologies, it encourages the Information Society development and the university technologic innovation and cooperation.

It is here where the Servidor de Tesis Doctorales en Red (TDR) has its space. The Servidor de Tesis Doctorales en Red (TDR) born as the result of an agreement signed in 1999 between the Universitat Digital de Catalunya (1999-2003) and, at that time, Comisionados de la Sociedad de la Información de Universidades e investigación and the Universities of Cataluña: Autónoma de Barcelona, Politécnica de Catalunya, Universitat de Barcelona, Universitat Pompeu Fabra, Universitat de Lleida, Universitat de Girona, Universitat Rovira i Virgili and finally, the Universitat Oberta de Catalunya, besides the Fundació Catalana per a la Recerca, CESCA and the Consorci de Biblioteques Universitaries de Catalunya (CBUC). Later other universities of the so called "Arco Mediterráneo" have been incorporated: Universitat Jaime I (July 2002), Universitat de les Illes Balears (December 2002), Universitat de Valencia (April 2003),

[8] Centre de Supercomputació de Catalunya. CESCA. Available in Web document: http://www.cesca.es. Consultation: 28,07,05.

Universitat Ramon Llull (October 2004), Universidad de Murcia (July 2005) and also the Universidad de Cantabria.

The free access TDR server contains nowadays in digital format all the Theses that are composed and given directly in their electronic format to the Universities that belong to this project and that include, as we have said, the Universities of Cataluña, Valencia, Castellón, Islas Baleares, Murcia and Cantabria. At the moment, other universities are studying ascription protocols to the mentioned project. The TDR is managed by the Centro CESCA and the Consorcio de Bibliotecas Universitarias de Cataluña (CBUC) and promoted by the Generalitat de Catalunya in particular for the Departamento de Universitats, Recerca i Societat de la Informatió. Its objectives are specific and concrete enough to the purpose that wants to reach; generally, the free Doctoral Theses consultation without the necessity of a password or any similar process to log in the system:

- University research results diffusion trough the whole world by internet.
- Contribute to improve the Doctoral Theses bibliographic control.
- Strengthen and stimulate the research itself as well as its use.
- Support and develop the electronic edition at the same time that the new digital library model it's strengthened.

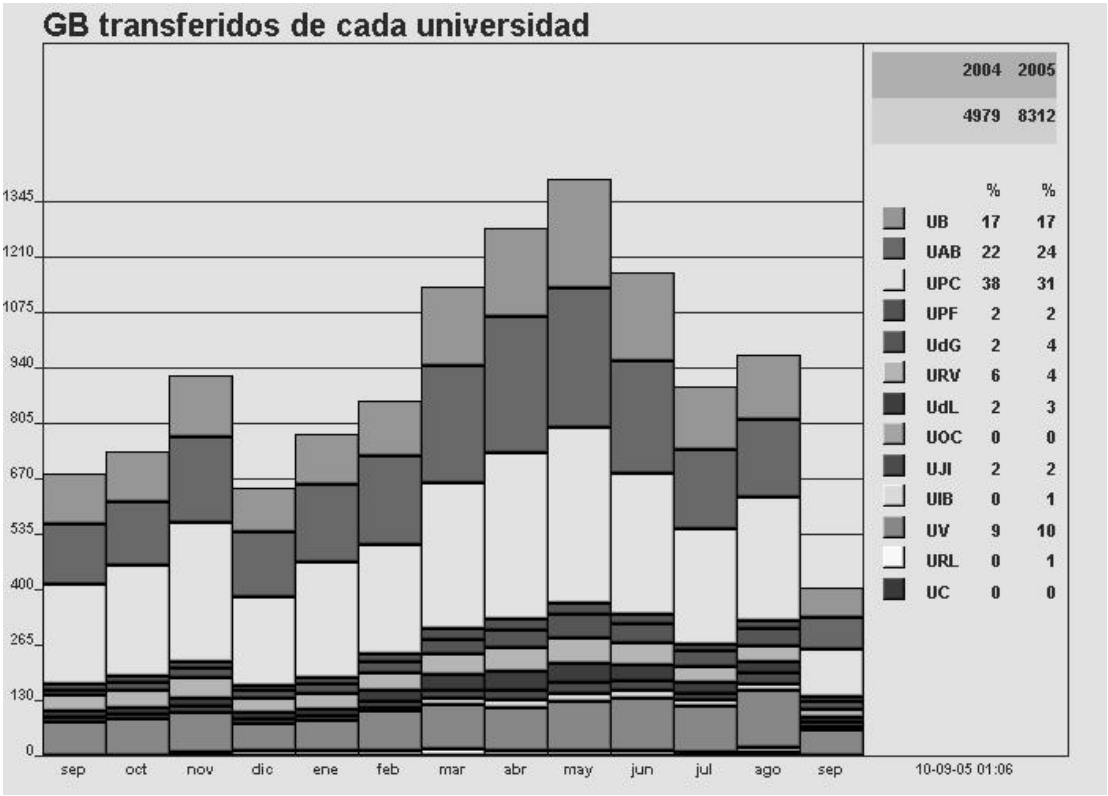
  Finally, another of its objectives is:
- Create a tool that for the authors and their works to be more visible and accessible.

The Theses author rights are protected by a contract that avoids possible irregularities regarding the violation of the author's intellectual property. The integrity of the text is guaranteed trough the security options that incorporates the storage format used: PDF. To visualize the whole text the Adobe Acrobat Reader programme is provided.

With respect to the Theses diffusion, the universities belonging to the project are the ones who diffuse them between the researching communities and, in particular, to the Doctorate students, to inform them, once the Theses are approved, about the allowed formats to their incorporation and edition in the TDR server. It's also finished the restructuring of digital format of Theses presented before the beginning of this initiative and that were archived in other formats.

As well as the open access to the whole Doctoral Theses text, the TDR server allows to know, trough the web statistics, the Theses incorporated by the universities, Theses by field and university, Theses consulted of each university, PDF card index and gigabytes transferred by each university, as we can see in Figure 1.

FIGURE 1



Information quantity, in gigabytes, transferred by Universities

The added value also allows knowing the percentage of the stock consultation for country/domain and Universities.

After Spain, Mexico is with the 10% the country that has a bigger consultation's number, followed by Peru with the 7 %.

The TDR through CESCA allows knowing the percentage of the historic evolution from the 30 more consulted university Theses between the years 2001 and 2004. It also allows knowing the percentage of evolution from the Theses' consultations done in each institution. The Universitat Politécnica de Catalunya is the one that makes more consultations, with a 40% from all of them, followed by the Autónoma de Barcelona and the Universidad de Barcelona, third in the ranking.

The server supplies the 25 consulted theses in 2005, indicating the title, number of consultations and the university where it was presented. If we have just seen that the Universidad Politécnica de Catalunya was the most consulted one in 2005, we also see in this year a wide variety of fields consulted. It's not a surprise that the most consulted Theses in 2005 where about environment, because of its actual projection, that shares the ranking with the research in fields such as education, information technologies and communication, engineering, sociology, architecture, documentation, etc.

The link with CESCA allows the server, through the Anexo IV, to supply the references of the publications done by the super computation projects, with generally difficult access, classified by author/year and publication. The TDR server belongs to the Networked Digital Library of Theses and Dissertations (NDLTD) and uses the protocol of interoperability Open Archives Initiative (OAI). OAI is a project of the Michigan University, specifically from the Production Service from the Digital Library with 5.704.392 registers and 510 collaborating institutions[9].

3.1.2. The Digital Theses in the Spanish Universities.

This Project is coordinated by the Red de Bibliotecas Universitarias (REBIUM)[10] that collaborates with the Conferencia de Rectores de las Universidades Españolas (CRUE). It's an ambitious project, which compiles all Doctoral Theses initiatives in Spain with a complete text. This project, as the one we have seen before, and collaborating with TDR verifies that the Spanish universities, on the whole –as well as the particular actions that we will see afterwards- go for the creation of systems to diffuse in OAI their Doctoral Theses and other academicals and scientific documents, which contribute to spread this typology of GL.

In this way, it motivates the creation of their own open archives or it recommends to use the TDR server to deposit them and even the e-prints server of the Universidad Complutense (now in testing phase).

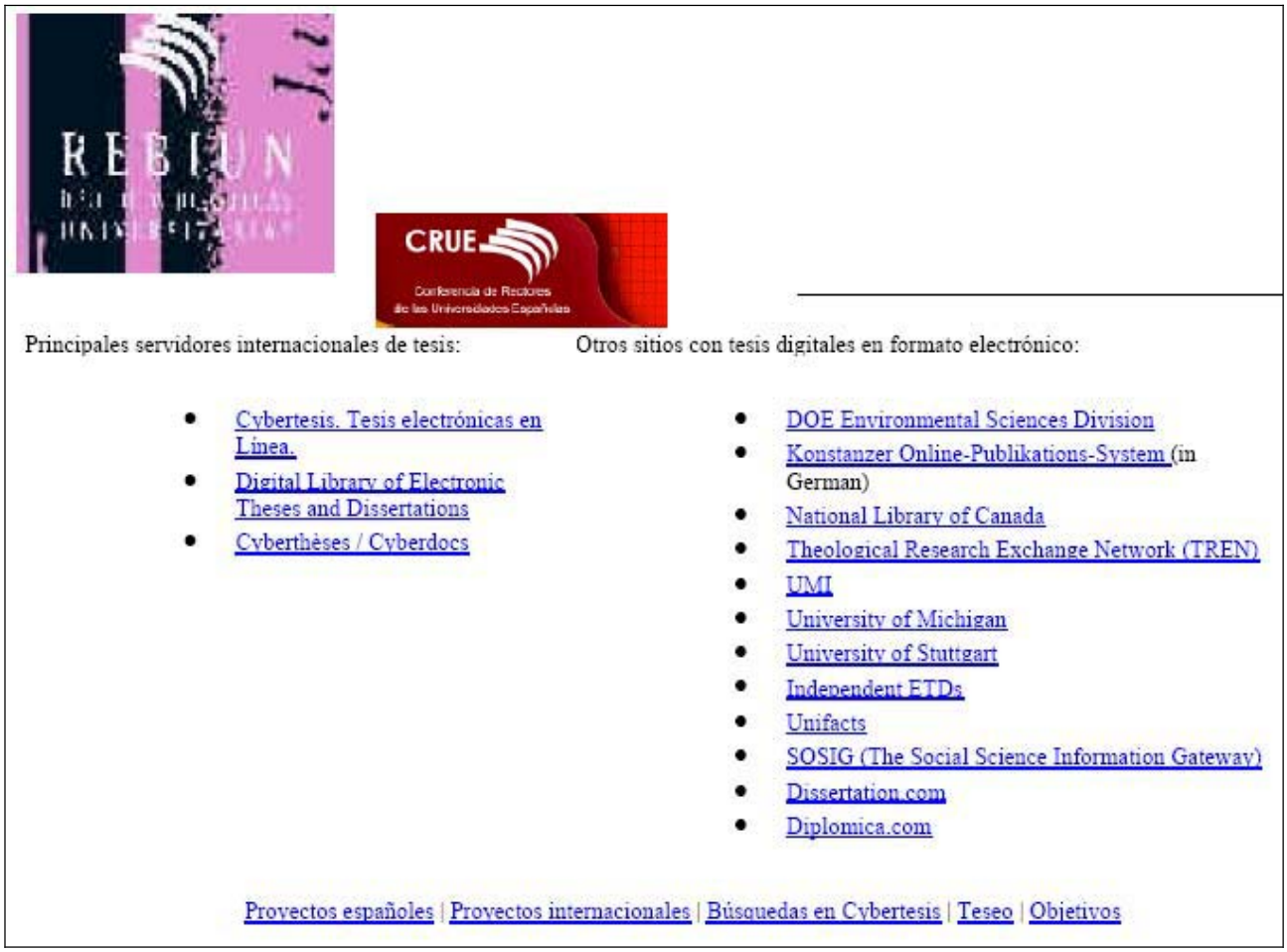
To retrieve the information it's used the online Doctoral Theses Cybertheses researcher, as well as to supply information about the TESEO Data base and the Electronic Theses and Dissertations Cybertheses/ Cyberdocs Digital Library.

Regarding the online Doctoral Theses international projects, those systems which help to spread the Theses and other academicals works information in OAI are chosen, knowing the access difficulty even with the efforts that are being made to solve it, but above all aware of the importance of these documents to the economic, researching and also social and industrial activity.

It is recommended, as well as the international servers that we have just mentioned, others in electronic format, some of them, very specialized, that help the user with the work and the information retrieval.

[9] OAIster. Available in the Web: http://oaister.umdl.umich.edu/o/aister. Consulted 28, 07, 05. The information given was actualized on 22 July 2005.
[10] Red de Bibliotecas Universitarias (REBIUM). Available in Web: http://www.ucm.es/BUCM/tesisdigitales/02.htm. Consulted: 28, 07, 05.

FIGURE 2



Access Web to Doctoral Theses of the Spanish Universities.

As it can be seen, this is a very ambitious project with a national vocation, whose strategic plan is to establish a free access window to the Spanish universities, with protocol OAI, which was the institutional and scientific reference from the Theses and other academicals documents. In the same way as the TDR, it pretends the widest spread through internet of the University researching results, to promote the Digital Libraries and electronic edition at the same time that it offers a tool that channels the scientific production. This server not only enriches the Theses information sources world but also the one of communication and spread of scientific knowledge [Ayuso-García, Martínez Navarro, 2004b: 86].

3.2. University and/or digital libraries own projects.

As well as the most important cooperative projects developed in Spain, it's necessary to say that from the seventy two registered Spanish Universities, a meaningful part have in their Library Catalogue the Doctoral Theses reference from the Ministry, in printed format and limited access. Nowadays, this situation is getting solved with the cooperative digitalization projects and the restructuring projects that we have already mentioned. However, there are also initiatives from the own universities or their libraries that recommend that Doctoral Theses should be presented in digital format, which allows a free access to consult them, at the same time that the digital stock of their libraries is incremented.

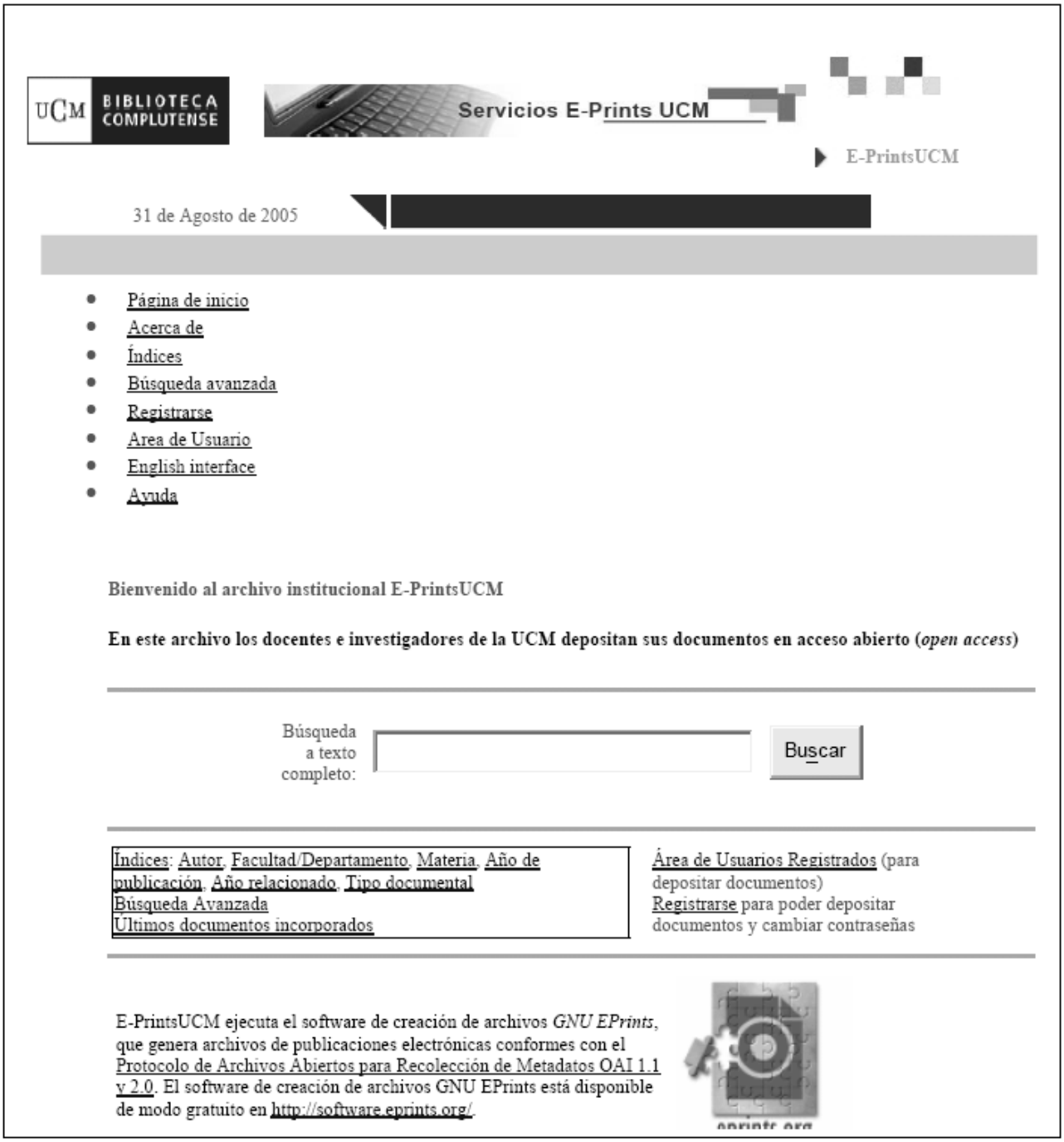
The examples are varied, but we are going to indicate only the most important ones. We will also point out the important initiatives promoted by the digital libraries that from the beginning take the Theses in digital format and make them free access. The most meaningful case, as we will see, is the one of the Biblioteca virtual Miguel de Cervantes.

As we have said, in general, Doctoral Theses have a very limited diffusion, in spite f their interest. Traditionally, the complete publication is not frequent. In the majority of the cases the edition is partial and the methodology doesn't usually appear. For this reason, the original researching works have difficult access to potential scholars and researchers of these fields.

- Universidad Complutense de Madrid (UCM).

Through the Biblioteca Complutense, users can have access to the digital collection of the Theses of the UCM. This collection has more than seven thousand digitalized Theses from 1990-2000. However, only 3000 from the total of the stocks can be freely consulted by Internet. The open access "archivo institucional E-print UCM"[11] has been created; it allows the consultation from a complete text from every searcher in Internet, which is undoubtly very interesting. The search can be of a complete text, of register, and advance. It also can be done through index, author, Faculty, Department, etc. The information recover is of a complete text in PDF format. As it can be seen in Figure 3, it's an open access archive for every user. It's an ambitious project that we should point out and that has among its objectives the restructuring in digital format of the printed Theses.

FIGURE 3



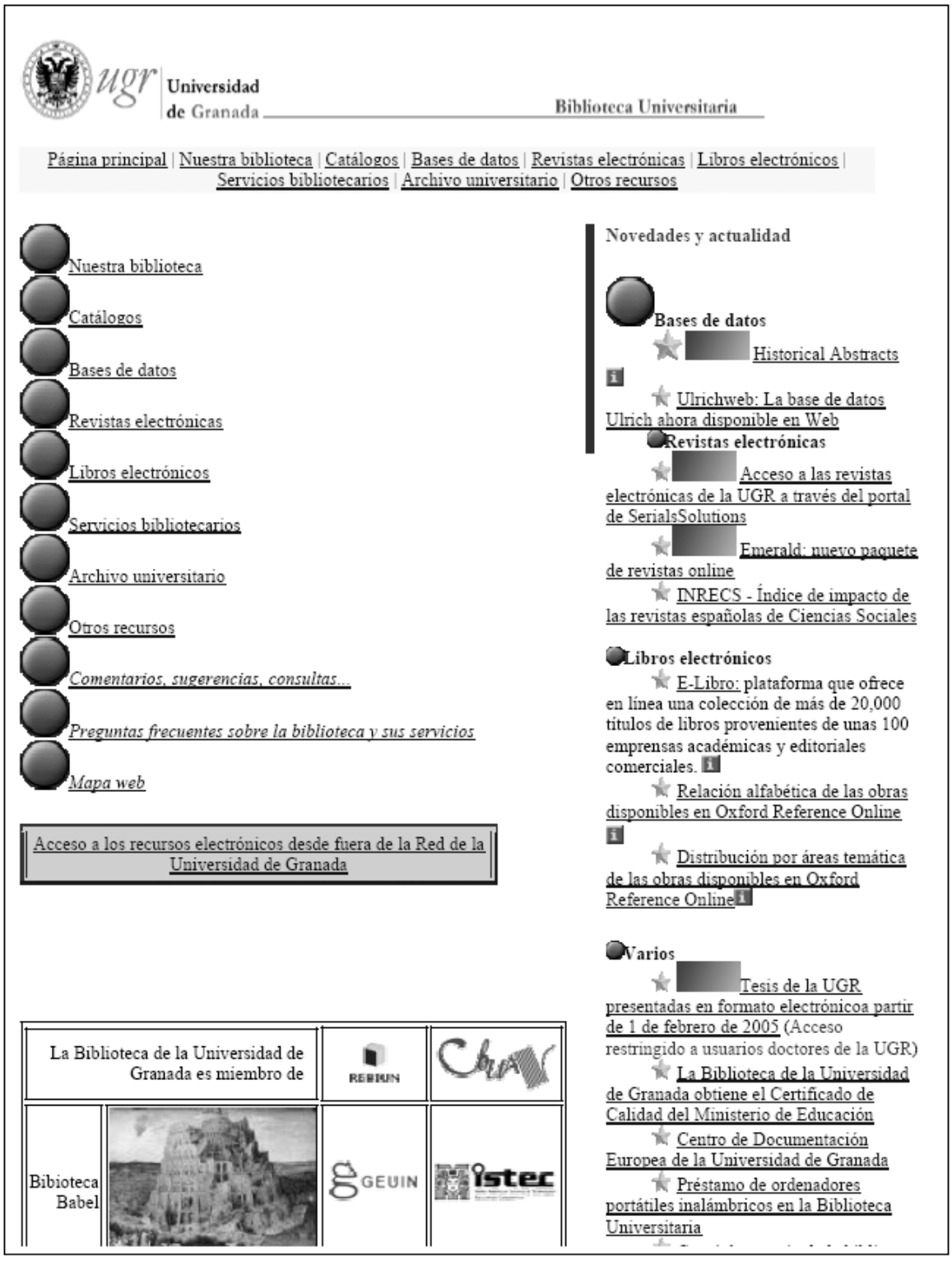
EPrints Services of the Universidad Complutense

[11] Welcome to the institutional archive E-Prints UCM. Available in Web: http://www.ucm.es/eprints. Consulted: 31, 07,05.

- Universidad de Granada (UGR)[12].

This university allows access through its University Library. In the link "Varios" (quite surprisingly) we access the Theses of the UGR in digital format from the 1st February 2005 only through the Departments. It's interesting to note it indicates the number of Theses that have been read and their reference, but it's disappointing that the access to approximately 200 Theses is not free and it's restricted to UGR users, as we can see in Figure 4.

FIGURE 4



Access to the Biblioteca Universitaria de la Universidad de Granada

[12] Universidad de Granada (UGR). Available in Web: http://www.ugr.es/rebiblio/. Consulted: 30, 08, 05.

- Universidad Politécnica de Valencia (UPV)

This university creates the Doctoral Theses Digital Library[13] in the electronic edition inside the PQ Digital Dissertations database. It allows searches of bibliographic references and abstracts and also the first 24 pages of the Theses. The access is free and without any cost in PDF format. It's interesting to note external users have the possibility of buying the Thesis in digital format.

FIGURE 5

Access to the Servicio de Información Bibliográfica de la UPV

Maybe the problem is the language used is Catalan and not everybody, even in Spain, can understand it.

- Universidad de Murcia[14].

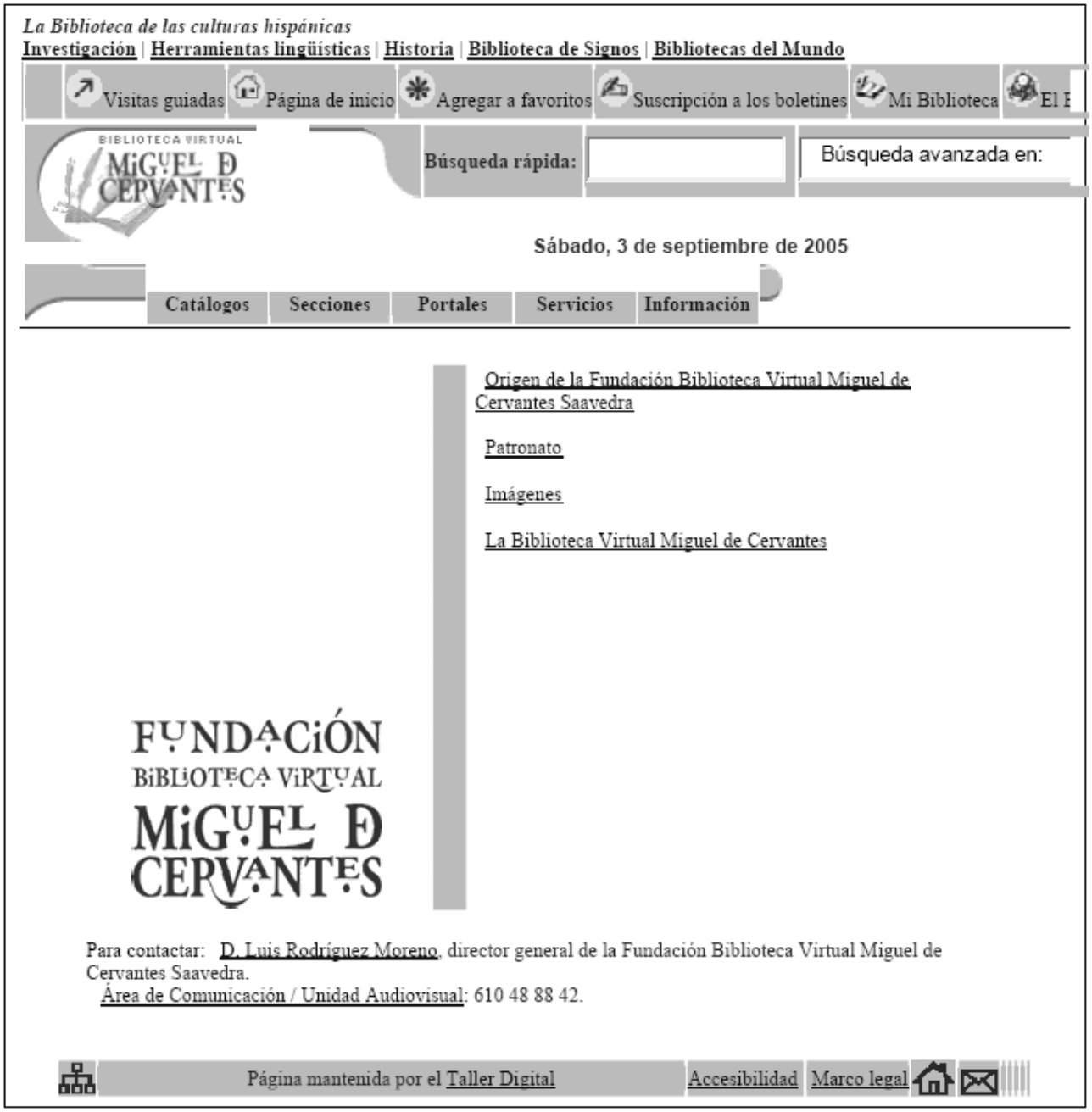
As many others this university is included in the network Project and the institutional page. Theoretically all universities have a tool to edit their Theses in digital format, through TDR or through the REBIUM Project, and this fact is helping without doubts to permit the free access and the information retrieval. However, and bearing in mind the data that we present in this paper, we are still far away to reach optimal results regarding the Doctoral Theses free access.

3.3. The Digital Doctoral Theses of the Biblioteca Virtual Miguel de Cervantes (BVMC) case.

Although it is note university institution, it's necessary to point out the role that this Virtual Library plays in the GL access, in particular in Doctoral Theses, and the importance its free access has in this project since it was created. In general, we can state that the BVMC[15] is a major project of digital edition of the bibliographic, documental and critical Spanish and Spanish-American heritage, which pretends the Hispanic cultures universal strengthening through the use of more advance technologies (Figure 6).

[13] Doctoral Theses of the Universidad Politécnica de Valencia. Available in the Web: http://www.upv.es/bib/digital/dig_tesis_c.html. Consulted: 31, 07, 05.
[14] Universidad de Murcia. Biblioteca Universitaria. Available in Web: http://www.um.es/biblioteca/. Consulted: 30, 07, 05
[15] Biblioteca Virtual Miguel de Cervantes. Available in Web: http://www.cervantesvirtual.com.

FIGURE 6



Biblioteca Virtual Miguel de Cervantes Link

This Digital Library was inaugurated in 1999 as an initiative of the Universidad de Alicante and the Banco de Santander-Central Hispano, in collaboration with the Fundación Marcelino Botín. It's designed as a cultural, educative and researching offer to universities, private or public institutions involved in Hispanic culture. It's a singular project that constitutes a great web access window to several digital cultural and bibliographic webs of Spain and Spanish-America. It allows the visit to the Libraries´ Webs in both continents. It also offers a complete directory of the library and bibliographic electronic resources in the entire world, but above all, and that's why it's so interesting for us, it acts as an editor and headquarters of some works inside the GL: Doctoral Theses and educative and scientific text publications. As we can see, it's a different offer that gives Doctoral Theses information not only from Spain but also from Spanish-America.

The general catalogue allows users to access its contents, from authors and titles alphabetic lists to the field process (CDU). It is also possible to locate the stocks through formularies in advanced researches and to link to works available in the network in other languages (Catalan, Galician, English, French, Italian, German and Portuguese).

The commitment of this institution on researching has made the researching department very interesting for researches from the entire world. Archives, critical works, newspaper libraries and Doctoral Theses, researching and informatics applied to digital libraries are also presented in different sections.

The BVMC knows since its creation that Theses and academic studies are an indisputable information source and a fundamental base for the scientific knowledge development. In this way they have decided from the beginning to accept and edit in their domain all the defended and approved Theses from every country in the world in Spanish language.

FIGURE 7

Author index card

The electronic edition of the Thesis and its diffusion in Internet have specific characteristics because of their contents and because of the users: Hispanic language researchers.

The access Is allowed by author, title, university, UNESCO field classification and Clasificación Decimal Universal (CDU)[16]. The access in PDF format isfree and has no cost for the user. As we can see it is a different project with a huge interest for the Hispanic and the Spanish speaking researchers.

CONCLUSIONS.
Summing up we would like to go back to the concept we have already explained at the beginning of this work by saying that this part of the GL, Doctoral Theses, is not really widespread and often when they are very limited. That's why the access to scholars, researchers and users in general is very difficult, even though this situation is getting a little bit better nowadays.
We would like to point out that, in spite of its importance, the TESEO DataBase is not enough. Nowadays, in Spain, maybe a little bit later than in other countries, networked Doctoral Theses

[16] Ayuso García, M.D. Curso de Módulo de Evaluación de recursos en bibliotecas digitales. Lima (Perú), 6-16 de Junio de 2005.

digitalization projects are being made Because of this tendency of limited access is luckily changing very fast, since approximately between the 20% and the 30% of the Theses in our country are digitalized and have free access.
From the seventy two registered Spanish Universities, acording to our data, only approximately a quarter of them contribute, cooperating with other institutions or universities or on their own, to Doctoral Thesis digitalization and free access. However, we have to say that this number has been a considerable improvement and the instruments and tools to allow the access and spread with the TDR (CESCA) and REBIUN projects and the proposal of the UCM, etc. already exists.
It's necessary to consider the quality of the Doctoral Theses retrieval instruments and to promote the quality control of them, specially those who are published in Internet and have free access (Ayuso-García, Martínez Navarro, 2004c: 135).

The future is encouraging.

REFERENCES

- Ayuso-García, M.D. (2000). Nuevas formas de comunicación científica en la sociedad global. Aproximación a la Literatura Gris. Actas del XV Coloquio Internacional de la Asociación Internacional de Bibliología. Salamanca, 9-11 de Mayo, 2000. CD Rom, II.
- Ayuso-García, M.D. (2003). La consolidación de las nuevas/actuales formas de comunicación científica en la sociedad del conocimiento. En: Actas del I Congreso Internacional sobre Tecnología Documental y Conocimiento. Madrid. 28-29 Enero 2003.
- Ayuso-García, M.D. y Martínez Navarro, V. (2004a). La Literatura Gris en entornos digitales: estrategias de calidad y evaluación. Revista Interamericana de Bibliotecología. Vol. 27, nº 2 (Julio-Diciembre) 2004, p. 49 – 70.
- **Ayuso-García, M.D. y Martínez Navarro, V.** (2004b). Propuesta de auditoría de recursos de información en Internet. SCIELO y la ciencia perdida. Investigación Bibliotecológica. Vol. 18, nº 37 (Julio – Siembre) 2004, p. 85 – 109.
- **Ayuso-García, M.D. y Martínez Navarro, V.** (2004c). La recuperación de la ciencia en Internet. Información digital y conocimiento científico. Boletín de la Confederación de Archiveros, Bibliotecarios, Museólogos y Documentalistas (ANABAD), nº 3, julio- septiembre, 2004, p. 133-171.
- **MacKenzie Owen, J.S.** "Value and use of a grey literatura: the implications of networked publishing" En: First Internacional Conference on Grey Literatura. Ámsterdam, 13-15 de Diciembre, 1993. Sesión III. Suplí and demand of Grey Literatura. There and ceser, p. 308 – 319.
- **Ramos de Carvalho, E.M.** La literatura Gris y su contribución a la Sociedad del Conocimiento. "Conferencia general de la Internacional Federation of Library Association". 2001.
- **Sorli Rojo, Á. y Merlo Vega, J. A.** (2002), Bases de datos y recursos en Internet sobre tesis doctorales.- In Revista Española de Documentación Científica, 25 (1) enero-marzo 2002, pp. 95-106.- ISSN 0210-0614. Puede también consultarse en línea: http://exlibris.usal.es/merlo/escritos/tesis.htm.
- **Torres Ramírez, I. de** (1994), La llamada Literatura gris. Reflexiones sobre su naturaleza y desarrollo e instrumentos bibliográficos para su identificación y localización (II).- In Boletín de la Asociación Andaluza de Bibliotecarios, Año 10, n1. 35, junio 1994, pp. 9-26.- ISSN 0213-6333.
- **Torres Ramírez, I. de** (1995), Fuentes nacionales para el conocimiento de tesis españolas. Reflexión breve. -In Actas de las VIII Jornadas Bibliotecarias de Andalucía, Huelva, Diputación Provincial, 1998, pp. 297-307.

Author Information

Isabel de Torres Ramírez is PhD in Roman Philology and Professor of Bibliography and Information Resources at the University of Granada. Her research activity is focused in the field of the General and Specialized Bibliography on which she has published many articles and books, some of the more significant ones are the following: Bibliografía. La palabra y el concepto (1990), Diez años de Tesis doctorales de la Universidad de Granada (1993),)Qué es la Bibliografía? Introducción para estudiantes de Biblioteconomía y Documentación (1996), Las Fuentes de Información. Estudios teórico-prácticos (1998), Fuentes de Información para los Estudios de las Mujeres (2000) y Miradas desde la perspectiva de género. Estudios de las mujeres (2005). Some of her most cited articles are: "De nuevo sobre la Bibliografía general, nacional, periódica en España. Historia y presente" (1991); "La llamada 'Literatura Gris'. Reflexiones sobre su naturaleza y desarrollo e instrumentos bibliográficos para su identificación y localización" (1994); "Colecciones monográficas españolas dedicadas al tema mujer (1975- 1997)" (1997) y "Los estudios de Bibliografía en el último cuarto del siglo XX" (2002). The last three lustrums, she has coordinated the Section "Mujer" of the journal Crítica, and nowadays she is a member of the Editing Board of the Feminae Collection of the Institute for Women Studies of the University of Granada.
Email: idtorres@ugr.es

TEF: Metadata for French dissertations

Dalia Boudia and Rosa María Gómez de Regil
Institut National des Sciences Appliquées, INSA de Lyon, France

Abstract
Since 2000, universities and research institutions have been encouraged to promote and preserve scientific and technical productions by disseminating digital theses and dissertations. This national program implies that new practices and new tools impacting thesis identification are going to be established at different levels (administrative circuit, theses production mode, theses submission methods, document processing chain, dissemination platforms, digital preservation systems…).It is important to describe theses and dissertations using metadata in Internet shared formats to gain information visibility. TEF (Thèses Électroniques Françaises) standard objective is to permit quality metadata production, exchange and dissemination for French digital theses and dissertations.TEF defines a metadata set for French digital theses. It proposes two standardization levels: a metadata set and an XML schema (exchange format and validation tool). TEF will be composed of descriptive metadata (bibliographical data) and administrative metadata (administrative circuit, rights and preservation data). The XML schema explicitly defines valid XML documents structure. TEF standard has to facilitate metadata exchange at local level (universities) and at national level (SUDOC, CINES). At local level, it will manage theses production and validation, metadata production and full text access. The descriptive metadata set derives from Dublin Core standard, which guarantees a good interoperability with different kinds of documents or different dissemination environments (cf. OAI-OMH protocol). It is also compatible with NDLTD (Networked Digital Library of Theses and Dissertations) Dublin Core application profile, named ETD-MS (Interoperability Metadata Standard for Electronic Theses and Dissertations).

Introduction

The French national Department of Education and Research[1] desires to promote the development and long-time preservation of French theses[2] and dissertations, disseminating them in an electronic format[3]. This national position implies that new practices (theses production and validation) and new tools (publishing platforms, metadata management) are going to be established. The objective is to respond to digital environment stakes. This attitude will impact thesis referencing and identification.

Today, French theses are referenced in the SUDOC national catalog (Système Universitaire de DOCumentation), in library catalogs or in specialized bibliographical tools (subject databases, institutional databases, search engines). If we want to extend theses internet referencing they should be described following a shared metadata schema.

Metadata (any data defining or describing other data) facilitate information sharing, contribute to minimize data loss, help in the searching and data storage processes. There are several metadata models used to describe networked electronic resources. The Dublin Core metadata set is used to make a brief and basic description of an electronic resource through 15 elements. Nevertheless, theses (research official document) need to be described with some other detailed elements. TEF (Thèses Électroniques Françaises) standard will respond to specific thesis description and its administrative needs. A standardized approach is necessary to facilitate systems interoperability and data exchange.

TEF standard is formed of
- a vocabulary (metadata set) that defines description elements and,
- a XML schema, an exchange format and validation tool useful to process data in computer environments.

TEF standard has been developed in two stages and as a consequence, it has two versions :
- TEF 1.0 proposes descriptive metadata (published on May 2005)
- TEF 2.0 is still under development and will propose administrative metadata (publication expected for January 2006).

Descriptive metadata have a bibliographical function in TEF. They include the Dublin Core metadata set. Administrative metadata will manage the administrative submission and publication thesis process (from subject submission to diploma attribution), the legal control associated to the document (copyright) and the long-time preservation.
This paper presents the TEF standard metadata set through descriptive and administrative metadata.

[1] Ministère de l'Éducation Nationale de l'Enseignement Supérieur et de la Recherche
[2] In this article we are going to refer to Doctoral thesis.
[3] Cf. *Circulaire du 21 septembre 2000*.

Descriptive metadata

French libraries have to reference theses defended in their institutions[4]. Theses are referenced and indexed in the national bibliographical SUDOC catalog. This catalog is structured in MARC format (MAchine-Readable Cataloging).

It is important to increase national research value, so, theses have to be well referenced on the web. As a consequence TEF standard uses Dublin Core metadata as a basic/primary description.

TEF standard will allow to reference simultaneously the electronic version of the thesis on the web, and the paper and electronic versions in the SUDOC catalog. TEF associates Dublin Core elements with UNIMARC fields (DC – UNIMARC mapping).

With a DC-UNIMARC mapping TEF entries will be converted to SUDOC catalog entries without data loss. The goal is to reference them in the SUDOC catalog and to use them in a web environment avoiding a double data entry.

The Dublin Core metadata set is not enough to obtain a thesis complete bibliographical description for the thesis. Some elements were borrowed from other metadata schemas:

- elements with specific data about theses (discipline, granting institution, thesis degree type... or thesis.degree element children)[5]. These elements come from the Networked Digital Library of Theses and Dissertations - Electronic Theses and Dissertations Metadata Schema (ETD-MS).
- qualifiers of dc.contributor element (entity responsible for making contributions to the content of the thesis). Some of these qualifiers come from MARC 21 format (marc.thesisAdvisor, marc.opponent, dc.contributor child elements beginning by « marc »)[6].
- new elements were declared to express all the specific French concepts, such as the elements « école doctorale » (ecoleDoctorale) and « numéro national de thèses »(NNT).
- authority data applying to legal entities or other entities mentioned in the bibliographical description can refer to external authority data entries. These external authority entries can be issued from a controlled vocabulary or an authority list (by default the SUDOC catalog authority entries). They can be linked to other schemas like MADS (Metadata Authority Description Schema).

The next table presents TEF 1.0 elements in a general approach (descriptive metadata)

Table 1 – TEF 1.0 descriptive metadata

DC Element	***Sub-element (child)***	***Encoding schema***	***Status Occurrence***	***Namespace***	***UNIMARC Correspondences***
thesisRecord	All other elements		O, N R	TEF	UNIMARC entry
dc.Title	mainTitle dcterms.alternative		O, N R Fac	DC DC	Zone 200 $a $e Zone 541 $a $e $z
dc.creator	name autoriteInterne autoriteExterne		O, N R N R	MADS	Zone 700, Zone 701, function code $4070 (en zone 700 et 701)
thesisID	NNT nationalThesisPID		O N R N O, N R	TEF DC	Zone 029$b Zone 856$u
dc.subject	KeyWordF KeyWordOther **indexationCTRL**		O, R Fac, R Fac, R	DC	Zone 610$a Zone 610$a
indexationCTRL	VedetteConstruite	Rameau, Fmesh,	Fac, R		
dc.description	abstractF abstractE abstractOther dcterms.tableOfContents		O, N R O, N R Fac, R Fac, N R	DC	Zone 330$a Zone 330$a Zone 330$a Zone 327$a + zones 463 or 488 for work theses
dc.contributor	**marc.thesisAdvisor** **marc.opponent** **ecoleDoctorale** **marc.researcher**		O, N R	DC	
marc.thesisAdvisor	name autoriteInterne/autoriteExterne		O, R O, N R	MARC 21	

[4] Cf. *Arrêté du 25 septembre 1985.* Available at : http://www.sup.adc.education.fr/bib/Acti/These/textregl.htm
[5] *Table 1* presents TEF elements in an abridged and clear form.
[6] TEF 2.0 will re-structure dc.contributor child elements.

marc.opponent	name autoriteInterne autoriteExterne		O, N R	MARC 21	Zone 314$a
ecoleDoctorale	name autoriteInterne/autorite Externe		O, N R	TEF	Zone *314*$a
marc.researcher	name autoriteInterne/autorite Externe		O, N R	MARC 21	Zone *314*$a
dc.date	dcterms.accepted	W3C-DTF	O, N R	DC	Zone 328 (Ind. 2=0)$d, Zone 100 $a positions 9-12, Zone 210$d
dc.type		DCT1	O, R	DC	Label pos. 6 (code : " a "), Zone 106 $a position 0 = "s", Zone 105 $a position 0-3
editionsGroupe	**edition**		O, N R		
edition	dcterms.medium dcterms.extent URI otherEditionID	IMT Selon type d'id	O, N R O, N R O, R	DC	Zone 856 $q, Zone 135 position 0, Zone 230$a for data type Zone 856 $s Zone 856$u Zone 856$f
dc.publisher	name place autoriteInterne/autorite Externe		O, N R O, R	DC	210 $c 210 $a
dc.language		ISO 639-1	O, R	DC	Zone 101 $a
dc.relation		URI	Fac, R	DC	Bloc 4XX
dc.coverage	dcterms.spatial dcterms.temporal	dcTERMS dcTERMS	Fac, R Fac, R	DC	Zone 610 $a Zone 610 $a
dc.rights			Fac, R	DC	Zone 300$a, Zone 310$a
thesis.degree	**thesis** **degree.discipline** **thesis.degree.grantor** **thesis.degree.level** **thesis.degree.name**		O, N R	ETD-MS	Zone 328 (Ind.#2) $c
thesis.degree.Grantor	name autoriteInterne/autorite Externe		O, N R		*Zone 712 $a*
thesis.degree.level			O, N R	ETD-MS	Zone 328 (ind 2=0) $b
thesis.degree.name			Fac, N R	ETD-MS	
madsAuthority	personMADS		O, R	MADS	
personMADS	*MADS* schema childs		O, N R		

Towards TEF 2.0

TEF 1.0 is not a complete version of the TEF standard, it contains only descriptive metadata; administrative metadata are not included. TEF 2.0 (expected at the beginning of 2006) will cover descriptive metadata and integrate administrative metadata.

Administrative metadata

Thesis administrative metadata section will describe thesis administrative follow-up (from subject submission to thesis status attribution), respect of legal obligations related to the document (copyright, university authorization for dissemination) and long-term preservation.

TEF 2.0 will use the METS schema[7] (Metadata Encoding and Transmission Standard), providing a flexible mechanism for encoding descriptive, administrative, and structural metadata for a digital library object, and for expressing the complex links between these various forms of metadata. A XML schema will be associated to a Schematron schema permitting the validation of some technical constrains.
The METS standard is advantageous for TEF because its "structure map" is organized in a hierarchical way. So, it reflects the TEF's FRBR structure[8]. The modular organization of metadata allows managing metadata blocks independently.

[7] Cf. METS site : http://www.loc.gov/standards/mets/
[8] TEF conception included a FRBR data representation. This is useful to express *work*, *expression*, *manifestation* and *item* concepts. Cf. http://www.ifla.org/VII/s13/frbr/frbr.pdf

Rights management metadata

Rights management metadata provide a way of declaring copyright information and controlling users' rights. TEF 2.0 will use METS Rights to encode rights management metadata (copyright, third rights, disseminating authorization ...). METSRights is an XML schema permitting to encode ownership of the intellectual content associated to a digital document.

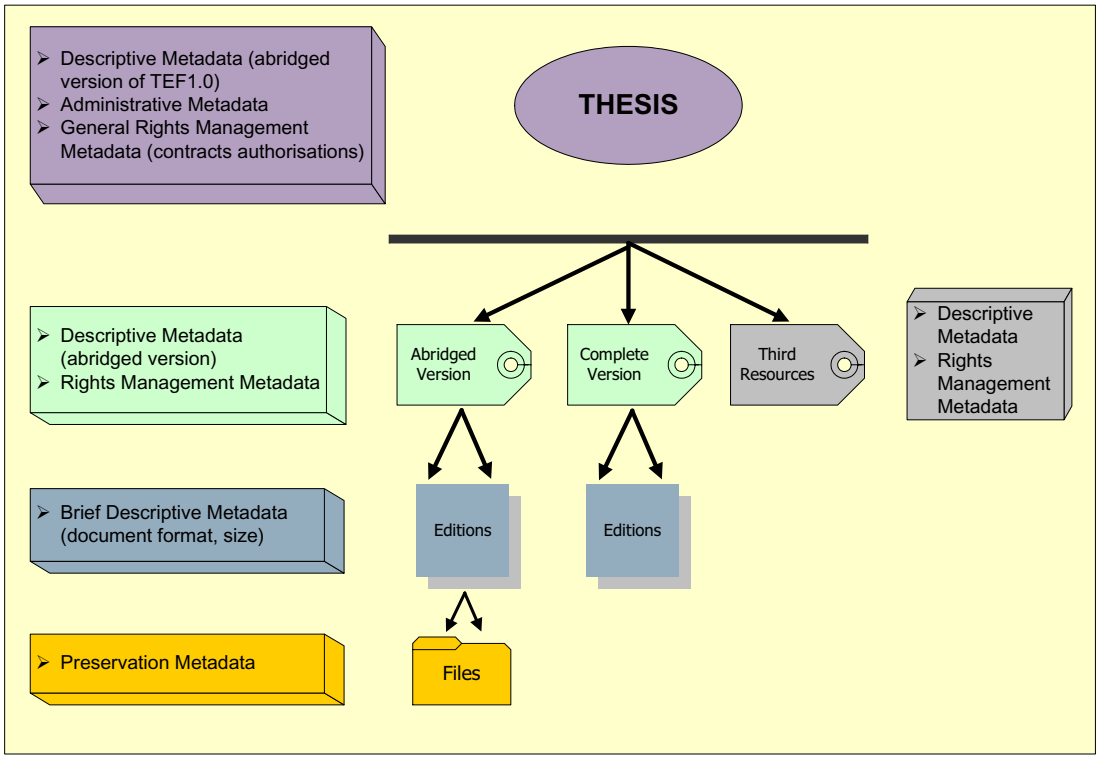
Preservation metadata

TEF defines a set of preservation metadata that will permit long-term thesis preservation. These metadata will accompany theses that will be stored at CINES (Centre Informatique National de l'Enseignement Supérieur) or at other storage centers / servers. Preservation metadata will facilitate theses management at CINES (documentation of the preservation of physical conditions, management of resources, documentation of actions taken to preserve physical and digital versions of resources: data refreshing and migration).

These metadata use some Dublin Core elements that are already defined in TEF 1.0 (dc.creator, dc.contributor …). Some elements of preservation metadata have been taken from the MoReq (Model Requirements for the Management of Electronic Records) specification, such as preservation dates and storage lengths. CINES has created new preservation elements to manage their own data, e.g. CINES archives identifier,...

The next figure shows the structure of TEF 2.0 version. Metadata are organized by blocs (corresponding to different XML schemas) and according to the FRBR structure.

Figure 1 – TEF 2.0 representation



Metadata exchange and OAI/PMH protocol

TEF standard offers a XML schema to facilitate metadata exchange between networks using the OAI/PMH protocol (Open Archives Initiative Protocol for Metadata Harvesting). This protocol defines rules to transfer metadata to different kinds of data repositories simplifying resource access in distributed architectures.

Dublin Core guarantees interoperability because it represents the "minimum required" to conform to OAI/PMH.

TEF metadata may be harvested and supply metadata to repositories. This action can ensure a larger theses referencing.

TEF standard will supply, in its second version (2.0), a subject classification permitting to assign one or more discipline subjects to published theses. This subject classification will facilitate metadata filtering for harvested theses by OAI/PMH protocol. This is also a way to create subject-based metadata repositories.

The classification is issued from a French theses context adaptation to Dewey Decimal Classification (22nd edition).

Conclusion

French theses are preserved in the libraries of the granting universities. They are referenced in local library catalogs and in the national SUDOC catalog.

Thesis referencing and discovering on the web are still limited. TEF work and a new legal national document (departmental order)[9], about grey literature referencing and proposal, are essential elements to have better theses online access to theses and to promote French research.

ABES (Agence Bibliographique de l'Enseignement Supérieur) and CINES are going to propose new global software. The purpose is to help universities to implement TEF in a rapid way. The software is called STAR (Signalement des Thèses, Archivage et Recherche). STAR will manage theses files and will have submission forms for associated metadata (descriptive and administrative).

[9] Expected for the beginning of 2006.

References

ABES. Système universitaire de documentation. [online]. Available at : < http://corail.sudoc.abes.fr/ > (accessed 23.11.2005).

Arrêté du 25 septembre 1985 relatif aux modalités de dépôt, signalement et reproduction des thèses ou travaux présentés en soutenance en vue du doctorat JORF n° du 21 novembre 1985

CINES - Centre informatique national de l'enseignement supérieur. [online]. Available at : < http://www.cines.fr > (accessed 23.11.2005).

Circulaire du 21 septembre 2000 sur la diffusion électronique des thèses BOEN n°34 du 28 septembre 2000 [online]: Available at : < http://www.sup.adc.education.fr/bib > (accessed 23.11.2005).

Dublin Core Metadata Initiative. [online]. Available at : < http://dublincore.org > (accessed 23.11.2005).

Haynes, David. *Metadata for information management and retrieval.* London : Facet, 2004, 186 p.

IFLA. *Functional Requirements for Bibliographic Records.* [online]. München : Deutsche Bibliothek, 1998, 144 p. Available at : < http://www.ifla.org/VII/s13/frbr/frbr.pdf > (accessed 23.11.2005).

Library of Congress. *METS – Metadata Encoding and Transmission Standard.* [online]. Available at : < http://www.loc.gov/standards/mets/ > (accessed 23.11.2005).

NDLTD. ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations -- version 1.00, revision 2 [online]. Available at : < http://www.ndltd.org/standards/metadata/current.html > (accessed 23.11.2005).

NICOLAS, Yann. *STAR, carrefour des thèses électroniques françaises.* Arabesques [online]. 2005, Nº 40, pp. 4-5. Available at : < http://www.abes.fr/abes/documents/ABES%20N40.pdf > (accessed 23.11.2005).

Office Québécoise de la Langue Française. *Le Grand Dictionnaire Terminologique*. [online]. Québec : OQLF, 2002. Available at : < http://www.granddictionnaire.com > (accessed 23.11.2005).

TEF. Métadonnées des theses françaises. [online]. Available at : < http://www.abes.fr/abes/documents/tef/index.html > (accessed 23.11.2005).

Introducing Electronic Theses and Dissertations in Universities: An Indian Perspective

J.K. Vijayakumar
American University of Antigua, AUA, West Indies

Dr. T.A.V. Murthy
Information and Library Network Centre, INFLIBNET, India

Dr. M.T.M. Khan
Bundelkhand University, India

Abstract
The changing electronic environment is forcing Indian Universities, to think about digitizing their theses and dissertations and introduce ETD Systems. A national level study was undertaken to analyze the present scenario in Indian Universities with focus on Ph D Research Scholars, Research Guides and University Librarians. The exact picture of doctoral theses collection, its usage, difficulties in access, academic and research community's attitude towards digital archiving and electronic publishing are discussed. Few Universities have already started ETD Projects, and majority of them will be starting the projects soon. The need for an experienced national level agency for coordinating the work, national level policies from agencies like UGC, necessary financial and technical guidance etc are identified through the analysis. Most of the academics are supporting the idea of ETDs, but still a moderate majority is supporting its Global access through Internet.

Introduction

Indian Higher Education system built upon the British system, is one of the largest and oldest systems of higher education found anywhere in the world. As of now there are 320 Universities, including affiliating Universities, deemed universities, open universities, universities under I. C. A. R, I. C. M. R etc and institutions of national importance like I. I. Ts and I. I. Ms which offer Doctoral Programs [1]. The creation and use of digital information is increasing phenomenally all over the world, and the electronic academic and scholarly communication process, becomes a reality in Indian Universities through University Grants Commission's UGC Infonet Program [2]. Therefore the stake holders are becoming more aware and started looking for more resources and latest information. Librarians in India are also slowly setting their minds to take up the challenges and actively getting involved in e-content creation of grey resources and hosting them on institutional repositories for open access. This changing environment forced Indian Universities, to think about digitizing their theses and dissertations, since these resources belong to the Universities and are not widely accessible out side the campus or Libraries.

1. Need of the Study

A world wide Literature search showed the increase in Ph Ds worldwide and its importance in furthering academic research. The present scenario in the dissemination of Doctoral research results, the problems faced during accessing them, the archival and publication practices of Ph D Theses are solved by the introduction of ETDs as a solution, world wide. It is with this context, there was a need to study the present scenario in Indian Universities with the recent trend in electronic publication and open access. Suggestion of an appropriate model was the part of the project, and this study will be helpful for the model.

2. Methods of the Study

A survey was conducted at national level with focus on Ph D Research Scholars, Research Guides and University Librarians, through specially designed Questionnaires. Research Scholars and Guides were selected from the participants of INFLIBNET's E-Resources awareness programs conducted at INFLIBNET and at various Universities across the country [3]. They were familiar with latest IT developments taking place in Information transfer and working with Universities, where electronic information culture has already emerged. 65 University Librarians, 88 Ph D Supervisors and 173 Ph D Scholars participated in this survey (See table, below).

Category	Number of Participants	Number of Universities
University Librarians	65	65
Ph D Supervisors	88	27
Ph D Researchers	173	27

3. Discussion

3.1 Average Number of Doctorates Awarded

It is found that out of 65 Universities, 29 (44%) awarded more than 2000 Doctorates and 56% Universities awarded less than 2000 Doctorates. Further study, though personal communications and Theses database at INFLIBNET, shows that, there are Universities who have awarded more than 20,000 Doctorates (like University of Calcutta) and more than 10,000 Doctoral Theses (like Banaras Hindu University). Through further analysis it is learned that newly established Universities or Universities with less number of departments, have less than 200 Doctorates awarded. It very clear that, Indian Universities altogether hold a huge volume of Doctoral Theses in their Libraries.

3.2 Importance of Ph D Theses in furthering doctoral research

Librarians rate their Ph D Theses collection, as a source for furthering University research. The value of information containing in these sources is underlined by 46 Librarians (71%), who have rated their collection with High Importance. Only 19 Librarians (29%) rated it with Medium Importance.

45 Guides (51%) rated the Ph D Theses with High Importance and 43 (49%) rated them with Medium Importance, in furthering research in their disciplines. But we should understand the fact that most of them consult the Ph D Theses during the conceptualization of research problems and literature research.

99 Scholars (57%) rated Ph D Theses with high importance, 67 (39%) rated with Medium Importance and 6 (3%) said that it does not have any importance in furthering the research. 2 Scholars have not responded to this question. Further analysis shows that Research scholars from Science and Engineering faculties, especially from Chemical Sciences, rated Ph D Theses with Medium importance or no Importance. The factual and experimental based research style in these faculties can be the reasons or the less usage of Ph D Theses in these disciplines.

3.3 Awareness about the concept of ETDs

Category	Aware	Not Aware	No Response
University Librarians	56 (86%)	9 (14%)	0 (0%)
Ph D Supervisors	53 (60%)	34 (39%)	1 (1%)
Ph D Researchers	96 (56%)	75 (43%)	2 (1%)

It is important to notice that Librarians in India are aware about latest developments like ETDs, and it is because of the wide availability of Internet and Electronic resources which provides up to date information faster than ever before. The less awareness among academics shows that enough awareness programs are required before initiating ETD programs in Indian campuses.

3.4 Availability of a Policy to collect e-format

Even though 86% of Librarians are aware about ETD Initiatives, only 22 Universities (34%) have adopted a policy to collect e-format of theses along with the submission of print copies. It has to be implemented in 43 Libraries (66%) and it is sure that they also have to adopt the policy in par with the emerging e-culture in Indian academic arena.

3.5 Supports to Online Access of Ph D Theses

Category	Supporting	Not Supporting	No Response
University Librarians	31 (48%)	34 (52%)	0 (0%)
Ph D Supervisors	78 (89%)	9 (10%)	1 (1%)
Ph D Researchers	144 (83%)	26 (15%)	3 (2%)

34 Librarians (52%) are still not supporting the idea of online access to the Ph D Theses and 31 Librarians (48%) are strongly supporting the idea. The confusions created by the fetus nature of ETDs and the gap in national level policies and frame works in this direction may be the reasons for their non-supportiveness. But it is sure that online access will be the ideal solution, which will be supported by all Librarians in near future.

The majority of 78 Guides (89%) are supporting and only 9 (10%) are not supporting online access. One Guide did not respond to this question. This shows that the fear of academic resistance towards ETDs is not an issue in India [4].

144 Research Scholars (83%) are willing to provide online access to their Ph D thesis through a Digital Library, but a minority of 26 (15%) are not willing to do so. 3 Scholars (2%) have not responded to this question.

3.6 Online Access Policy of Ph D Theses

Out of 31 Librarians supported Online Access, 12 (39%) suggested the access can be provided only on Library LAN, 9 (29%) suggested Campus Intranet and only 10 (32%) are suggesting for Internet (Global Access), which is the only solution to avoid current difficulties in accessing Ph D Theses of Indian Universities. If we consider the total number of participants in this survey, only 10 out of 65 (15%) supporting Global access to their theses collection. If we analyse the existing ETD systems at International level, the Global access is provided by only 50% of the Institutions where others have restricted the access only to their campuses. Open Access Initiatives are getting momentum in India and it is therefore can be predicted that Global access will emerge as the choice for avoiding the obstacles in physical access.

Out of 78 Guides supporting online access, 50 Guides (64%) are supporting for Global access over Internet, 19 (24%) are supporting access on Campus Intranet and 9 (12%) are supporting access only on Library LAN, to their student's Ph D Theses. Majority of faculty members supports Global access to Indian Theses literature, and the negative opinions will get vanished along with the emergence of digital information environment in Indian campuses when they are fully covered under UGC Infonet programme.

Out of 144 Scholars who support online access, 96 Scholars (66%) preferred the Global access through Internet, 27 (19%) preferred access on Library LAN and 21 (15%) preferred Access on Campus Intranet. 44% Scholars are still not ready to provide Global Online access to their Theses. They may be ready to provide online access after a time gap, to publish articles, papers or books based on their research. If somebody wants to apply for a patent based on the research, especially in Technology disciplines, the research has to be kept unpublished.

3.7 Present Availability of ETD System

It is a matter of happiness that 18 Indian University Libraries (28%) already started creation of ETD Systems, but the majority of 47 (72%) still have not started. The unavailability of Infrastructure and technical expertise are the reasons, and it is sure that it will be vanished along with the emergence and full operation of projects like UGC Infonet.

3.8 Future Plans for ETD System

It is clear that the majority of 30 Librarians (61%) are planning to create ETD System by the year 2005 and 19 (39%) do not have immediate plans to go for ETD System. It is a matter of fact that the wide awareness of ETDs and its benefits among Librarians and the ever demanding electronic culture in Universities accelerate them to go digital in terms of resources and services.

3.9 Supports Needed for ETD System

Out of the 65 respondents, 36 Librarians (19%) demand sufficient policies from Universities, 32 (17%) demand Infrastructure support, 28 (15%) demand Policies from Government bodies like UGC and 27 (15%) demand for Technical expertise. Few of them also demand supports from academics and financial support, where 12 Librarians said they do not require any kind of supports. These Universities have sufficient infrastructure, funding and technical expertise, which are in fore front of digital revolution.

3.10 Obstacles in Creating ETD System

Obstacles identified by Librarians in the creation of ETD System in India. Out of 65, 33 Librarians (51%) identified Copyright Problems are the main obstacles to create ETD Systems. 25 (37%) identified Infrastructure Problems, 23 (35%) identified Plagiarism, 20 (31%) identified Lack of Funding, 15 (23%) identified Lack of Expertise or Lack of Administrative supports as the obstacles. Few of them identified Lack of Supports from Students and Faculty and a very few declared that there are no obstacles in creating ETDs in their Universities. The above mentioned problems need to be resolved by Governmental agencies like UGC, ICAR, ICMR, AICTE etc by having common accepted policies and frame work, sufficient financial support and spreading the benefits of ETDs among academic and researcher community. There may be obstacles which have to be addressed at Institutional level or regional level.

4. Conclusion

The survey was aimed to initiate a discussion and to know the attitude towards going electronic theses among Indian academic and research community, where the idea of electronic theses are gaining much attention day by day. Through this survey and data analysis in this article, an attempt is made to reflect the exact picture of doctoral theses collection, its usage, difficulties in access, academic and research community's attitude towards digital archiving and electronic publishing etc in Indian Universities. It is found that access to Ph D Theses is still facing lots of problems, especially for accessing collection from other Universities. Few Universities have already started ETD Projects, and majority of them will be starting the projects soon. The need for an experienced national level agency for coordinating the work, national level policies from agencies like UGC, necessary financial and technical guidance etc is identified through the analysis. Most of the academics are supporting the idea of ETDs, but still a moderate majority is supporting its Global access through Internet.

Even though there are issues and concerns from every corner, the study clearly shows the trend towards the creation, organization and dissemination of information in electronic means. In the near future, every Indian University will collect electronic format of Ph D Theses and host them in a Digital Archive, which will be accessible through campus intranet or through Internet, according to their access policies. Since this idea is in the initial stages, it is better to have a common policy accepted by all Universities, in terms of format, workflow, software, accessing and archiving policies, which will help them to share this valuable resource through a national level platform identified by Governmental agencies working in this area, like INFLIBNET. University Grants Commission is already working for a policy stipulation to be implemented in all Indian Universities, as mandatory to collect and host Theses in e-format [5].

In the last few years, the Open Access phenomena has strongly embarked upon using ICT for sharing of rich and valuable content through different modes like; computer networks, intranet, internet and interoperable web sites. In this context, sincere effort is made in identifying the issues and problems and lack of coordination and uniformity in the Indian scenario, which prompted to investigate into the system while making a comparison with the already developed scenario and come out with an ideal model suiting to Indian circumstances and be a partner in the global village of Open Access.

5. References

1. UGC, University Grants Commission, Academic Mobility, *(Visited November 30, 2005).* URL: http://www.ugc.ac.in/inside/acad_mobility.html
2. UGC, University Grants Commission, UGC Infonet Project, *(Visited November 30, 2005).* URL: http://www.ugc.ac.in/new_initiatives/infonet.html
3. INFLIBNET, Information and Library Network *(Visited November 30, 2005).* *URL:* http://unicat.inflibnet.ac.in/econ/mindex.htm
4. Vijayakumar, JK; Murthy, TAV and Khan, MTM. Indian academia on copyright and IPR issues of Electronic Theses and Dissertations. In Murthy, TAV and others, ed. Conference papers of International CALIBER-2005, Kochi, 2-4, February 2005. Ahmedabad: INFLIBNET Centre, 2005, Pp 697-704.
5. UGC, University Grants Commission-New Initiatives in ETD, June, 2005 *(Visited November 30, 2005)* URL: http://www.ugc.ac.in/new_initiatives/etd_hb.pdf

Author Information

J. K. Vijayakumar is Assistant Director of Library at American University of Antigua in West Indies. He holds MLISc with First Rank from Annamalai University and submitted Ph D thesis at Bundelkhand University. He was with INFLIBNET Centre in India since 1998 and contributed to its various innovative programmes aimed at modernizing Indian academic libraries. He is the recipient of IFLA Travel Grant - 2004 and participated in IFLA Conference-2004 in Buenos Aires, Argentina and IFLA Pre-conference in Sao Paulo, Brazil, and selected as IFLA/OCLC Fellow - 2002 for participating a specialised training programme based at OCLC in USA. He also received InfoShare Membership Award of ASIS&T and Travel Grant from IATUL. He co-edited International CALIBER 2004 proceedings and widely published 30 technical papers and reports at international and national level. He is standing committee member in IFLA – UN Section, life member of SIS, ILA, IASLIC, KLA, GGSS, ISOC etc. His areas of interest are Digital Libraries, ETDs, Web Integration, Digital Repositories, Internet and Online Resources.
Email: vijayakumarjk@yahoo.com

Dr. T. A. V. Murthy is the first LIS professional to become the Director of INFLIBNET with the status of a Central University Vice Chancellor and elected President of Society for Information Science in India. He is also the council member of IASLIC and Secretary of Ahmedabad Library Network. He holds B Sc, M L I Sc, M S L S (USA) and Ph.D from Karnatak University and carries with him a rich experience and expertise of having worked in managerial level at a number of libraries in many prestigious institutions in India including National Library, IGNCA, IARI, University of Hyderabad, ASC, CIEFL etc apart from Catholic University and Casewestern Reserve University in United States. His highly noticeable contributions include KALANIDHI at IGNCA, Digital Laboratory at CIEFL, UGC Infonet E-Journal Consortium etc. He has been associated with number of universities in the country as visiting professor and has guided number of Ph.Ds and actively associated with the national and international professional associations, expert committees and has published good number of research papers and books. He visited more than 22 countries on professional activities and organized several national and international conferences and programmes. He is the recipient of SIS Fellowship, SATKAL Librarian Award and SALIS Harish Chandra – Sushil Chandra Best Librarian Award.

Dr. M. T. M. Khan is the Professor & Head of Ranganathan Institute of Library and Information Science, Bundelkhand University, Jhansi. He holds M.Lib.Sc. Degree from Aligarh Muslim University and earned Ph D from Jiwaji University. He has rich teaching and guiding experience of 25 years and guided several scholars for research work and Ph Ds. He widely published in the form of books, conference papers and journal articles and organized several conferences and training programs. He is also actively involved in professional associations in India.

Indexing grey resources: considering the usual behaviour of library users and the use of Dublin Core metadata using a database of specialised vocabulary

Jacques Cuvillier
Université Paris 8
Le laboratoire Paragraphe, groupe Document Numérique et Usages, France

Abstract

Grey resources spread over a large set of matters. In each field of interest, people have their own concepts, each group will use their own vocabulary.

Most authors of Grey resources were not able to determine the Decimal or Dewey classification themselves, nor are they able to fix themselves the appropriate terms to embed in the DC elements.

It is then worth providing the webuser with an interface that makes him feel at home, by adapting automatically the terms designing the fields of the search form to those which are in use in his domain of activity.

The whole classification proposed cover about 110 different topics, each one having a set of terms chosen to entitle the fields of the interface of thematic search as well the submission forms.

Moreover, seven entry points are proposed in the first step of the connection, driving the user towards the appropriate context in accordance with the purpose of the document to find : teaching and research ; literature ; technological documentation and data sheets ; vulgarization and everyday life ; citizen and social activities ; arts , sports and spare time ; bank of images.

Indeed, the system must allow the flexibility that allow the replacement of any term by another if necessary. So the entire set of terms is put in a database. This system is also able to adapt the language chosen by the user.

The submission forms filled by authors are parsed according to the topic selected and the Dublin Core metadata file - as an XML file - is generated and stored in the indexing server. It is delivered on simple metadata request, and is reachable by an hyperlink included in any page of search results.

The experimental system has been built and evaluated over a period of five years, and was implemented in the site of the university of Poitiers since June 2005. The software is to be released under free software license (CECILL).

1 Introduction

Sharing Grey resources leads to conciliate the role of several partners : authors, library keepers and library users. In the following, the three will be considered as webusers.

Regarding authors and library keepers, the purpose of collecting a significant amount of Grey literature - even from authors among whom a part have various concerns for rigorous metadata - leads to conceive an adapted system that makes the indexation process as non-constraining as possible.

Moreover, we should keep in mind that classifying a document or embedding metadata within DC elements are not tasks that everyone could manage. Nevertheless, filling up the set of basic Dublin Core metadata is considered as a the lowest goal. The method used for collecting these metadata should naturally lead to convergent interpretations.

Regarding library users, the search task has different aspects depending on the widespread use of concerned documents. In a narrow scope, one or two fields of metadata should be sufficient, for example the author's name or the reference of a course. Also, the screens used for document search are up to the library keepers, and smart means of building up on-demand interfaces[i] have been implemented on the site <Document Libre>.

Nevertheless, thematic search based on the full set of available metadata is considered as the typical and more powerful solution, and will be considered in the following.

2 Grey resources and Grey behavior

Presenting to library users a convenient way to find the required resources is our main concern. A comparison can be done about the ways they got in use till now whether they are stepping in a real library or using an Internet browser.

When he steps in a real library, the customer refines his selection step by step : finding the "right store", where the targeted item should be, and then the "right bookshelf" which might group the documents of a given scope. Usually, the bookshelf is named with words closely related to the topic, according also to the public for which the documents are intended. All is done for the customer to feel confident.

Searching via an Internet browser, the webuser enters some keywords in a search engine that returns a tremendous number of suggestions and has to filter the results in a rather randomly manner.

Indeed, he should make better use of a specialized website allowing a thematic search. We will examine the means of conciliating the user's wellbeing and the constraints of efficient indexation system.

3 How to conciliate Grey behavior and efficient indexation?

"Grey literature has been studied and discussed since the 1970s but users are rarely mentioned ... a lack of widespread and consistent cataloging negatively impacts grey literature use." [ii]

The wellbeing of authors, library keepers and users is only a part of the goal. The essential is the availability of resources and the relevance of the indexing system. The exploration of this problem started in the very beginning of the last decade[iii] but the approach must now take advantage of standardized metadata.

We are aware that Dublin Core metadata brought a decisive stride forward in the area of resources sharing[iv], and is of peculiar interest regarding Grey resources. As most Grey repositories are too small to be able to satisfy customers of their own, DC metadata allow to spread the research over a federative network provided they stay in a compatible framework and develop along a concerted and coherent process. Nevertheless, the witty use of this system belongs to experts, and its misuse would lead to helplessness.

The solution suggested here is based on know-how and skills of contributors nested in specialized contexts, and on a self growing base designed to avoid drift and mismatches, instead of a though-before sieve.

We could still wonder why people feel comfortable in a local library, where all documents are well-ordered according to peculiar local concepts, but would feel a lack of awareness in a standardized library environment ?

On the one hand, we can weight the scheme of classification itself. In a sharp-specialized context, users apply to the classification some refinements that they feel pertinent, but which could not match with the formal (e.g. Decimal) classification, and this one is perhaps not fully understood, nor is it necessarily fully approved.

On the second hand, as nested vocabularies apply to given scopes of activity, the dialogue with people of these domains could be improved if acquainted terms were used.

Therefore, it seems interesting to make the system able to adapt the vocabulary with the webuser. Words from this vocabulary should be used in the interface screen for labeling the form fields, and to give textual metadata as well.

4 A crucial step for thematic search : choosing the entry point

The first step in entering an e-library interface is to select an element in the entry table. Several ways are possible. The ones like Grisemine[v] propose to select first the "document family", others, like Castore[vi] and many others present a list of scopes, but the selection of a class based on both a thematic class and a type of document is also recommended for example by the Deutsche Initiative für Netzwerkinformation (DINI) :

"A rough textual classification facilitates technical structuring and selection. The structure below complies with the functional groups of the German National Bibliography (Deutschen Nationalbibliographie DNB) of the Die Deutsche Bibliothek . Each element should be defined via its functional group number. Descriptions should be output in English (<setName>). An identification via the number will make it easier to convert into the Dewey Decimal Classification (DDC) at a later date at least as soon as the Die Deutsche Bibliothek has converted into the DDC as well." [vii]

A priori, the principle adopted by <Document Libre> could remind the DINI recommendations, and this suggested scheme could take place in the <Document Libre> database without any difficulty. Nevertheless, differences exist in that the main concern of <Document Libre> is less to match closely with a given classification scheme, than implementing a sieve able to cluster the entries so as scope and adapted vocabulary match. In other words, we suggest a usage-based classification.

In a rational approach, this attempt should be preceded by a huge semantic study which could define the smart and clear way in clustering scopes.

We approached the challenge in a more realistic process, starting from a generic scheme and staking on a dynamic growth. As we will see below, it will be made use of practices of different kinds of users whenever they are given directly by suggestions or gradually capitalized from the records of the interactions between the system and the webusers.

5 Entry points : stores and themes

The context is designed according to the image used above, dealing with a customer in a usual library. Terms have been chosen on purpose. Selecting a store leads to the first rough choice. In the very beginning, seven classes have been implemented as "stores" : teaching and research ; literature ; technological documentation and data sheets ; vulgarization and everyday life ; citizen and social activities ; arts , sports and spare time ; bank of images.
This first step is a manner of arrowing the path. Just after, a "theme" must be selected. The latter is the actual cluster - the real thematic entry point - whereas the "store" is the title of a group of themes. Since themes are more than hundred, selecting first a "store" limits the sub-item number to about sixteen themes.

6 Beyond the entry point

6.1 Vocabulary fitting

A convenient effect of having selected the entry point is to have selected at the same time the vocabulary tables which best fit to the theme. These tables are intended :

- for labeling form fields in the interface screens
- for recording already-used keywords in this theme.

In addition, all records of nouns and expressions given by users when submitting a document or managing the classification will be clustered specifically to be associated with the theme in the future. Eventually, these records will turn in recommended vocabulary for new submissions. Library keepers which have a privileged access to the database are able at any time to amend these records in order to reorganize or facilitate the convergence.

6.2 Bookshelves as thematic nodes

We would introduce the key principle retained to facilitate the indexation along whith thematic nodes relying on the user classes, these classes being typically based on the purpose of the document as well as the supposed level of knowledge of the reader. This approach matches with the instinctive know-how of library assistants who put together the books intended for a given public, on a given subject and coverage.
This leads to consider that the most efficient way to express the thematic situation of a document could be a compound of three axes along the topic, the coverage and the required level of knowledge. Related Dublin Core element "Subject", "Coverage" and "Audience" should be the right crossing dimensions.

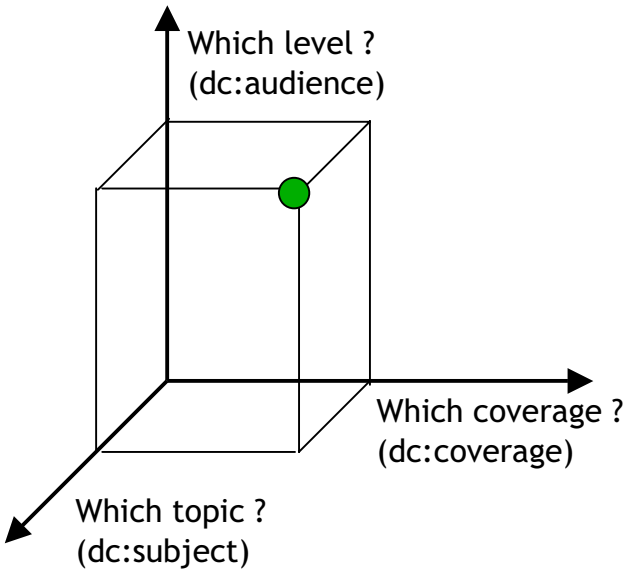


Figure 1 : bookshelf : a three-axial thematic node

Obviously, the matrix built along the full three-axial space of known values of both elements would lead to a tremendous number of nodes. A rather small share of nodes should be relevant.

Let us take an example : technical data sheets do not refer to the whole set of thematic scopes, and the typical definition of coverage[viii] does not apply to such resources which are neither concerned by geographical areas nor by dates.

The resolve power of the crossed metadata would be cut down for each domain which is not concerned by the typical definition of coverage.

This is the case each time geographic zones, dates or jurisdiction are both irrelevant. This argues for new refines of the coverage element that could apply to other kinds of coverage. For example, "Sensors" are drawn to cover different physical scales : force, pressure, strain... The main advantage of this extension would be keeping the same metadata structure and processing methods for all themes.

6.3 Bookshelves as textual metadata

Textual metadata are words or sentences able to describe explicitly the document which they are related to. Terms used to name stores and themes can be considered as textual metadata. In the three-axial space discussed above, each node considered as bookshelf is given a title as explicit as "*Intellectual property in US - Master, Ph.D.*" which represents a textual reminder of the attributes along both axes. These titles are on the one hand an essential guide for the users. They are means for them to feel like in a traditional library. On the other hand, they become also a helpful guide for inserting new documents without breaking the former order.

Indeed, the bookshelf name cannot replace of a DC element. Subject, coverage and audience should have their proper content. Consequently, bookshelf title can be seen as redundant. But since a kind of fail indicator is required to avoid classification mismatches, this redundancy will probably be found very convenient in practice.

7 Development of specialized vocabulary

Specialized vocabulary is divided in two classes whether nouns and expressions are "thought before" or are recorded following the use of the system. Several implementations have been made in order to improve :

a) the scheme along which clustering scopes is processed
b) the vocabulary used in the interface screen for labeling the form fields
c) the textual metadata typically used to name bookshelves
d) the set of recommended keywords

Points (a) and (b) lead to static vocabulary ; (b) and (c) lead to dynamic vocabulary.

7.1 Development of static vocabulary

As stated before, a starting scheme has been adopted as first entries for the thematic search system. Whenever a whole set entry points could be established after an assessment of typical user groups, many clusters could remain empty for a long while ! It is probably more realistic to build a flexible set based on existing resources, keeping in mind that this set could be extended gradually provided the indexing system is not hard-coded but has been thought to allow convenient and quick modifications. Nevertheless, a starting grid has been proposed. The choices have been made carefully, but are drawn to be amended continuously. Also, it has been thought that specialists from anywhere could bring their own contribution - according to the usages of free software developing process. As a body, this kind of development community should bring the required expertise to amend efficiently the initial scheme. Typically, this collaborative organization should involve library keepers of the sites aggregated in the repository network.

In order to collect suggestions, a *blank table*[ix] has been set with first entries listed whereas all other cells have been blanked. It is available on the <Document Libre> site.

Table 1: excerpt of suggestions table (blank table)

rayon	***theme***	***sujet***	***champ d'application (espace, temps, juridiction...)***	***public de destination***
Enseignement-recherche	informatique, systèmes information			
16 thèmes	philosophie psychologie			
	religion théologie			
	religions			
	sciences sociales			
	sciences économiques			
	sciences juridiques			
	sciences de l\'édudation			
	mathématiques et sciences exactes			
	sciences naturelles			
	sciences médicales			
	sciences appliquées,technologie			
	gestion comptabilité			
	linguistique, philologie			
	hist géo, archéologie géologie			
	Autres thèmes			

7.2 Development of dynamic vocabulary

Dynamic vocabulary develops from records of nouns and expressions entered with submissions of new documents.

Each term entered is potentially a metadatum to be. The same words would probably not be used for characterizing on the same "subject" both a scholar work and also a popular work, or for characterizing documents of different types (e.g. a photography vs a non fictional prose. The best vocabulary is then theme dependent.

At the time of submitting their offer, authors are induced to fill a submission form (in three successive screens). Certain nouns must be typed directly, for example the author's name, others are chosen from a popup list, for example the access type ("direct" or "via" access for downloading the document) but in most cases, authors have to choose between using an already recorded word or expression - a suitable set extracted specifically from the database for a given form field being presented in a popup list - or suggesting a new one in a text window. According to this principle, the classification could be suggested by author but this is not mandatory. The decision in matter of classification can be committed to the library keeper.

Figure 2 : Selecting already recorded keywords ans suggesting new ones

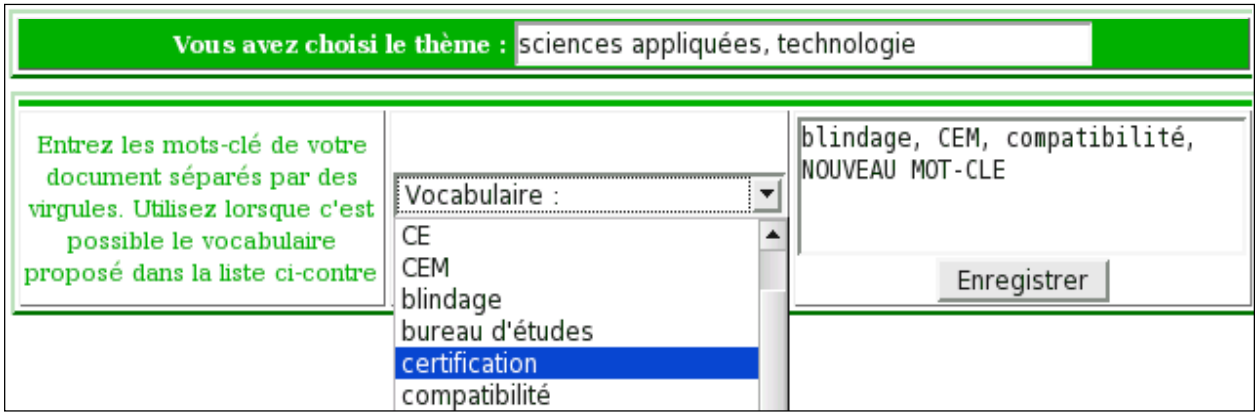
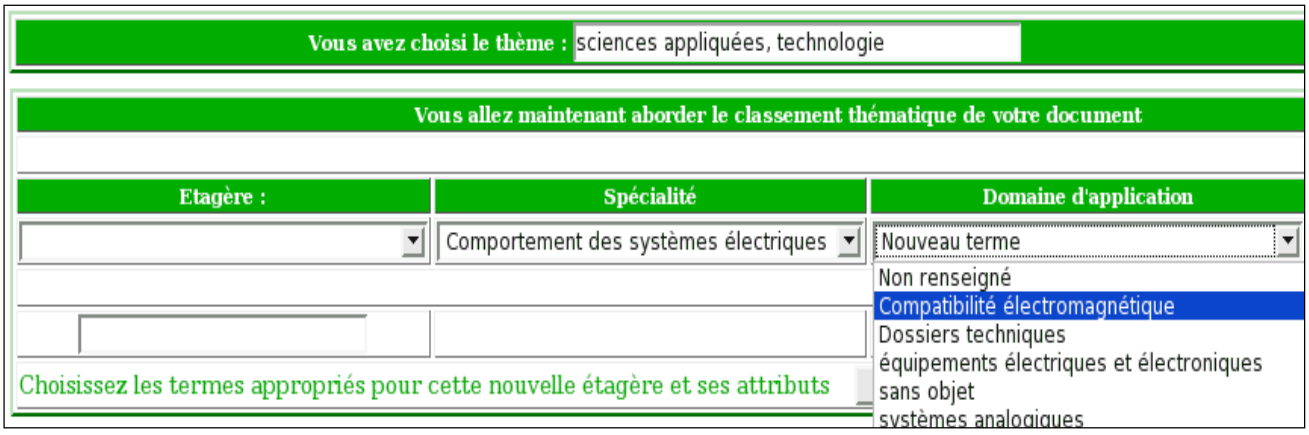


Figure 3 : Selecting already recorded labels and suggesting new ones



Submission records are XML translated and stored temporarily on the indexation server. An automated e-mail is then sent to the library keeper of the site to which the submission applies, with an included hyperlink that shows him the screen intended for reviewers. He is then able to correct or accept new records that complete the database as soon as the accept clic is entered.

7.3 Maintaining vocabulary tables

A comprehensive check up of the vocabulary database is possible and should rely on a collaborative work. The database can easily be managed via phpPgAdmin which is a free software tool allowing an on-line interface to the database tables.

Since it is not worth allowing several people to tinker the database, the latter might be modified at once after concerted decisions. Representatives of repositories aggregated to the <Document Libre> network will be welcome to form a development team.

Improvements can be experienced from local implementations for research purpose. Since the whole system is based on free software[x], local implementations are free of charge. Moreover, mirroring the database is easy via an export file in text format. Pgsql format is native, XML could be performed as well.

Figure 4 : Online interface to the database

theme	cderayon	tspec	tci	tannee
Arts plastiques , photographie	2	Activité	Genre	Public destinataire
Cinema, video	2	Thème	Genre	Public destinataire
Collections, généalogie	2	Centre d'intérêt	Epoque, lieu	Public concerné
Gastronomie, art de la table	2	Sujet	Lieu, époque	Public concerné
Jardin, art floral	2	Activité	Espèce	Public concerné
Jeux de société, jeux interactifs	2	Jeu	Catégorie	Public destinataire
La maison	2	Activité	Domaine d'application	Niveau de compétence
Mode et décoration	2	Activité	Domaine d'application	Niveau
Modélisme	2	Activité	Catégorie	Niveau
Musique	2	Activité	Instrument ou Genre	Niveau
Nos animaux	2	Sujet	Espèce	Public concerné
Sports	2	Sport	Type ou Pratique	Niveau
Vehicules et bateaux	2	Activité	Appliqué à :	Expérience

8 Implementation of the indexing system

In order to clarify the framework of the indexing system, we have first to precise few key points about what is aimed.

- The Indexing system is by no means a library in itself. It offers an indexation service available for independent e-libraries which tend to share their content without being merged in a depersonalized context.
- The share of documents must be done without impacting negatively the role of public relations that a given site is supposed to fulfill. Home page is the normal step to reach a site, and documents are not to be extracted by deep links if this way is not accepted.
- The catalogue of a given e-library and the extent from external repositories are distinguished. An intentional action must be done to extend from the local to the shared content.
- Setting up and managing an e-library or a document collection should be a light work which requires a minimum knowledge and investment. Typically, a basic Internet site and a computer with an e-mail and Internet browser should be sufficient.
- Each indexed file is tracked day by day, and any revision leads to an automated e-mail to the author with an already filled form in order to record metadata modifications if any.
- The system is intended to process a set of requests which return directly as XML or html files the metadata related to a document of a given identifier, or other data like the bibliography related to an author, the result of a search...

As a whole, a given e-library using the system is based on three pillars

- The repository : server on which a group of authors has a writing access and can replace their files if necessary ;
- the e-library : piece of a given Internet site (institutional site... school site...) ;
- the indexation service itself.

Though its address vanishes from library sites[xi] the indexing server performs most tasks in the document search and delivery. Typically, the e-library site must record very short html files such as system calls, one or more personalized styles sheets (css type), perhaps a certain number if images, and a short configuration file. An installation kit is available to make the setup as simple as possible. A ready-to-use standard interface can be chosen. If not, a special tool[xii] helps in designing personalized screens.

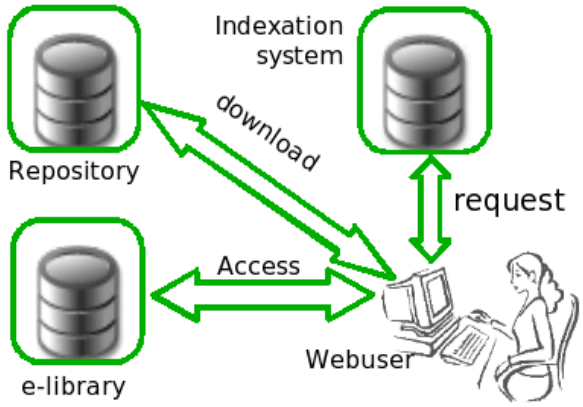


Figure 5 :a Three pillars based system

The indexing system composes html pages at once to fit the requirements of any client along the parameters following the system calls.

The full description of these facilities would be beyond the scope of this paper. We can therefore give below the functional schematic graph of the typical composition process :

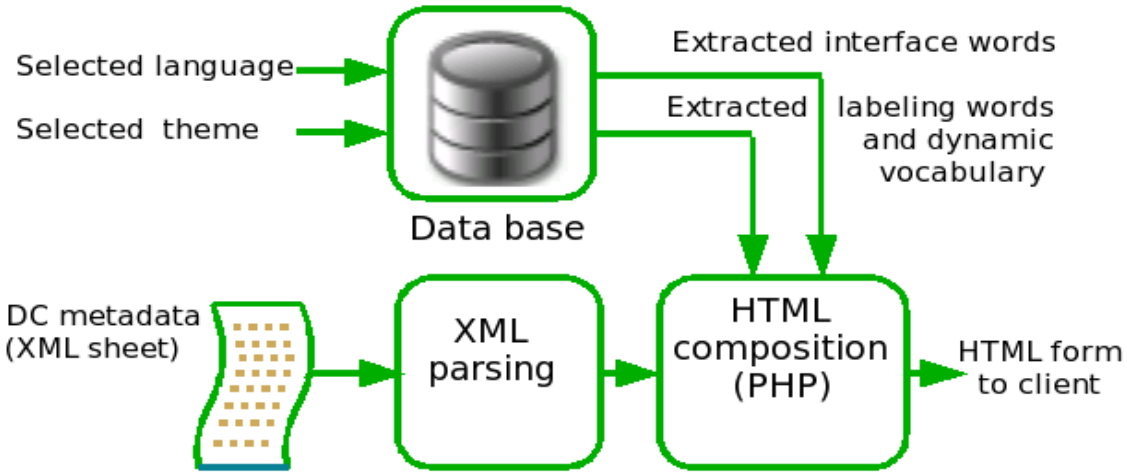


Figure 6: Typical composition process

9 Conclusion

A breakthrough in the promotion of Grey literature is possible with the concomitant effect of powerful communication, judicious rules in a straight framework , convenient processing tools, good human organization. The Web brought the first requirement, standard metadata brought about the second, we hope that improving indexing tools will help provided a sufficient number of protagonists get acquainted with the possibilities and agree together in a way to gather new sharing practices. The ball is now in their court…

References

i http://www.documentlibre.org/kit.html the site <Document Libre> is sponsored by the university of Poitiers and is implemented with the support of i-medias
ii Ranger, S.L. - Grey Literature in Special Libraries: Access and Use. - GL6 Conference TextRelease, 2005
iii http://www.iut2.upmf-grenoble.fr/RI3/Lille97_redoc.htm Vers un serveur d'indexation et de recherche sur le Web (1997) Gilbert EYMARD & Jean-Marc FRANCONY Groupe RI3 Université Pierre Mendès France
iv http://debora.enssib.fr/del63.rtf Metadonnées et bibliothèque numérique Youcef Amerouali, 1999 Université Claude Bernard (Lyon 1)
v http://cridoc.univ-lille1.fr/default.asp?bustl/grisemine, see also : Grisemine, une bibliothèque numérique de littérature grise, Marie-France Claerebout, Jean-Bernard Marino, GL5, dec 2003
vi http://www.emn.fr/x-info/bn/index.html see also : A study of teachers and researchers practices with digital documents - grey or not Author's Name(s): Céline Bourasseau, Cédric Dumas Organization(s): EMN (Ecole des Mines de Nantes) - GL6 Conference Proceedings. TextRelease, 2005
vii Electronic Publishing in Higher Education How to design OAI interfaces - Recommendations - for Data-Provider at german Universities - Deutsche Initiative für Netzwerkinformation e. V. Juillet 2003
viii "Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity)."
see : http://www.dublincore.org/documents/dcmi-terms/
ix Http://www.documentlibre.org/blanktable.html table intended for collecting suggestions
x OS (Linux), database, page processors are under GNU/GPL or similar licence, the server(Apache) is under Apache licence whenever the indexing system itself is under the CECILL licence.
xi Framing technique allows to make the real server address not visible in the address bar
xii Http://www.documentlibre.org/svg2css.html using OpenOffice.org draw for designing an interface screen.

An innovative concept to disseminate scientific communications

Sylvie Grésillaud
Institut de l'Information Scientifique et Technique, INIST
Centre National de Recherche Scientifique, CNRS France

Different information elements compose a congress: before it, a program is available. During a conference, we follow the speech of the orator, some slides and sometimes a text which is published in the detailed program. After the congress, we can read or see the intervention by consulting the conference proceeding or the web site when it exist. On this site we could find the intervention video and slides. But it is not easy to reach the good information: we have to consult the agenda to read the full text or to display all slides. The video and slides could be synchronized. So, by consulting a part of the conference plan, we display the corresponding video part.

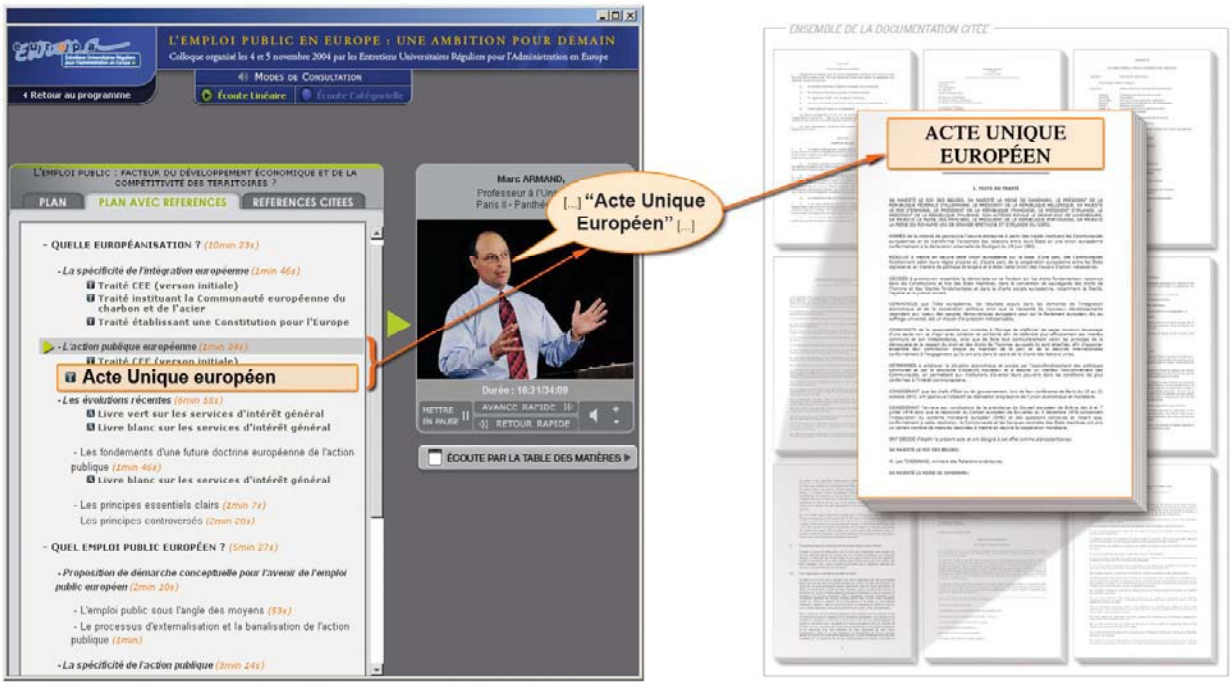
Droit In-Situ (www.droit-in-situ.net), a French private society created in 2000, has developed a new concept based on the oral speech and e-learning. All the technology, which is patented[i] , corresponds on a process and a system for the production of a multimedia edition on the basis of the oral presentations. Today, a collection in Law Sciences is available and commercialised.

The Concept

The innovation is related to an indexed multimedia combination: speech diffused on a video, conference slides, full text and all cited information. The user has two possibilities to consult the final product: a linear listening (from the speech to the cited documents) or a indexed categories browsing (from the cited documents to the speech).

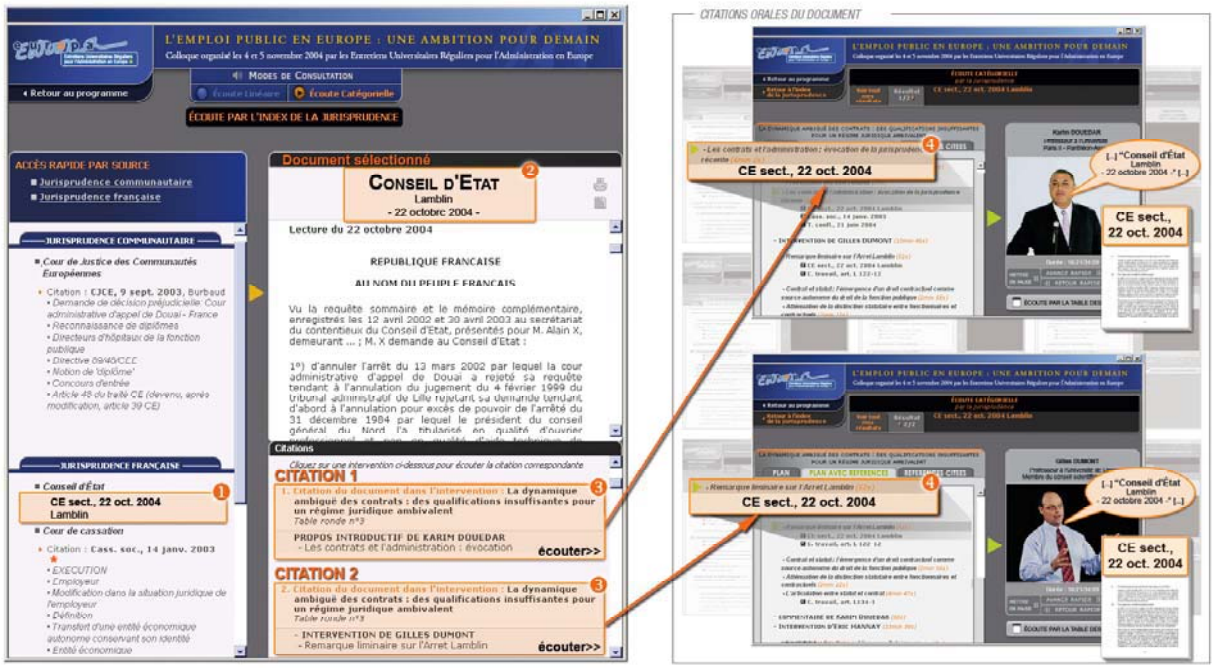
The linear listening

A classical navigation is available: from the plan we can access to the corresponding video sequence. When the orator cites a document, we can consult it.



The categories browsing

All cited information are classed and displayed as a knowledge and e-learning product. From the cited documents, it is easy to consult the correct part of one or several speeches.



The worflow conception

To have such result, a specific workflow conception is needed. The speech and all related information are indexed: time, category, relation with cited documents... Based on these timing steps, sequences are dynamically recomposed.

A partnership with INIST Diffusion

INIST Diffusion is the private subsidiary company of CNRS, in charge of the distribution of INIST-CNRS products and others publishing services. A partener ship between Droit In-Situ and INIST-Diffusion has been contracted to distribute the Law Sciences collections and developp new Sciences collections. In that way, in March 2006, a first Chemistry CD will available: 25 conferences of the EuroChem Nancy 2005[ii] Congress will be published on this innovative technology: the first volume will correspond to the "Prevention of the Chemical Risk" session, the second one will be diffused all plenary lectures, espacially the speech of Pr. Richard R. SCHROCK, who has obtain the Chemistry 2005 Nobel prize.

To obtain more information on this innovation, do not hesitate to consult the web site http://www.inistdiffusion.fr/in-situ/.

Author Information

Sylvie Grésillaud graduated as a Dr Engineer from the French "Grandes Ecoles" National Graduate Chemistry School of Montpellier" in 1986. After some months in working at the private pharmaceutical industry, she joined the National Industrial Property Institute" as a documentalist. Since 1989, she has several activities in the National Institute of the Scientific and Technical Information (INIST-CNRS), mainly as a project Manager. Today, she is the head of the electronic publishing service. She also teaches edition methodology and technology at the different universities (Lyon, Besançon and Nancy) and CNRS doctoral school. Email: Sylvie.GRESILLAUD@inist.fr

References

[i] WO2004/062285 – process and system for the production of a multimedia edition on the basis of oral presentations

[ii] http://www.sfc.uhp-nancy.fr/eurochem2005/

Managing OA Multimedia Multi-type Digital Documents in the OpenDLib Digital Library Management System

Pasquale Pagano and Donatella Castelli
Istituto di Scienza e Tecnologie dell'Informazione, ISTI, Italy

*Abstrac*t
In self-publishing philosophy, authors provide access to their works by depositing them into an OA archive. Before being "published", documents may be reviewed and edited within a process where authors, reviewers and library administrators interact. At the start of this process, authors must upload a copy of their work. It is very important that the self-publishing service allows authors to easily deposit their documents however complex they are.
This paper describes the OpenDLib service for self-publishing. OpenDLib is a Digital Library Service System that allows the publishing, maintenance and dissemination of documents that conform to the DoMDL document model to represent multi-edition, structured, multimedia documents that can be disseminated in multiple manifestation formats.
After briefly presenting the DoMDL Model, the functions for submitting documents by the authors and for managing the publishing process by reviewers (if requested) and administrators are described with some details.

1. Introduction

Nowadays organizations of very different nature– whether academic, industrial, commercial or public ones - are becoming more and more interested in constructing institutional repositories. It is expected that by federating the information published by these repositories a wide comprehensive information source accessible to the interested community at any time and anywhere can be created. In this way, people might operate on this common source by exploiting services that allow them to retrieve, access, transform and produce new information that, in turn, can be disseminated to the others, thus giving rise to a process of continuous knowledge enrichment.
Until now, however, this high potentiality of the institutional repositories has been exploited only to a very low extent. New functionality must be made available to fully support their content that is already going to become of very different nature. In fact, most activities are already producing documents composed of different parts in different formats: lessons/seminars with talks and slide presentations, tables and graphics, videos reproducing experiments, surgical operations or new products and the like. Any single component of these multimedia documents may require great intellectual and technical effort for its design and realization as well as large memory resources for its storing, so it is very important that they can be managed by functionalities that allows people to easily create them, make them searchable through specific metadata and reusable for different purposes in different forms. The OpenDLib system [1,2] makes such functions available by providing services able to support a very rich and powerful document model: the *Document Model for Digital Library* (DoMDL) [3,4]. After briefly presenting the DoMDL Model (Section 2), the functions for submitting documents by the authors and for managing the publishing process by reviewers (if requested) and administrators are described (Section 3). Final remarks regard additional advanced functions offered by OpenDLib by exploiting the DoMDL model (Section 4).

2 The Document Model for Digital Library

In OpenDLib, DoMDL plays the role of the logical document[1] model that is shared by all the services that implement the functionality of a digital library (DL). DoMDL can represent structured, multi-editions and multimedia documents that can be disseminated in multiple manifestation formats. As sketched in Figure 1, documents submitted by the authors, or harvested from different sources, are logically represented and known to all the OpenDLib services as DoMDL documents, although they may be of very different type and format. At the same time, DoMDL documents can be perceived by the DL users under different views. For example, DoMDL can be used to represent a lecture as the composition of the teacher presentation together with the slides and the summary of the talk transcript; this lecture could be disseminated as the MPEG3 format of a video or as the SMIL document synchronizing all parts of the lecture.

[1] Throughout this text, the term "document" is used to mean "multimedia digital object"

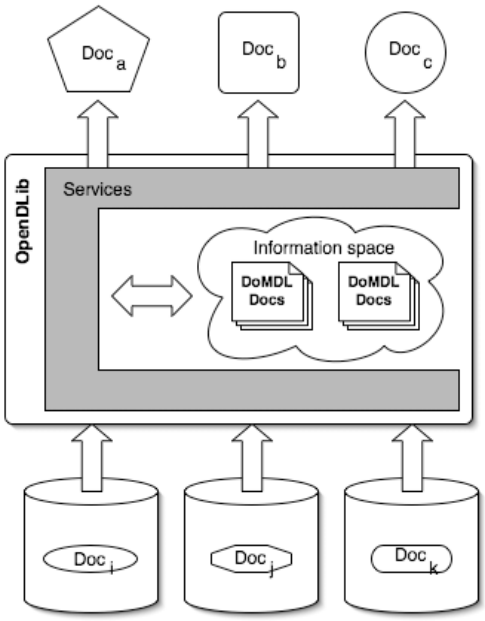


Figure 1: DoMDL in a DLMS

In order to be able to represent documents with completely different structures, DoMDL distinguishes four main aspects of document modelling and, using terms and definitions very similar to those coined in the IFLA FRBR model [5], represents these aspects through the following entities: Document, Edition, View, and Manifestation (see Figure2).

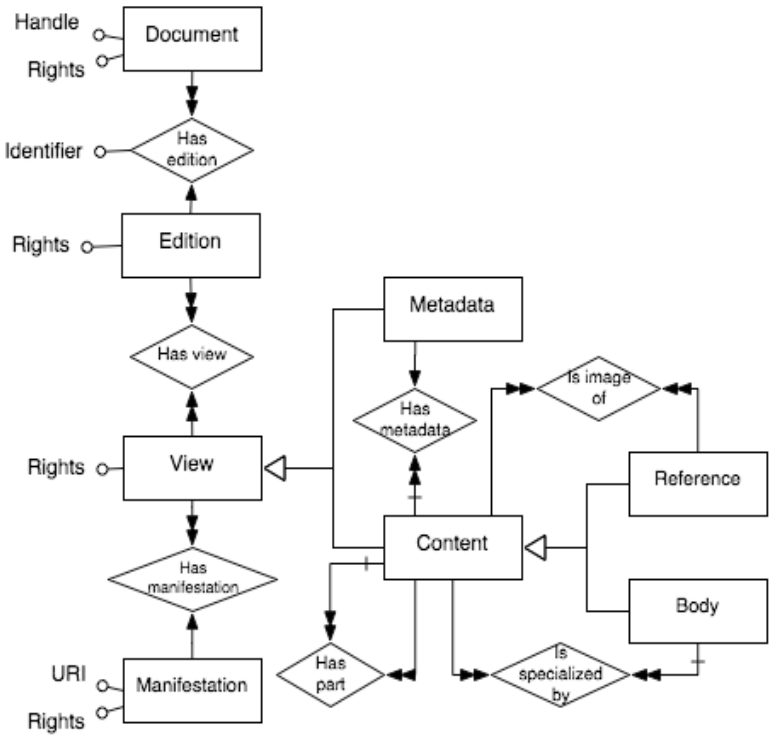


Figure 2: Document Model for Digital Library.

- The *Document* entity is an abstract entity that represents the document as a distinct intellectual creation, capturing the more general aspect of it in very abstract terms; as such, it is very similar to the Work entity defined by the FRBR model.
- The *Edition* entity represents a specific expression of the distinct intellectual creation, thus being able to model an instance of the document along the time dimension. A Document can be recognized only through its individual editions or *expressions.* As such the *Edition* entity is very similar to the Expression entity defined by the FRBR model. The boundaries of *Edition* are defined, however, so as to exclude aspects of physical form, such as text or image formats, that are not integral to the intellectual or artistic realization of the *document* as such.
- The *View* entity has no FRBR counterpart. It has been conceived to model the different ways in which a digital information content can be organized, viewed and disseminated. For example, the *View*

entity related to the original edition of the proceedings of a workshop might modelled in such a way that this edition be perceivable under three different views: a) a "structured textual view" containing a "Preface" created by the conference chairs, and the list of thematic sessions containing the papers presented to the conference, b) a "presentation view", containing the list of the PowerPoint slides used in the presentations, and c) a "metadata view", containing a structured description of the proceedings. This is made possible by how the View entity is specialized, as shown in Figure 2 and described in detail below.

- The *Manifestation* entity, as the homonymous FRBR entity, models a physical formats by which a document is disseminated, i.e., any medium in which any View may be physically embodied constitutes a *manifestation* of that View. Examples of manifestations are: the MPEG file containing the video recording of the lecture made at a conference, the AVI file of the same video, the postscript file of a lecture given by another teacher at the same school, etc. Physical formats are accessible via URIs, used to associate local or networked file locations.

These entities are semantically connected by means of a set of relationships. The relationships *Has edition*, *Has view*, and *Has manifestation* link the different aspects of a document. Note that these relationships are multiple, i.e. there can be several objects in the range associated with the same object in the domain. This means that there can be multiple editions of the same document, multiple views of the same edition and multiple manifestations of the same view.

As shown in Figure 2, the View entity is specialized in two sub-entities: *Metadata* and *Content*. The former allows a document edition be perceived through the conceptualization given by its metadata representations. These may be a flat list of pairs (fields, values), as in the Dublin Core metadata records [6], or more complex conceptual structures, such as in the IFLA-FRBR records. Typically, the metadata view is indexed to support attribute-based querying and browsing operations, but it may otherwise be used. For example, it may be disseminated free of charge while the document contents are regulated by fee access, or disseminated on a mobile device. By using the *Has metadata* relationship it is possible to model that also content views can be described by one or more metadata records in different formats.

The Content view has the following sub-entities:

- *Body*, that is a view of the document content when it is to be perceived either as a whole or as an aggregation of other views. For example, a textual view of the proceedings of a workshop is built as the aggregation of the textual views of its component articles. The relationship *Has part* links a Body view with its component views. A Body view may be specialized by other views that represent more detailed perceptions of the same content. For example, an article of the cited proceedings may be specialized by two views related to the French and English version of that document, respectively. A view is related to all its specializations through the relationship *Is specialized by*.
- *Reference*, that represents a view that does not need to have explicitly associated manifestations because of its being equal to an already registered one. This entity has been introduced to represent the relationship between views of different document editions. Articles presented at the same workshop, for example, can be modelled as documents and grouped together by the workshop proceedings document that contains only the references to them. It is also important to note that this entity, bringing together parts of real or virtual documents, makes it possible to manage virtual documents that are not explicitly maintained by the storage system. For example complex reports, or training lectures, can easily be modelled as a composition of parts extracted from real documents. A reference view is linked to another one via the relationship *Is image of*.

Each of the entities described above has a set of attributes that specify the rights on the modelled document aspects. This makes it possible, for example, to model possibly different rights on different editions, different access policies on different views or on different parts of the same view, and so on.

DoMDL can be specialized to respond to the needs of the DL being created , so that also the services offered by OpenDLib can be configured to support/work with predefined document structures/models

3. OpenDLib services for self-publishing

In self-publishing philosophy, authors provide access to their works by depositing them into an OA archive. Before being "published", documents may be reviewed and edited within a process where authors, reviewers and the system administrator interact. OpenDLib self-publishing services, i.e. the **Submit-, Review-** and **Admin-Service,** are tailored to these different user profiles, although such profiles are virtual ones depending on which rights a DL user is entitled with.

Any institution has its own publishing politics. Some institutions might publish only research papers, others may include books or workshop proceedings, and other might produce journal issues at regular time intervals.

The DoMDL model can be specialized in different ways so that OpenDLib publishing services can allow users working with document models predefined by the DL Administrator. Further DoMDL permits creating personalized document structures (or "free" models) so that authors can occasionally define document structures conforming to the document they want to upload and DL Administrators can pre-define new permanent models to cover changes in or expansions of the institutional editorial politics.

3.1 - Submit-Service

At the start of the self-publishing process, authors must upload a copy of their work and enter metadata that describe it. Obviously, complex documents require more complex procedure, so that it is very important that the self-publishing service allows authors to easily deposit documents and provide them with descriptive, structural and administrative metadata.
The Submit service manages requests submitted by the authors in order to:

- create, update and submit new documents
- create, update and submit new editions of already published documents
- make revision of documents already published
- withdraw documents from the repository

The most significant function surely is the first one, thus it will be described with some details.

3.1.1 - Creating, updating and submitting new documents

The Submit Service provides authors with two submission modes, i.e., whether authors use one of the document models predefined by the DL Administrator (as many models as to cover all the needs of the institutional publishing politics), or they freely construct a new document structure conforming to the document they want to submit.
Using predefined models, authors have only to enter metadata and upload/refer content file/s that represent their documents, however complex their structure or composition be. For allowing authors to upload documents, the system graphically displays the structure of the selected model in an *Edit Page* where also the functions for operating on the structure are made available.
Figure 3 reproduces the *Edit page* of the OpenDLib Interface that shows the structure of the 'simple' model, i.e. a model for non-structured documents such as papers or technical reports, consisting in information content and accompanying metadata. The users can operate on this structure (and can repeatedly update/correct their operations) in order to:

1. identify which information they have to enter (e.g. metadata) or to upload (the file/s of the document information content) and easily make such operations
2. add an image file to be used as a cover (optional)
3. update default values associated with a specific model (optional): for example, an author uploading a book-like document could prefer to update the default name of some document structural elements, i.e., could prefer to replace "Metadata " with "Cataloguing data", an so on
4. update access right defined for each individual structural element by default (optional)

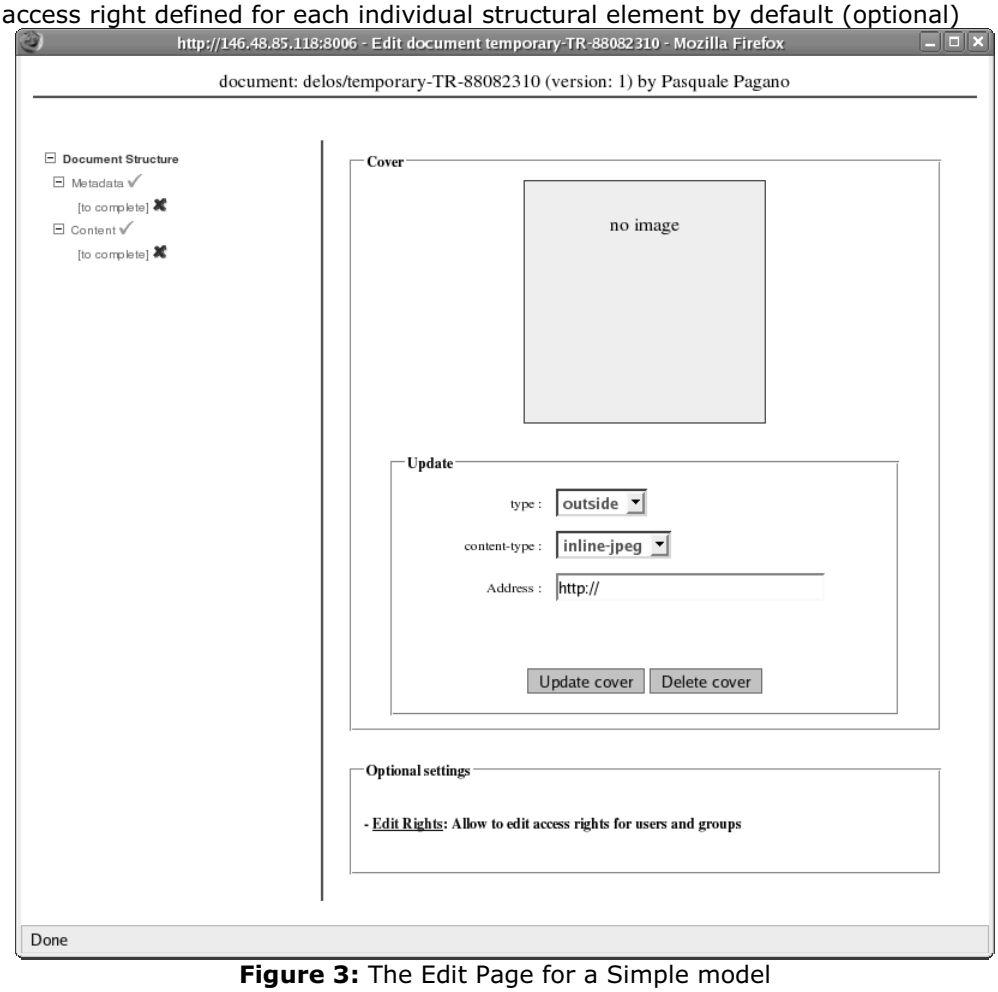


Figure 3: The Edit Page for a Simple model

In most cases, the only needed operation is the first one, that can be executed in the following very simple way: (i) click on the *[to complete]6* [2] structural element related to *metadata* to make a form to appear where metadata can be registered; (ii) click on the *[to complete]6* structural element related to *content.* This opens a window that allows you to insert the related ***manifestation***, i.e. the physical data file.

In the case of a structured model such as "Book" or "Collection", operations 1) to 4) above need to be preceded by the operations to complete the structure. In fact, the structure of such models are necessarily "open", i.e., their hierarchical structure is to be completed by the authors to conform this structure to the document they want to upload (for example, authors have to define how many sections/subsections their book has, and so on). An example of such operations is given in Example 1. The freely creation of new models, instead, is made available by a very advanced function. This function has a twofold objective (i) to permit the DL Administrator to define permanent models to be presented to the authors; (ii) to allow authors to define an occasional, very specific model and submit their document accordingly. A very interesting feature of this function is its capability to reuse pieces of information (whether texts, or images, or tables or videos) already existing in the DL. These pieces of information can virtually be "inserted" in the structure of a new document by simply making reference to them. An example of free model definition is given in Example 2.

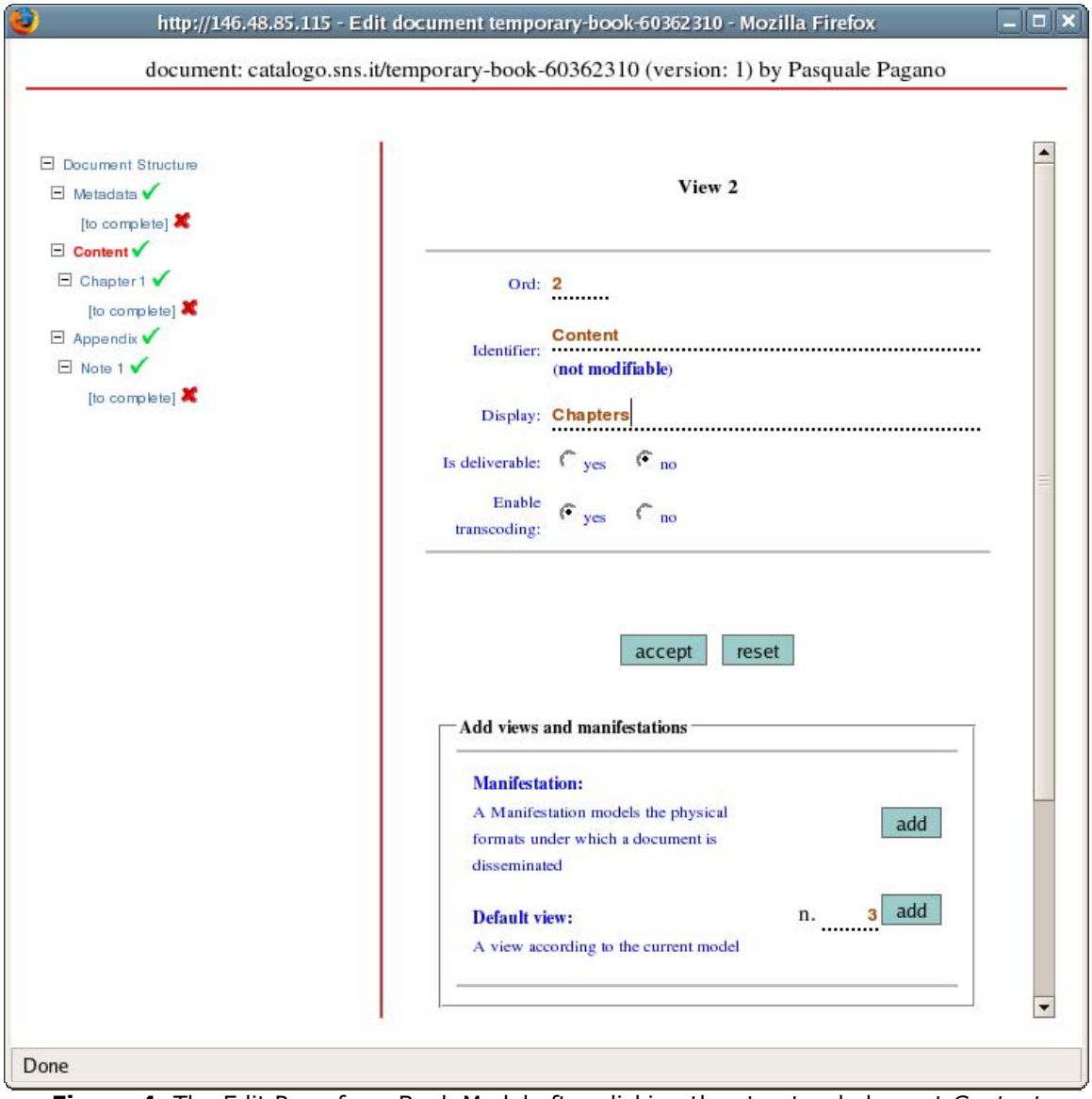


Figure 4: The Edit Page for a Book Model after clicking the structural element *Content*

Example 1: Completing the structure of the "Book" model

Consider, for example, the structure of the model "Book" as it is represented in the *Edit page* of the OpenDLib Interface (see Figure 4). Structural elements of this model are *Metadata, Content* and

[2] Names in italics refer to items in the figure

Appendix. *Content* and *Appendix*, in their turn, are further structured into *Chapters* and *Notes*, respectively. The check symbol √ that marks some document structural element means that information for that element has been completed by default values (such values can be optionally changed). The right-hand side shows which functions are available to insert/update/delete structural elements, as follows: when one clicks on any element of the *Document Structure* then the right-hand side shows which operations can be execute on that element. For example, to add three chapters, the author has to click the structural element *Content.* In the right hand side of the page the *add* button corresponding *default views* allows the author to add as many chapters (and related *[to complete]6* items) as are needed, thus updating the document structure accordingly. Further, if the information content of a *Chapter* structural element consists in more than one file of different type, then that Chapter needs to associate more than one manifestation. For this, after clicking that *Chapter* structural element the author has simply to click the *add* button related to *manifestation*, thus adding as many *[to complete]6* items as the supplementary document files are. Figure 5 shows the result of the above e operations. In the right-hand side of the page the author can personalise default values related to *Chapter 1* by simply updating such values. Similarly, default values for the other structural element can also be updated. It is to be noted that any structural element and any piece of information related to the *Document Structure* can repeatedly be updated by clicking on the respective item. The editing procedure can be interrupted at any stage and then restarted.

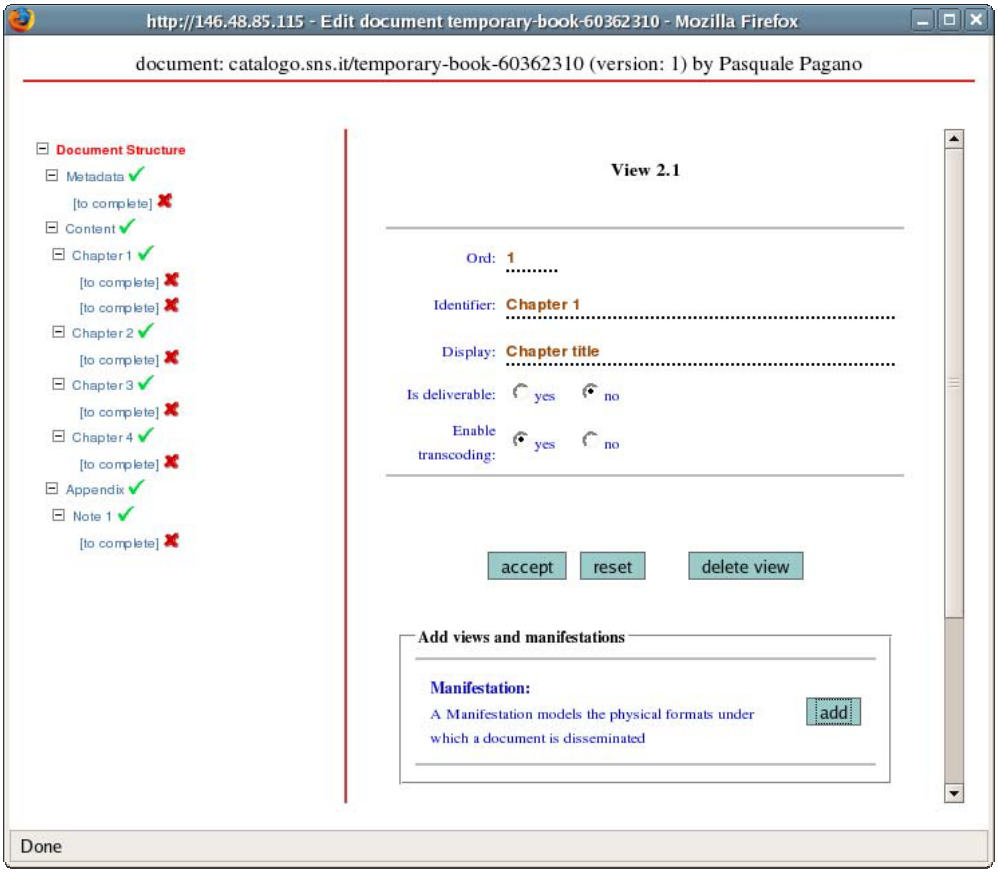


Figure 5: The book model structure completed for a given book

Example 2: Creating a "free" structure

In this case, the *Edit page* shows no predefined structure, rather it displays how such a structure is being created/updated by combining different structural elements. In Figure 6 these elements are shown in the *Add view* box, namely:

- *Body*: refers to document content when it is to be perceived either as a whole or as an aggregation of other parts
- *Metadata*: refers to bibliographic information regarding the overall document or one of its parts
- *Reference:* makes a reference to a structural element of another document

They are called *Views***,** a *View* being a way how a document can be perceived. The *Document Structure* is created by selecting a set of Views while organizing these Views in such a way that the resulting structure corresponds to how authors want their document be perceived.

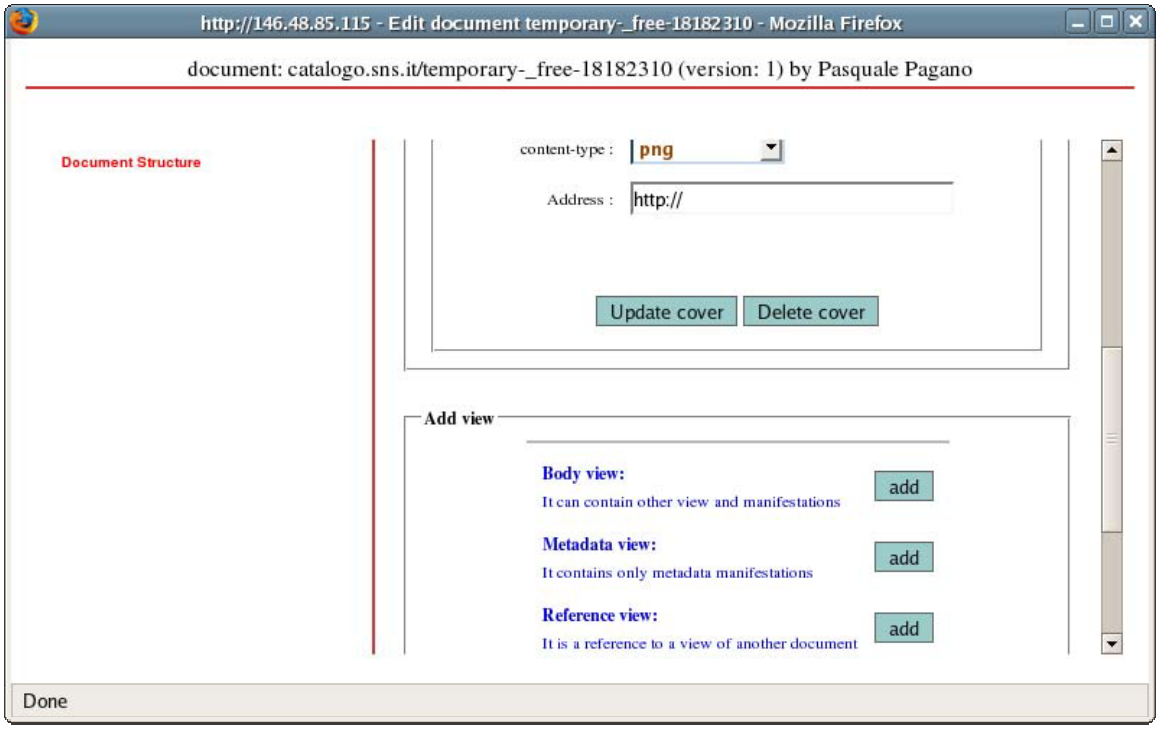


Figure 6: The Edit page for starting the creation of a personalised (free) structure

The name of any newly added *View* - whether *Metadata* or *Body*, etc. - is inserted into the *Document Structure* together with a progressive hierarchical number that specifies its position in the structure tree. The accompanying symbol 6 means that further information related to this view is to be given.
Figures 7-9 show the main steps for creating a document structure/model. Figure 7 offers an example of a body view aggregating another view. This structure is a result of the following actions by the author: (i) Clicking *Document Structure* and then clicking the *add* button corresponding to *Body view*. Thus a *New View body_1* is added to the *Document Structure*. (ii) Click *New View body_1* and, again, the *add* button corresponding to *Body view*. Thus a *New View body_1_1* item is added to the Document Structure. With the same mechanism it is possible to add *New View body_2, New View bod_ 3,* … to the *Document Structure* as well as to further detail any item.

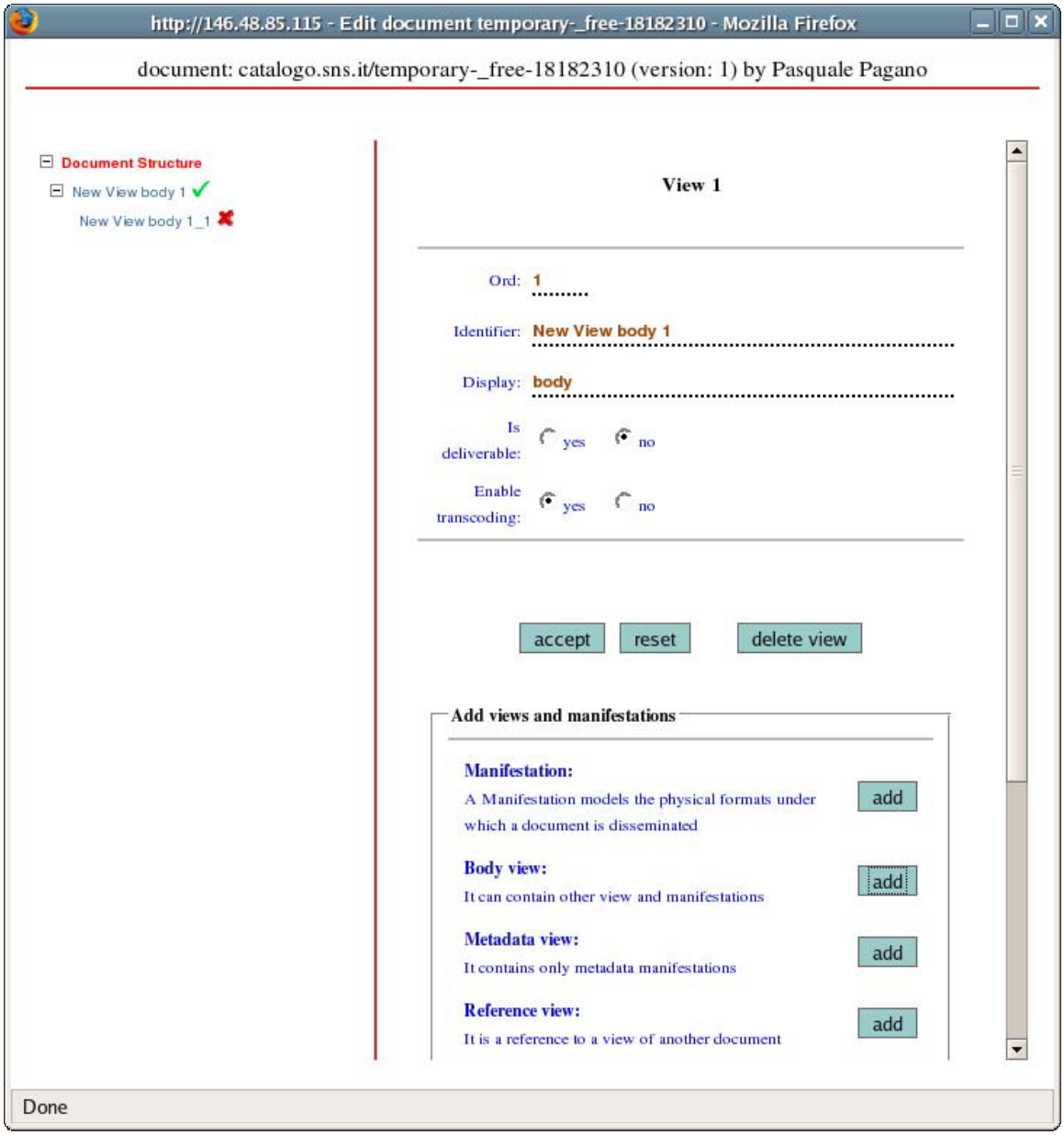


Figure 7: Inserting a *Body view* aggregating another view

Figure 8 shows the result of adding a *manifestation* , i.e. the structural element to upload information content related with *New View body 1.1*. This result is obtained by two steps, the first of which is not exemplified in the figure: (i) Clicking *New View body 1.1* and then clicking the *add* button corresponding to *manifestation***.** Thus a *[to complete]*6 item is added to the document structure meaning that a physical document file is to be uploaded or referred to; (ii) Clicking the *[to complete]*6 item. This makes the window for uploading files to appear in the right side of the Edit page. Figure 8 shows the status of the Edit page after selecting the file to be uploaded and clicking the *accept* button. The detailed description of the uploaded file is displayed. In case of errors, all operations can be repeated after clicking the *delete manifestation* button.

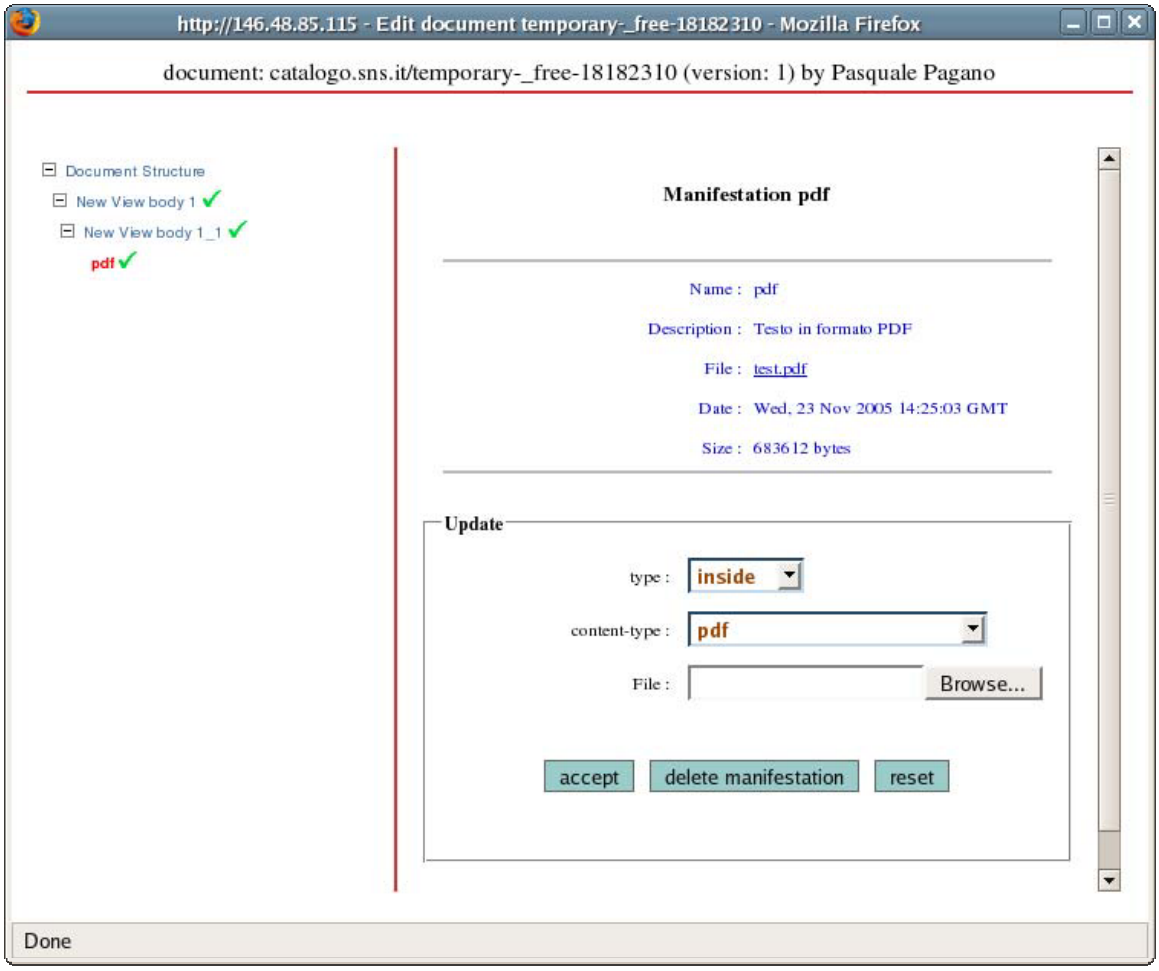


Figure 8: Inserting a *Manifestation* related to *New View body 1_1*

To complete the example, let us show how the use of the above mechanisms could permit an author to simply create the structure for a project report. Such a report is thought of as consisting in: i) a presentation by the project scientific coordinator [a video file to be uploaded]; ii) a section (with related metadata) composed by three referenced views, i.e., the descriptions of the three project objectives and related tasks [text files, already existing as specific *Views* in another document of the DL]; iii) a section dedicated to the presentation of two system prototypes realized in the project [PowerPoint files to be uploaded].
The structure conforming to such a document appears in Figure 9. The right hand side of this figure explains the Reference View mechanism, that makes it possible to "reuse" pieces of information embodied in already existing documents and virtually insert them in the structure of the document under construction. The figure displays the status of the Edit page after clicking *New View reference_2_1*. and registering data in the fields that identify the referenced piece of information through its Reference handle/version/view. As can be seen, *Paragraph 1-1* of the document *delos/2005-DD-004-version2* is to be uploaded every time the *New View reference_2_1* is accessed..
Note that in this example all the default names *New View body* and *New View metadata* have been updated with personalised ones, so that for each view it is easier to know which file has to be uploaded or referred to. The names *New View reference* have not been updated to better exemplify the Reference mechanism, but Figure 9 shows that these names can also be personalized by updating the value of *Identifier* (that is used by the system to identify the view) and the value of *Display*, that is the name used by the system when displaying that view to the DL users.

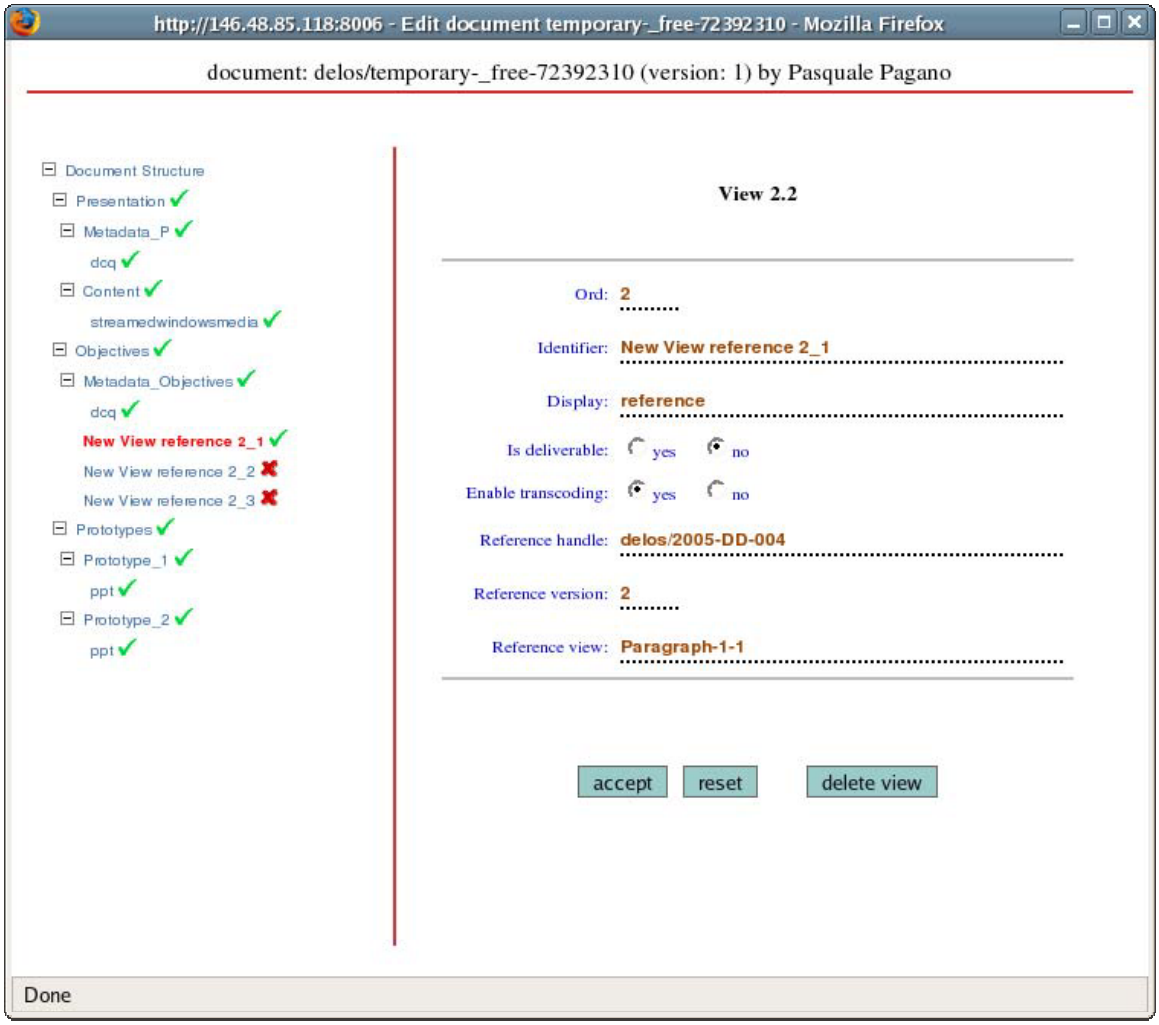


Figure 9: Inserting manifestations through the reference mechanism

3.2 – Administration Service

The ADMIN service allows the DL Administrator to have a general view of the status of all the documents submitted for publications by different authorities, and to manage them accordingly.
This service concisely displays the status of the different types of submission requests - Submissions, Editions, Corrections, and Withdrawals - by the different authorities (see Figure 10) and allows the Administrator to examine their detailed description and operate on any described document with different actions depending on the type of request, as detailed below.

- New Submission/Edition Requests: To manage submission requests the following operations are available: (i) View the detailed structure of the submitted document; (ii) Edit the document; (iii) Evaluate the documents and, accordingly, whether approve or reject the document or submit it to selected reviewers (see Figure 11). Option (iii) also permits the Administrator to select approved papers and automatically organize them in a collection such a Journal issue or a Proceedings.

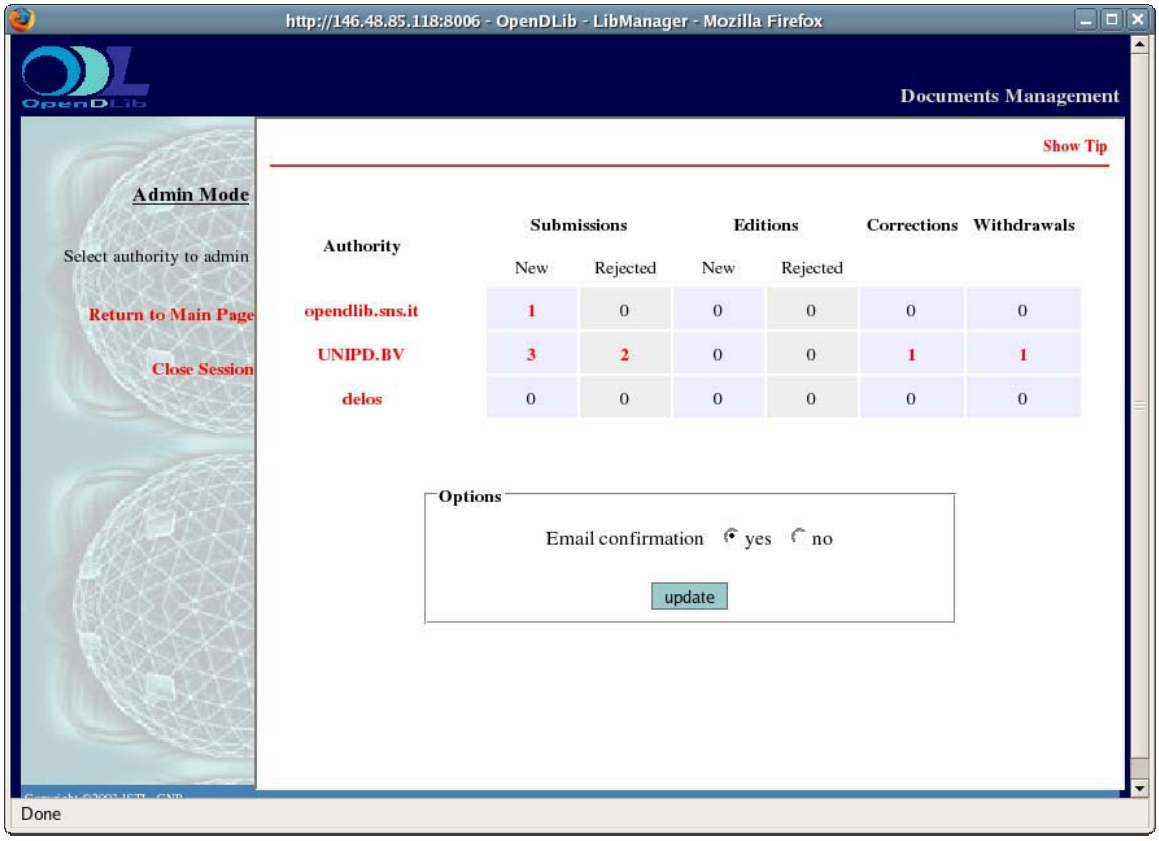


Figure 10: Displaying the status of the different submission requests

- Rejected Submissions/Editions. Documents rejected by the reviewers can be viewed, edited, unrejected or discarded while giving the user notice of these actions.

- Correction requests: To manage such requests the following operations are available: (i) View the detailed structure of the submitted document; (ii) Edit the corrected document; (iii) approve the correction, thus making the corrected document to replace the existing one; (iv) reject the correction, and give the user notice of this.

3.3 – Review Service

This service displays the list of the documents a reviewer is requested to review. Before filling in the review form supplied by the system, the reviewer can access reviews of the same document by other reviewers, if any.

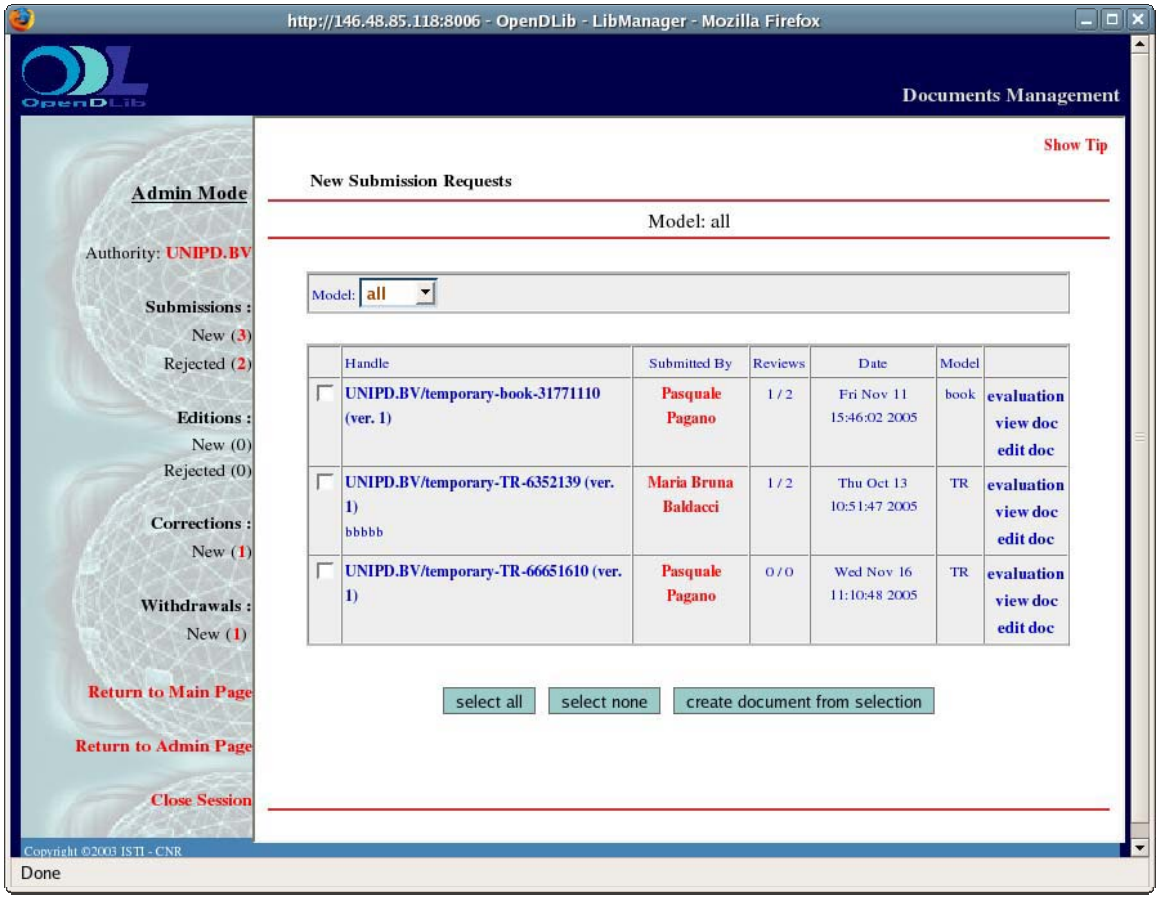


Figure 10: Options for managing new submission requests

4. Final remarks

The OpenDLib system has been able to exploit the DoMDL model by decoupling the definition of the document model instance from its real data, thus giving solutions to the following problems: i) how to describe the internal relations among document entities, ii) how to manage the related physical parts of each entity

With this approach a document is really composed by several files. The instance of the document model for a given document is expressed as an XML Schema and described in a separate file, named Structure file [7], which is the only mandatory element that must be provided. The goal of this file is to explain the composition and the relations among the other files that compose the document. It can put together different physical components to form an unique and coherent structured document. Different editions of the same document are not physically linked together in order to obtain a higher flexibility of the system. Rather they are logically grouped by the storage model which is able to manage them as a single entity since they share the same document identifier.

Finally, according to the document model specification, it is also possible to express a set of rules that regulate the rights on the document views via the properties child tag; in this implementation the rights to download, deliver, transcode or display a view may be, or not be, granted.

By exploiting the DoMDL model, the OpenDLib system is capable to offer new functionalities based on the following four main issues:

Document storage. According to the DoMDL specification, the storage model must be able to manage multiple metadata formats for the same document and multiple physical manifestations for the same view of a document. This makes OpenDLib capable (i) to create new Digital Libraries starting from existing heterogeneous information sources and (ii) to preserve documents from the technological obsolescence. Regarding manifestations, they are identified by URIs. A manifestation can be stored inside or outside the system, depending on the time in which the URI is dereferenced,.

The combination of the different options by which a manifestation can be directly uploaded, or automatically retrieved from an external location or maintained as an external manifestation makes it possible to build new structured documents that enrich the original ones by aggregating multiple parts of different documents from different heterogeneous information sources. Moreover, these choices promote an optimal utilization of the storage resources. This optimization is also supported by the reference view mechanism by which data duplication is avoided.

The last advantages to be noted are (i) the possibility to submit and manage documents that are modelled in very different fashions in the same OpenDLib instance, if they are compliant with the XML Schema presented; and (ii) the addressability, i.e. the granularity of documents that can be directly addressed or referenced. The basic addressable unit is the single manifestation. Moreover, the list of all views or manifestations as well as the list of editions of a document can also be addressed.

Document access. The access granularity, i.e. how a document or its components can be accessed, is closely tied to the storage model. OpenDLib is able to expose data according to the document model representation and to hide the representation and provide an interface to query the model in order to obtain the document parts.

Document discovering. Thanks to the document model, an OpenDLib instance can have multiple indexes able to index any format independently of their number or location. Also the graphical user interface provided to interact with the search subsystem has the capability to configure itself, depending on which index it currently interacts with, by automatically adding, removing or changing both its components and look-and- feel. In addition, the search subsystem offers the very new possibility to execute queries across documents handled by different information sources and expressed in different formats.

Document visualization. The visualization of documents is the last main issue strictly related with the document model. DoMDL gives a great number of opportunities for the presentation of complex documents. For instance, it allows document visualization be personalized by deciding who has the rights to view what.

References

1. Castelli, D., Pagano, P.: OpenDLib: A Digital Library Service System. In Agosti, M., Thanos, C., eds.: *Research and Advanced Technology for Digital Libraries, 6th European Conference, ECDL 2002, Rome, Italy, September 2002, Proceedings.* (Lecture Notes in Computer Science), Springer-Verlag (2002) 292–308
2. Castelli D., Pagano P. - A System for Building Expandable Digital Libraries. Joint ACM-IEEE Conference on Digital Libraries (Houston, Texas, May 2003). Proceedings, p. 335-345. Catherine C. Marshall, Geneva Henry and Lois Delcambre (eds.). IEEE, 2003
3. Castelli D., Pagano P. - A flexible repository service. The OpenDLib solution. elpub 2002. International Conference on Electronic Publishing (Karlovy Vary, Czech Republic, 6-8 november 2002). Proceedings p. 194-202. João Á. Carvalho, Arved Hübeler and Anna A. Baptista (eds.).
4. Candela L., Castelli D., Pagano P., Simi M.: From Heterogeneous Information Spaces to Virtual Documents. To appear in *Proc. of 8th International Conference on Asian Digital Libraries (ICADL 2005),* Bangkok, Thailand, December 2005
5. IFLA Study Group on the Functional Requirements for Bibliographic Records: *Functional Requirements for Bibliographic Records.* Final Report. München : Saur, 1998
6. Dublin Core Metadata Initiative: Dublin Core Metadata element set, version 1.1: Reference description. (http://dublincore.org/documents/dces/)
7. OpenDLib: DoMDL XML Schema. (http://www.opendlib.com/resources/schemas/domdl.xsd)

Grey Literature, institutional repositories, and the organisational context

Simon Lambert, Brian M. Matthews and Catherine Jones
CCLRC, Rutherford Appleton Laboratory, United Kingdom

Abstract
A wide variety of grey literature is produced during the work carried out at a large multidisciplinary scientific research organisation. This paper examines how the grey literature may be managed and the benefits that result. Trends in Technical Reports, which have always been an important medium for transmission of knowledge, are examined, and the use of an institutional repository is advocated for the future of the medium. Other kinds of grey literature produced in research projects are also described, and again the institutional repository is seen as an important mechanism for preserving and making accessible the knowledge they embody, particularly if it is coupled with other information systems in the organisation.

1. Introduction

CCLRC, the Council for the Central Laboratory of the Research Councils, is one of Europe's largest multidisciplinary research organisations supporting scientists and engineers world-wide. As well as participating in setting the priorities for the UK's science needs, it operates three research laboratories: the Chilbolton Observatory in Hampshire, the Daresbury Laboratory (DL) in Cheshire and the Rutherford Appleton Laboratory (RAL) in Oxfordshire, and employs 1800 people. These laboratories offer facilities and expertise including ISIS, the world's most powerful pulsed neutron source; high-power lasers; space science technology including satellite and ground-based instrumentation; and information technology.

The operation and development of CCLRC's facilities—which set the organisation apart from most universities—require a high level of specialised knowledge and expertise that has been built up over many years. Additionally, research is conducted by members of staff and users of the facilities. Collaborative projects are undertaken, sometimes of a very large scale: for example, the Space Science and Technology Department was strongly involved in the Mars Express and Venus Express planetary missions. In information technology, there is a history of many years of international collaboration in European research and development projects, involving partners (companies, research institutes and universities) in many countries and producing software, standards and academic publications.

It can be seen that CCLRC is a knowledge-intensive organisation with some special attributes and requirements. Grey literature, as well as literature for publication, plays a key role in its business, and this paper will study that role, link it to the context in which the business is conducted, and highlight the importance of CCLRC's institutional repository in supporting it.

2. Technical reports as grey literature

CCLRC and its predecessors have produced Technical Reports since the organisation was founded. These are formally published by the organisation and are deposited, according to UK law, with the National Deposit Libraries. The format was designed to capture the pre-refereed version of journal articles or to capture technical details for posterity and dissemination. The first point was especially relevant for particle physicists, as rapid dissemination of information was, and still is, important to advance the boundaries of the field. The second is a recognition that not all the knowledge gained from the scientific process is suitable for publishing in the scientific journal record but is still valuable. As one of CCLRC's main roles is the building and maintenance of large-scale scientific facilities, the organisation has detailed technical knowledge in very specialised fields. Building these facilities does not happen very often, but for example information contained in technical reports for the Neutron Spallation Source (ISIS) are being used twenty years later to aid the construction of a Second Target Station.

However both these roles have been affected by the changes in information dissemination that have occurred with the rise of the Internet. Pre-refereed paper versions have been completely overtaken by e-print repositories (such as arXiv[1]) and the print form is now redundant. Furthermore, the internal processes for producing technical reports have been a barrier in this electronic age and the format has been in decline. Figure 1 shows the trends in the production of formal reports over a ten-year timespan.

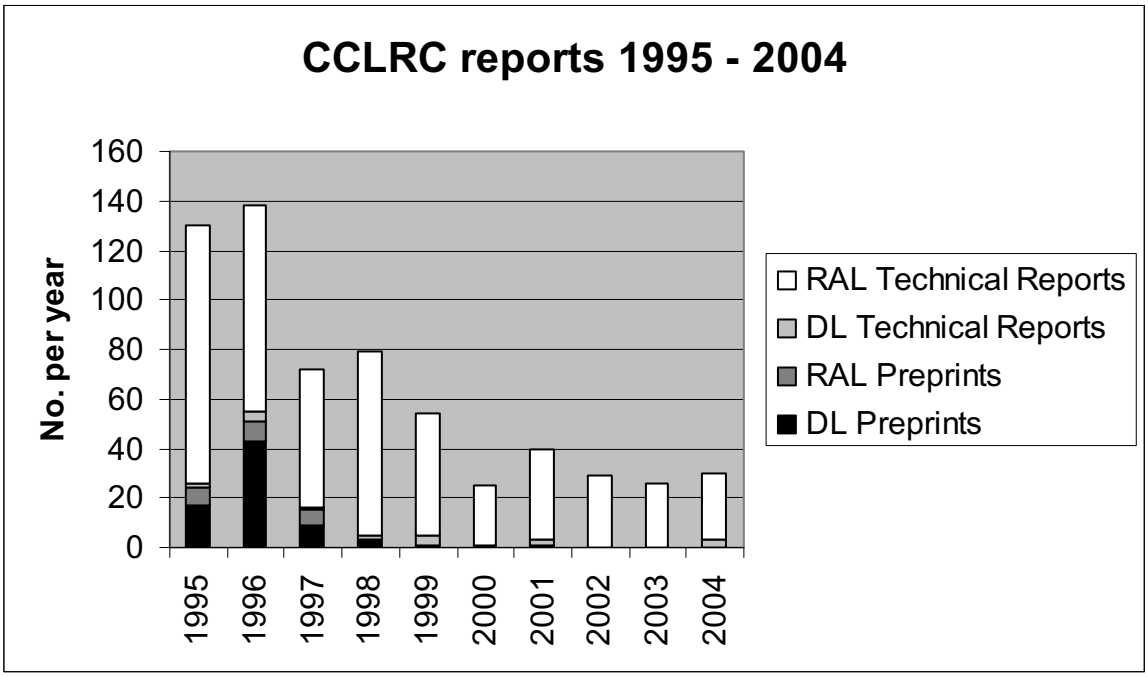


Figure 1: The decline of formal Technical Reports at CCLRC

The production of the technical knowledge has not stopped, but the barriers of the internal process have made it more attractive to produce 'grey' versions of the material. This limits the dissemination and resource discovery potential of the material, especially if it now resides in a variety of systems such as internal file stores. It can also lose the clarity of versioning: is this copy in location A the same as this copy in location B?

3. Reclaiming Technical Reports through an institutional repository

The previous section discussed some of the reasons that Technical Reports have been encouraged to go grey. It is a contention of this paper that by using an institutional repository the organisation and wider community can 'reclaim' Technical Reports. An institutional repository is a formal and managed archive of research output in the form of digital documents that is operated by a particular institution such as a university. In some cases the focus is only on e-prints[2]—pre- or post-publication versions of refereed papers. However, the remit of CCLRC's institutional repository, known as ePubs[3], is the scientific and technical output of CCLRC; it does not distinguish in its collection remit between formally published or grey literature. It is concerned with the dissemination of the knowledge discovered by CCLRC and the users of its facilities rather than the final location of the information.

By encouraging technical report writers to deposit their material within ePubs, and making the internal process recognising that work as a technical report more efficient and effective, it becomes more visible and useful to the wider community. ePubs is committed to the long term preservation and curation of Technical Reports and by depositing such work in ePubs then its long term future is assured. Further, ePubs uses the IFLA Functional Requirements for Bibliographic Records to represent the work and its version relationships. It is possible to distinguish between different versions to bring clarity back into this area. This is discussed further in section 5 below.

4. Varieties of grey literature

Technical Reports exemplify a type of literature with a potentially long useful life becoming progressively greyer as time goes on. The longevity of this material suggests a refinement to the established definition of grey literature for a research institute like CCLRC. The established definition is: *'Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body.'*[4] It is accepted that this definition needs narrowing for particular contexts, and such a refinement is: *'information produced in a specific working context which is, or might be of value outside that context.'* This additional specification is intended to capture the idea that the material, though not intended for publication in the formal sense, may profitably be 'published' to different working contexts.

There are many other varieties of grey literature produced at CCLRC, but one that is worth singling out is that produced by collaborative R&D projects in the form of deliverable documents, working reports, etc. Normally the distribution of these documents is specified by the contract under which the project is conducted, but it will usually be the case that some of the documents have unrestricted distribution.

Furthermore, they embody valuable knowledge that might be applicable in wider contexts. For example, final reports of one project may serve as input to the baseline for the next project, or surveys of products or state of the art produced for internal decision making may retain their value for some time and have a wider applicability beyond the project that created them.

Figure 2 shows a schematic (and incomplete) depiction of parallel processes in an organisation such as CCLRC, and the ways in which grey literature is generated and consumed among them. These processes are represented as coarse-grained work flows, which, while not constraining the activities of scientists or managers, describe the stages that they pass through in a particular project.

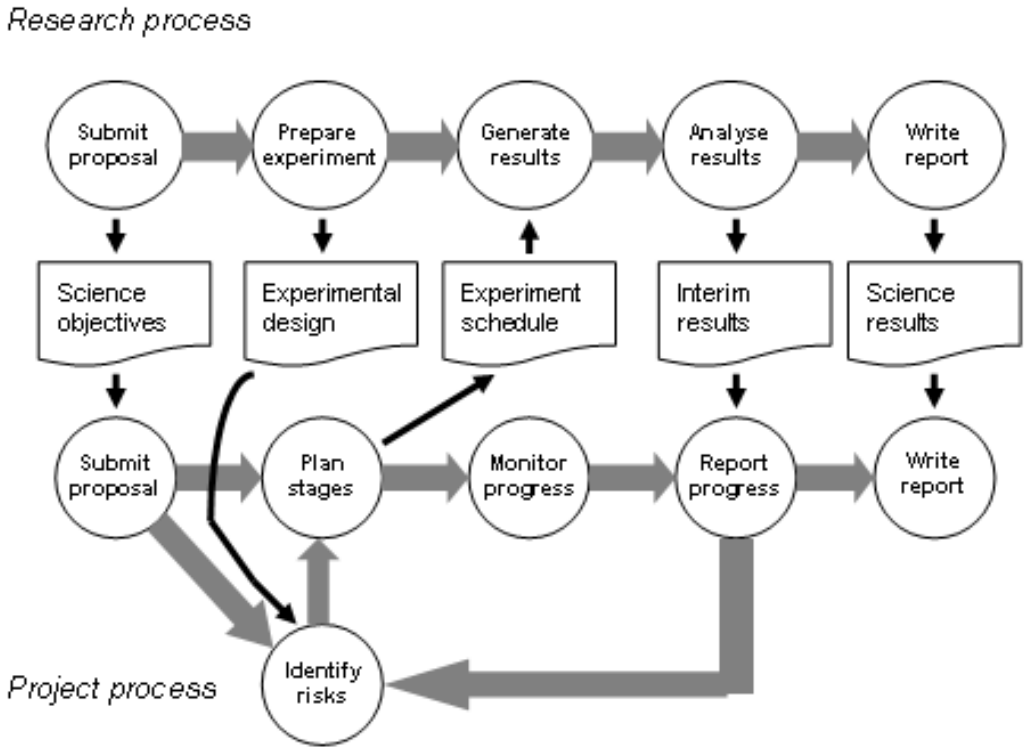


Figure 2: Exchange of grey literature between processes

The research process is what drives the production of scientific knowledge, and interlocked with it is the project process, which can be understood as part of the organisation's *business process* since it is concerned with receiving and using resources with which results are returned and disseminated.
These processes interlock in the sense that outputs from one stage serve as inputs to another stage in a parallel process. The analysis of results feeds into the progress monitoring; the scientific report forms part of the final project report; and so on.
It is also possible to characterise in a similar fashion other processes such as the digital curation process[5] and the publication process.

Inasmuch as the output of the stages is recorded in some form, it is by definition grey literature. As noted above, an additional specification may be introduced that the information is, or might be of value outside the context in which it was produced. This can mean two things:

- the information is useable elsewhere in a parallel process (for example, transfer from the research process to the project process);
- the information is useable in a different (later) process instance (for example, in a different project).

Figure 3 shows a variant of the diagram in which the research process is now concerned with collaborative development of software technologies (as opposed to the experimental science represented in Figure 2) and grey literature is generated that may be of value in other such projects. The broken arrows indicate such feeds into the future.

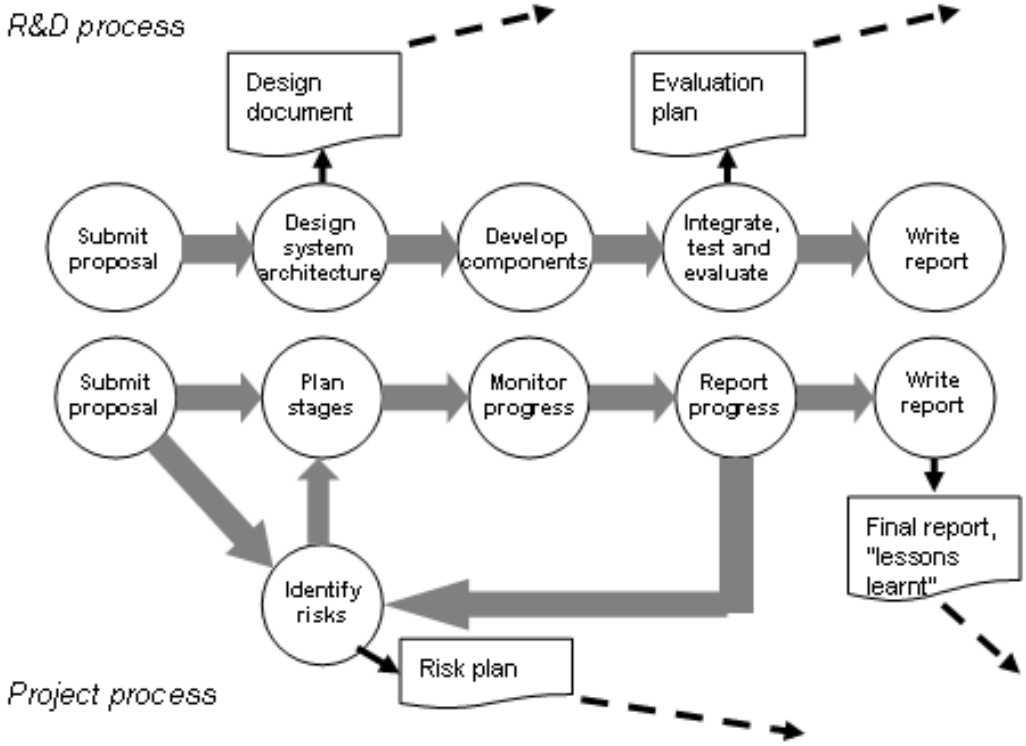


Figure 3: Grey literature available outside the immediate processes

5. Grey literature and the institutional repository

It is a contention of this paper that some of the grey literature produced within the organisation may be profitably stored in an institutional repository, so as to enable the kind of reuse outlined above. Particularly relevant are the institutional repository's capacity for long-term storage and for clarity in versions of material submitted. At CCLRC a programme of digitisation of Technical Reports is under way, with the ePubs repository playing a central role in storing and indexing the documents. This is part of the effort to have ePubs as a long-term archive for curation and preservation of the research output of CCLRC. The deposit of some project documents such as final reports is encouraged.

Regarding versions, ePubs partially implements the IFLA Functional Requirements for Bibliographic Records[6] model. This model specifies four levels of description for any particular work. The *work* itself is an abstract concept of a distinct intellectual or artist creation. A particular work can be realised as an *expression*, or series of *expressions*. Each expression will be physically embodied as a *manifestation*, or series of *manifestations*. A single exemplar of the manifestation is known as an *item*. ePubs implements the Work, Expression and Manifestation elements at present. Figure 4 shows an example of different expressions of a single work: in this case, a paper as published in conference proceedings, and the PowerPoint presentation that was actually shown at the conference.

Publications

Publication Type	Reference	URI	Local file(s)	Year
Presentation	*9th DELOS Network of Excellence thematic workshop: Digital Repositories: Interoperability and Common Services, The Foundation for Research and Technology - Hellas (FORTH), Heraklion, Crete, 11-13 May 2005*	http://www.ukoln.ac.uk/ev...	jones-mascord.ppt [Microsoft Powerpoint, 707 KB] [42 accesses]	2005
Paper in Conference Proceedings	Proc. *9th DELOS Network of Excellence thematic workshop: Interoperability and Common Services, Heraklion, Crete, 11-13 May 2005*		delos-jones-20050425.pdf [Adobe PDF, 134 KB] [19 accesses]	2005

Figure 4: An example of different expressions of a single work in ePubs

Of course storage of the grey literature is only one half of the problem. It is also necessary to retrieve it effectively according to context. There can be no substitute for human expertise here, but the expertise

can be augmented by suitable metadata describing content and the working context in which it was produced. Another contention of this paper is that the institutional repository could and should be integrated with other information systems of the organisation, so as to provide some of the context needed for retrieval. These systems include:

- personnel systems;
- financial systems;
- project management/monitoring systems;
- security infrastructure.

Such integration can offer benefits including accuracy of information (for example, allocation of publications to individuals), quality control (ensuring that all necessary stages are met), efficiency of metadata collection (for example, collecting metadata at project application stage and tracing through), adherence to policies and procedures (for example, approvals for paper submission), traceability and accounting, and report generation (for example, for performance assessment).

Integration of this kind will require integration of repository metadata with metadata associated with the other systems (HR-XML, CERIF, PRINCE 2 standards, …). Moreover, to assist with retrieval, it is also necessary to describe the stage in the processes at which the document was produced, and to characterise those processes themselves in terms of the domain (scientific specialism). There are two dimensions of similarity of context: the content itself (for example, documents relating to data mining or to neutron scattering spectrometers) and the stage in the process (for example, a risk assessment, or a final report). The ePubs system currently supports the first of these through optional keywords associated with entries, though not the second, which has clear links to the project process.

6. Conclusions

An institutional repository, being a central point within the organisation for literature and data, is a component of the integration of processes, which promises benefits both to the organisation itself and to the researchers within it. Grey literature can be recorded and retrieved according to accurate and up to date personnel, project and business unit structure, leading to benefits in accuracy of reporting, quality control, etc. From the researchers' point of view, the repository assists them by storing multiple expressions or manifestations for the different parts of the process, and so brings clarity to work and version relationships. It also allows retrieval of grey literature from other projects that might otherwise be invisible to those who could benefit from it. In short, the institutional repository is integrated as a part of the overall institutional memory.

References

1. arXiv.org e-Print archive, http://www.arxiv.org
2. Supporting Open Access, http://www.eprints.org
3. Jones, C., Mascord, M.(2005), 'Experiences of building an Open Access Institutional Repository in a UK Scientific Research Organisation', Proc. 9th DELOS Network of Excellence thematic workshop: Interoperability and Common Services, Heraklion, Crete, 11–13 May 2005. To access the ePubs system see: http://epubs.cclrc.ac.uk
4. Definition of grey literature at http://www.greynet.org, accessed November 2005.
5. Open Archival Information System (ISO 14721:2003): Reference Model for Data Curation
6. IFLA Study Group (1998), *Functional Requirements for bibliographic records: final report*, Saur, Munich, http://www.ifla.org/VII/s13/frbr/frbr.pdf

Author Information

Simon Lambert is a member of the Information Science and Engineering group in the Business and Information Technology Department at CCLRC Rutherford Appleton Laboratory. He has been working in the field of knowledge-based systems and related areas for many years, first with the companies SPL and Systems Designers, and then at CCLRC. He has been the CCLRC project leader on a range of research and development projects, including many under the European ESPRIT and IST programmes. In these roles he has conducted requirements analysis, knowledge acquisition, system design and implementation, user trials and the writing of documentation. Currently his interests include knowledge management and environmental information systems. Most recently he was the overall project coordinator for the IST project Pellucid, which produced a platform for experience management systems for public employees. He is currently involved in the European project CISTRANA, developing a Current Research Information System for national research activities across Europe. Email: S.C.Lambert@rl.ac.uk

Catherine Jones was born in the UK in 1964 and graduated from Thames Polytechnic with a degree in Computer and Communication Systems. She joined the Rutheford Appleton Laboratory as a Database Systems Analyst/Programmer and worked on internal management information systems. In 1994 she transferred to the CCLRC Library and Information Service as Systems Librarian and was responsible for Library IT strategy, policy and developments. She is now the Library Systems Development Manager. Since 2003 she has been the project manager for CCLRC's Institutional Repository.

Entering Grey Waters: Challenges and Solutions of Providing Access to Non-traditional literature in an Aquarium's library

Marisol Ramos-Lum and Steve Vogel
Virginia Reid Moore Marine Research Library - Cabrillo Marine Aquarium, CMA, USA

Introduction

The seventy year old library collection at the Cabrillo Marine Aquarium (CMA) has grown steadily from its humble beginnings as the original collection of the old Cabrillo Marine Museum at the Cabrillo Beach Bathhouse in San Pedro, California. Until recently, the library was run by the Education/Collections curator and assisted by one volunteer who helped to catalogue the collection using the *Library of Congress Subject Headings*, did inventories, data entry and other library-related activities. In 2004, the *Virginia Reid Moore (VRM) Marine Research Library* was opened as part of a CMA Expansion Project and a librarian was hired full-time to administer and manage the library and two part-time assistants have been hired to assist running the library.

The VRM Library is a small collection of about 2,000 + books, mostly in the marine sciences (marine invertebrates, ecology and fishes) and 40 titles of donated journals and magazines, of which about 5 are current paid subscriptions (*California Fish and Game Bulletin, Fishery Bulletin, Marine Mammal Science, Nature, and Science News)*. But, our strength lies in the grey area beyond books and journals, where a large quantity of useful information resides. One of the unique characteristics of the VRM Marine Research Library collection is the wide variety and quantity of *grey literature* and *ephemera* materials from a wide variety of sources: local environmental organizations, historical societies, governmental reports, other aquariums publications, etc… These documents have valuable and unique information which often is not found elsewhere. These materials are vital to the VRM collection since they are rarely collected by traditional libraries and offer a rare view of non-traditional materials by non-profit, private and community organizations with common interests with the Cabrillo Marine Aquarium mission of research and education about the marine life of Southern California.

Unfortunately, these materials are inaccessible to users. Although the materials have been finally relocated to the library, they are still housed in boxes without indexes or any other access point. To begin to address this problem, the library requested that the aquarium administration invest $4,500 to acquire high-density file cabinets to help preserve and provide access to this material. The cabinets have been delivered and installed in the library but the biggest challenge remains—indexing and making this collection accessible. Our goal is not only to secure the proper storage of these materials but also to centralize and supply access to these valuable resources. To achieve this goal, a three-pronged strategy was designed to give access and further develop our grey literature and ephemera collection: indexing and thesaurus development; database with online search capabilities and outreaching to local and regional organizations to update and expand the current collection.

This paper will discuss the challenges of collecting and preserving grey literature and ephemera in the marine science knowledge-domain, and the design, execution and expected outcome of this VRM library initiative.

Swimming in Grey Waters: Challenges and Solutions to Manage Grey Literature and Ephemera

Grey literature and ephemera are discussed in this presentation, because both types of publications are part of the collection and are worth preserving for their rich information content and uniqueness. Grey Net defines *grey literature* as "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing."[i] *Ephemera* is a term that has been defined as: "Items manufactured for a specific, limited use, and usually intended to be discarded thereafter, especially printed matter on paper."[ii] This definition does not convey the reality of most ephemera collections, since indeed instead of being discarded a great quantity of ephemera is collected and preserved. For the purposes of this paper, ephemera are defined as:

> Materials (physical or electronic) that regardless of appearance, quality or quantity, and that at some point were considered disposable and of little value or no value, through time, had become valuable in such a way that it had broadened their appeal and made them desirable to be collected and preserved by individuals, collectors and information institutions. These materials become the non-traditional, alternative evidence of mainstream and non-mainstream groups in society. (Ramos-Lum 2002, 9)

CMA has historically gathered and accidentally retained considerable quantities of reprints, clippings, curriculum guides, brochures and the like. Our diverse staff receives and shares proceedings and publications sent to them by colleagues from like-minded institutions and agencies. It has been beneficial for us as an institution to have materials such as these filed under subject headings or by institution or simply mainstreamed into our shelved holdings. When working on a given topic it has largely been beneficial to have similar materials co-located. In many cases, items such as reprints, posters and multi-media materials do not fit in easily. The main use for much of our grey literature has been for educational purposes. We temporarily display current newspaper clippings in our exhibit hall and afterward file them by subject. Folders by subject are used for docent training and as reference material for exhibit development and use in our publications. Articles stripped from duplicate magazines and filed by subject add to this resource and make it easy to browse a subject or organism instead of tracking down isolated shelf resources in various publications.

Our director has requested that files be kept on related institutions including examples of their printed materials at all levels from visitor guides to booklets, pamphlets, annual reports and newsletters. These materials have helped us develop content, program and design ideas as well as providing networking opportunities when we can become aware of similar initiatives at other organizations.

Because of the importance of these materials to the aquarium a wide range of materials have been collected and continue to be acquired. This situation presented three different challenges for the Education Curator and the Librarian:

- Determining what we really have (Inventories)
- Deciding what we want to collect from now on (Collection Development)
- Making these materials available to our users (Access)

Inventories

As a way to handle the first challenge, since May 2005, the library staff started conducting several inventories to get a sense of the size and variety of subjects in the collection. We have not finished these inventories but here are some partial results:

Table 1 (Partial Inventory totals) 2004-2005

Type of Materials	***Totals***
Newsletters	72 titles
Reprints	51 letter size documents boxes
Newspaper clippings binders	15 binders
Proceedings	10 titles
USA Gov. Reports	13 titles
State/Local Reports	14 titles
Environmental Reports (non-governmental)	9 titles
Cabrillo Master Plan	5 versions
CMA architectural proposals	10 versions
Catalogs	50 titles
Workshops/Training materials	5 titles
Miscellaneous (brochures, flyers, CDs, stripped magazine articles, etc…)	?

The inventories have been hard to do because most of the materials were catalogued as unique items in the regular book collection, even though they were part of sets of reports. While converting the old library book collection to a new library cataloguing system, we started taking out and placing together these different sets of reports, surveys and other grey literature materials. Other reports had not been catalogued yet and were stored in boxes and moved from the old library to the new library space. These partial inventories help us to see the patterns of collecting during the last 20 years, and inform one of our first solutions to handle the acquisition of these materials—a collection development policy for the library.

Collection development policy

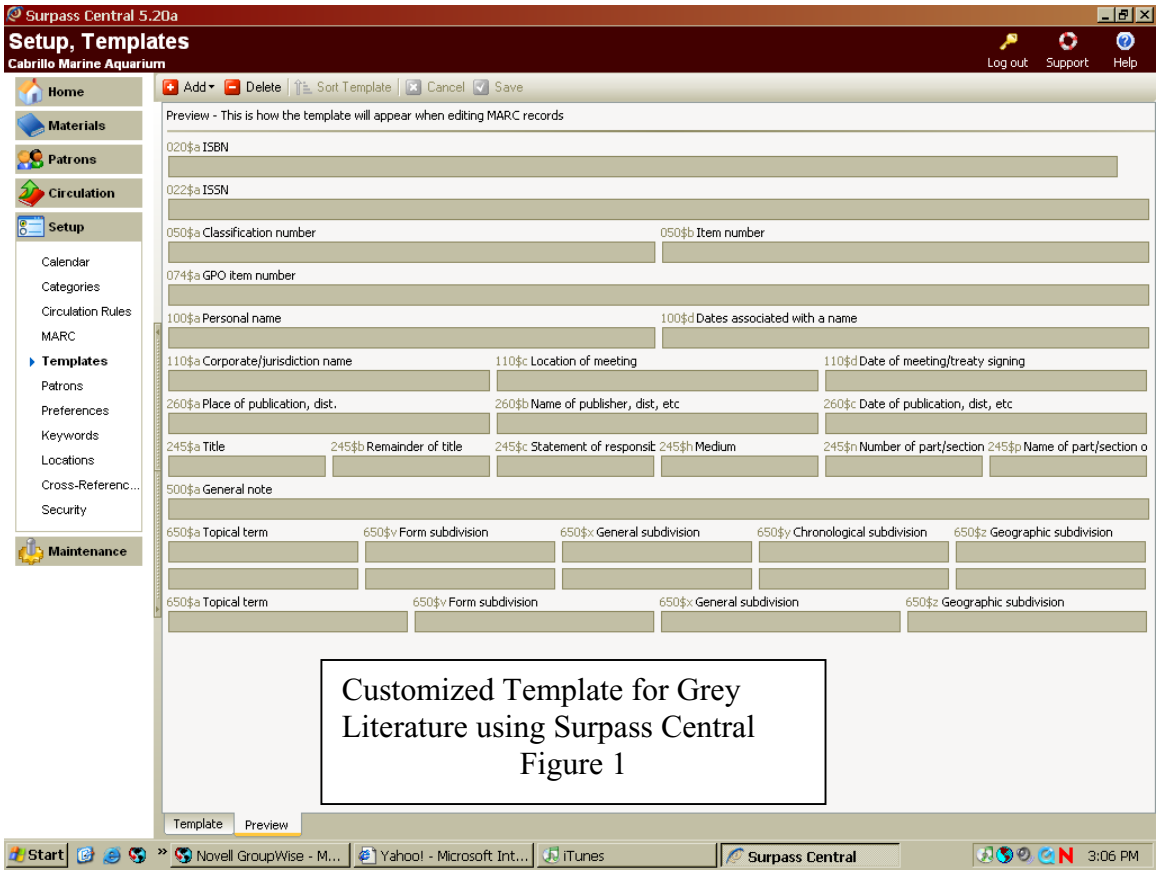
Until last year (2004), the library never had a collection policy that specifically identified what type of materials were acquired for the library but this year (March 2005) we submitted such policy and put in writing was has been the "unofficial policy":

> The following types of materials are acquired [for the library]: serials, monographs, books, newsletters, videos and other media, proceedings/transactions, technical reports, government documents, theses and student projects. (VRM Collection Development Policy 2005)

Our biggest sources of grey literature come from newsletters and reports (both governmental and non-governmental). We are starting to collect student projects to support the work of one of the new spaces at the aquarium, the Aquatic Nursery, and it is our hope to expand this service to other departments at CMA as a way to record the work done by student interns doing research at the CMA. We are still working to improve this policy, each year it will be revised to adapt it to our needs.

Access

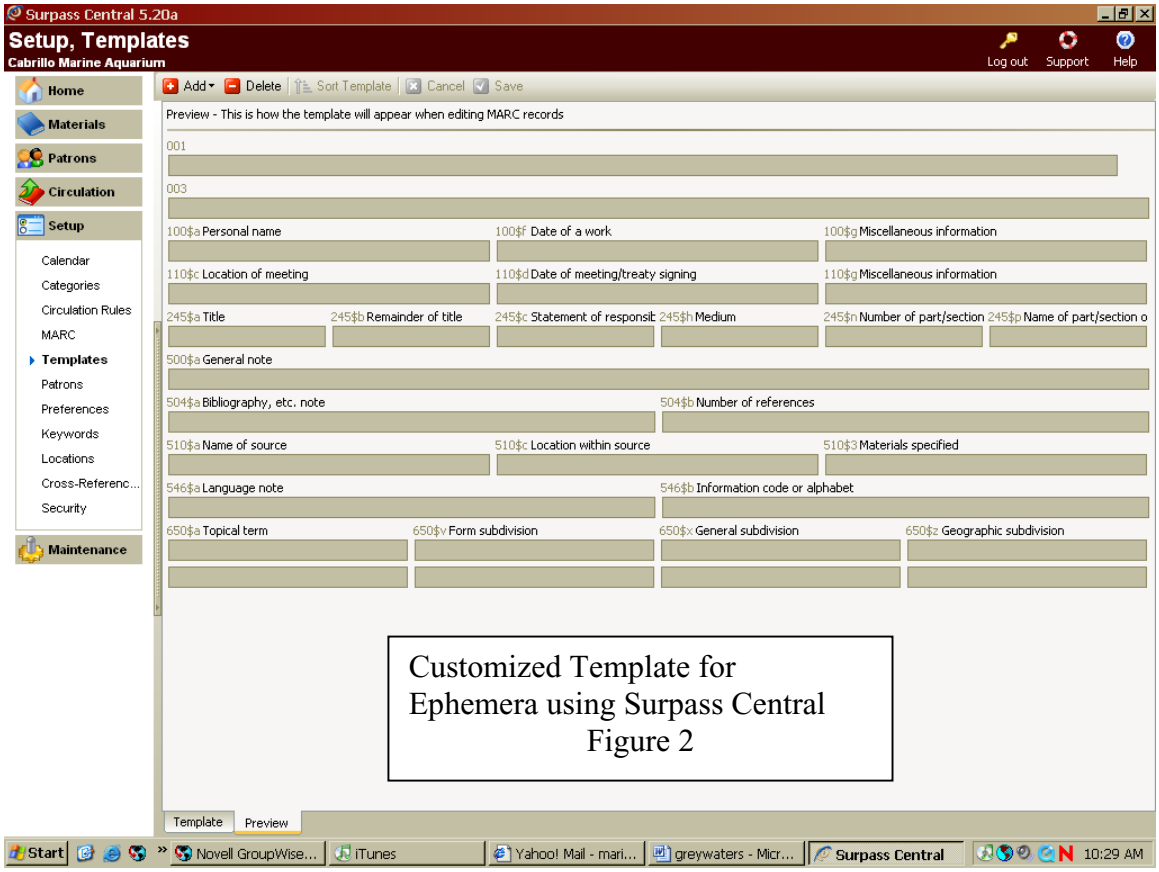
The biggest challenge left to our library is to make our grey literature and ephemera collection available to our users. Ideally, we would like to have one database that could manage all the materials that the library holds. Most of the technical reports, research studies and other governmental materials are not hard to add to our cataloguing database. Surpass Central, our library cataloguing and circulation program, allows the creation of customized templates and we are testing one for Grey Literature (See Figure 1).



Customized Template for Grey Literature using Surpass Central
Figure 1

Ephemera materials are very important for our organization and we have accumulated a wide array such as aquarium brochures, flyers, pamphlets, clippings, online news reports, and finally, reprints (sent by the authors to different staff

members, or received as donations from retired professors and scientists). As with grey literature, we will be testing a new template to capture this information and make it available through the library online catalogue (See Figure 2).



Customized Template for Ephemera using Surpass Central
Figure 2

The ultimate goal of adding these records for grey literature and ephemera is to make these materials searchable online. Ideally, we will be able not only to search books and journals, but also our subject-based collections. We want to make the staff and other outside users aware that we also have reprints of articles that may otherwise be "lost" inside journal issues. To be able to know what we really hold, we really improve our ability to serve all our stakeholders (staff, volunteers, docents, Friends' members and the general public).

Preservation

Based on our preliminary inventories we found a lot of preservation issues that we need to address to be able to protect our collection for future generations. Our biggest challenge is dealing with acidic materials such as clippings binders, stripped magazine articles and reprints. As we started the process of cataloguing these materials, we will start making preservation copies of the most endangered materials. This is not a difficult process but it is a very time consuming one. As part of the inventory, we are identifying the materials that need preservation or conservation measures and as time and resources allow it, we will continue the preservation process.

Conclusion

Making our collection of grey literature and ephemera accessible is an ambitious goal and we are still in the developmental phases of it. Computer programs do not always perform as advertised and there are many other challenges related to the basic problem of cataloguing non-book, non-mainstream materials (how to assign meaningful subject headings to improve browsing and a good cross-reference system especially since these materials can be catalogued and searched by more than one heading). We still need to ask ourselves what access point we should use when creating these records. We have just started indexing the reprint boxes and we hope to develop a thesaurus to facilitate searching. Selection is the key for a more coherent and comprehensive collection. Our challenge is to make selection a more conscious decision, based on our collection development policy, instead of a passive reaction. We had already started this process when we decided to engage in this discussion at this conference and we look forward to the future, when more people can enjoy these amazing resources.

References

Ramos-Lum, Marisol. (2002), The Ephemera Dilemma: When Definitions Fail to Describe their Objects- the case of Science Fiction (SF) Fanzines - *Unpublished paper*: pp. 1-27.

Virginia Reid Moore Marine Research Library (2005), Policies Part 1, Collection Development Policy - Cabrillo Marine Aquarium pp. 1- 7.

Endnotes

[i] Grey Literature Network Service 31 October 2005 http://www.greynet.org/pages/1/index.htm

[ii] Art and Architecture Thesaurus 31 October 2005 http://www.getty.edu/research/tools/vocabulary/aat/

Author Information

Marisol Ramos-Lum is the Librarian at the Virginia Reid Moore Marine Research Library, Cabrillo Marine Aquarium (CMA), San Pedro, California. She obtained her MLIS degree at UCLA with a specialization in Archival Sciences. As an archivist, Marisol has worked with Grey literature and Ephemera collections description and accessibility issues and working at CMA has been a great opportunity to continue exploring these issues. Email: MLum@rap.lacity.org

Steve Vogel is Education and Collections Curator for Cabrillo Marine Aquarium, San Pedro, CA. His acquisitive nature and UCLA biology background have added considerable diversity to our library and specimen collection holdings over 20 years. Cataloguing and other library disciplines are not his training or forte and CMA is fortunate to have a librarian/archivist on our team.

Yizkor Books as Holocaust Grey Literature

Faith Jones
New York Public Library, NYPL, USA

Gretta E. Siegel
Portland State University, PSU, USA

Abstract
Yizkor is a Hebrew word meaning 'you will remember', and Yizkor books are books that commemorate the vanished communities destroyed by the Holocaust. As the 60th anniversary of the liberation of the German concentration camps is commemorated this year, it seems fitting, with this conference being held in Europe, to call attention to this unique and interesting body of literature.

Yizkor books present an historic but interesting type of grey literature, with significant modern day interest. While the Yizkor Book had its origins in the 13th century, it re-emerged early in the 20th Century as a tool for documenting the declining villages and Jewish communities of Eastern and Western Europe. Following the Holocaust, these books were put to a new purpose in their memorialization of the destruction of Jewish life in Europe's cities and villages. These books originated as manuscripts, collected in bits and pieces by fraternal organizations and mutual aid societies, and were eventually printed non-commercially in very small editions in the 1950's, 1960's and 1970's. Like the world population of Jews, the collections of these books are diasporic; scattered throughout the world. They are of current interest to a variety of scholars and others, for example – as tools for genealogical research and study – both for families of those who perished and for genealogists; to historians as primary sources- rich with detailed information about the past; to sociologists and anthropologists, as studies of a segment of European village life; to psychologists with an interest in the documentation of memory, and, it is hoped, to scholars of information creation and dissemination, as an addition to the canon of grey literature.

This paper approaches Yizkor books from all of these angles, and focuses on contextualizing them as Holocaust-era grey literature. The creation of these books, their publication and distribution history, and collection development efforts are discussed. Like the vanished communities that they describe, these books themselves have become endangered due to their small production numbers and various preservation challenges. Currently, there are digitization efforts underway to preserve and to increase open access to this genre of literature. These efforts are described and discussed.

What Is Yizkor, What Are Yizkor Books, and What Makes Them Grey?

Jewish law teaches that death is to be acknowledged communally, and this tradition of collective mourning encompasses a wide range, from the loss of an individual to the loss of a homeland or location. The rituals for collective mourning are, to a large extent, prescribed and codified. When a loved one dies, and on the anniversary of the person's death, a prayer, called the Mourner's Kaddish is said in the presence, and with the support, of one's community, defined as a minimum of ten community members. During the year, with the observance of certain holidays, there are special services, called Yizkor Services, which serve the purpose of communal mourning for the loss of loved ones, for the loss of the first and second holy temples in biblical times, and for the losses suffered in the Holocaust. There are many other mourning rituals and practices, which serve to re-establish a world which has been radically de-stabilized by a death (1). The term 'yizkor' is Hebrew for 'you will remember'. During a Yizkor Service, individuals are remembered by name, with the names of the deceased generally spoken aloud. The origin of the yizkor book, the topic of this paper, can be traced to this tradition. Yizkor books can be described as a Jewish communal response to the Holocaust. There is little published about the actual production of this genre of literature, and most of what we know comes from an oral tradition. It is hoped that this paper will make a useful contribution to the literature on this topic.

The first yizkor book, or 'memorbuch' as it was called in old German, was written at the end of the 13th century as a record of the individuals and communities slaughtered during the 200 years following the First Crusade through Central Europe. The original purpose of the yizkor book was liturgical. By including lists of names, places, and dates, it allowed people to remember and to state specifics in the Yizkor prayers, and to acknowledge that these people lost their lives specifically because of their faith. The post-Holocaust yizkor books – the focus of this paper, are descended from this original one from the medieval era.

Yizkor books fit both the definition and the 'spirit' of grey literature, and possess many characteristics of materials that we generally think of as grey. Aside from being produced non-commercially, much grey literature can be described as being data driven. In spite of the potential for very large audiences via the web, most of what is generally considered to be grey has a very small and specific intended audience. Additionally, while some grey literature is peer-reviewed, much of it does not

go through a rigorous review process. Other characteristics of pre-digital age grey literature are very small print runs and, relative to commercial publications, poor production quality. Yizkor books share all of these characteristics.

After World War II, the Jews who had dispersed to the various corners of the earth began to find each other and to form communities of survivors. These communities created survivor organizations or joined with existing ones, called Landsmanshaften, that served as 'benevolent societies' helping people to re-settle and to connect with other surviving family or community members. These Landsmanshaften, from the U.S., Canada, Israel, Argentina, France, and many other countries - posted advertisements in newspapers around the world looking for other survivors from their villages, and asked them to send in the names of those who had perished. This information provided the necrology, the 'core data' necessary to create the lists that could be used for Yizkor services. The ads also asked people to send in profiles – of important people from the their village, but also of commercial, social, cultural and political activities of the village. In addition to this information, people sent in maps, geographic descriptions, and stories. In the field of history, there is often an 'official history' which may or may not be followed by revisionist histories. In the case of the yizkor books, all of these 'histories' are 'revisionist', because the histories of the towns were not documented until the towns no longer existed, or if they were, those written documents were destroyed. Because of this, not all of the descriptions of the former life of the village agreed with each other. At times, the editorial committees themselves would have disagreements, often along ideological grounds mimicking pre-war divisions (Zionist vs. Bundist, religious vs. secular, etc.) and for these reasons, there may be multiple yizkor books published for a single village, or, if the disagreements were not too strong, simply sections of the books that may not completely agree with others. In general though, the books are a combination of lists of names and detailed descriptions of community life, giving a multi-faceted view of not just the people who vanished but of the entirely vanished communities, in all of their social complexities.

The concept of 'memory' plays an interesting and important role in this story. One of the purposes of the Holocaust was to destroy memory, yet one of the strongest foundational tenets of Judaism is the belief in text as a force for perpetuating memory. It is this deep-rooted belief, coupled with the tradition of trying to 'restabilize' a radically altered world that likely was the impetus for people to respond to the Holocaust with the outpouring of information that made the yizkor books possible. According to Hall in her paper on the politics of memory, this response embodied a 'crucial moral role to play within the historical record' (2). She goes on to discuss how first hand accounts are essential to historical research and how empathetic identification, when used in conjunction with objective historical accounts give broader dimension to the experience.

When a Landsmanshaft had collected enough material, it assigned editors to cull through it all and to produce these books, originally in the form of manuscripts, and later published informally in editions with print runs no higher than 1000. The books were produced in whatever languages material was submitted in, thus many of the books are multi-lingual, including Polish, Russian, Hebrew, Yiddish, German, Romanian, and English. The size of a yizkor book ranges from 40-900 pages. The intended audience for these works were neither the general public, nor Holocaust scholars, but were the survivors themselves as part of the grieving process, whether private, communal, or liturgical. The books were printed **by** and **for** survivors and émigrés from a given town or village. Not all towns have a yizkor book. By a calculation done by the New York Public Library, about 850 of them do have at least one book that was developed in this way. Almost all of these books were published very cheaply, on poor quality paper, with poor photographic reproductions, as the books were not printed with the expectation that there would be any long-term interest. They were never intended as formal publications. Frequently the editorial committee found a Jewish-owned printing house which, in sympathy with their work, would give them a greatly reduced price. Sometimes they bartered a portion of the printing costs in return for adding the name and address of the printing house to the title page as a form of advertising. They were distributed through communal rather than formal bibliographic channels, and it is unlikely that the titles ever appeared in Books-In-Print, the veritable bible for commercial book acquisition.

Collection Development and Collection Management

The present collections of yizkor books came about in a variety of ways. Sometimes, a community member might approach their local library, or the library of a university or research institute with a personal gift of a book that they, or their forebears had owned. If it was a community with many Holocaust survivors, many of these individual gifts would amount to a collection and would be recognized as such. Other institutions and libraries recognized early on the value of collecting this material and were very proactive and conscientious in seeking donations and in contacting sources for acquisitions. New York Public Library took this latter route and now has the largest and most significant collection in the U.S. Many others came about as the result of U.S. Public Law 480 (1958), which allowed foreign debt to be paid to the U.S. through book donations. From 1964 to 1973, the Library of Congress administered the Israeli book receipt aspect of this program and distributed materials to many other libraries (3). Among the materials received from Israel were collections of yizkor books. This is how Portland State University received their collection, the largest collection in the Pacific Northwest region of the U.S.

Because of these various modes of collection development, it follows that collections are widely dispersed and of varying quality and quantity. The most important collections, those that have been developed conscientiously and pro-actively are those at Yad Vashem, the Martyrs' and Heroes' Remembrance Authority in Israel; the New York Public Library; the YIVO Institute for Jewish Research in NY; the Library of Congress in Washington DC; the Medem Bibliotheque in Paris, and the Jewish Public Library, in Montreal, Quebec. There was an important collection in Argentina, prior to the 1994 bombing of the Jewish community center in Buenos Aires, which destroyed the library. From Table 1, one can see that there are well over 60 institutions that collect these materials. Each collection, even a small one, has some individual books not found elsewhere. While most of the collections reside in the U.S., which is also where most Holocaust survivors fled to, it can be seen from the table that there is broad geographic distribution, including Israel, France, Great Britain, Canada, and Australia. The JewishGen website lists most of these collections (4). This resource however, must be used with care, as the holdings numbers actually represent several bodies of literature, and some of the most significant collections, for example the one at Medem, are not even listed on the site. A future project that would be helpful would be to investigate each of these collections and to provide scope notes linked to each holdings statement that would describe what the institution is defining as a yizkor book.

Table I: Libraries and Archives with Yizkor Book Holdings
(adapted from JewishGen website and other sources)

Location	Institution	Number of holdings
United States		
California		
	Hebrew Union College	450+
	Simon Wiesenthal Center	300
	UCLA (University of California)	Not given
	University of Judaism	Not given
	The Holocaust Center of Northern California	Not given
	Stanford University	Not given
Connecticut	Yale University	Not given
	National Yiddish Book Center	Not on list
Washington DC	National Museum of American Jewish Military History	Not given
	Library of Congress	Not given
	U.S. Holocaust Memorial Museum	460
Florida	Florida Atlantic University	450+
	University of Florida	450+
Illinois	Hebrew Theological College	~350
	Spertus Institute of Jewish Studies	Not given
Indiana	Allen County Public Library	200+
Maryland	Baltimore Hebrew University	Not given
	Jewish Museum of Maryland	Not given
Massachusetts	Clark University	40
	Boston Public Library	Not given
	Hebrew College	~125
	Harvard University	Not given
	Brandeis University	~500
Michigan	University of Michigan	Not given
	Holocaust Memorial Center	1000+
Missouri	St. Louis County Library	Not given
New York	Columbia University	Not given
	Cornell University	450
	Hebrew Union College	Not given
	Jewish Theological Seminary	Not given
	Leo Baeck Institute	Not given
	Museum of Jewish Heritage	~170
	New York Public Library	650+
	Yeshiva University	Not given
	YIVO Inst. for Jewish Research	Not given
Ohio	Hebrew Union College	Not given
	Ohio State University Libraries	700+

United States, cont'd		
Oregon	Congregation Neveh Shalom (Portland State University)	120+
Pennsylvania	Philadelphia Jewish Archives Center	Not given
	Gratz College	Not Given
	University of Pennsylvania	~175
	Reconstruction Rabbinical College	Not given
Tennessee	Jewish Federation of Nashville	Not given
Texas	University of Texas at Austin	~250
Australia		
Victoria	Jewish Holocaust Museum and Research Center	~90
Great Britain		
London	The British Library	110+
	Institute of Jewish Studies	Not given
	Jewish Genealogical Society of Great Britain	~65
	University College London	50+
	University of London	Not given
	The Weiner Library	~60
Oxford	Oxford Centre for Hebrew and Jewish Studies	350
Cambridge	Cambridge University Library	Not given
Canada		
Ontario	Jewish Public Library of Toronto	500+
	University of Toronto	500+
Quebec	Jewish Public Library (Montreal)	600+
France		
Paris	Centre de Documentation Juive Contemporaine	~200
	Medem Bibliotheque	Not on list
Israel		
Tel Aviv	Hitachdut Yotzei Polin	200+
	Moadon Ha'Bund-Brith Avoda	100+
	Tel Aviv University	~40
	Shar Zion Public Library	~500
	Ramb'm Library	Not given
Jerusalem	Hebrew University – Givat Ram Campus	Every Yizkor book published in Israel (#?)
	Yad Vashem Martyrs' and Heroes' Remembrance Authority	1040
Beer Sheva	Ben Gurion University of the Negev	300+
Haifa	University of Haifa	450+
Ramat-Gan	Bar-Ilan University	~420
Western Galilee	Beit Lohamei Haghetaot	~500

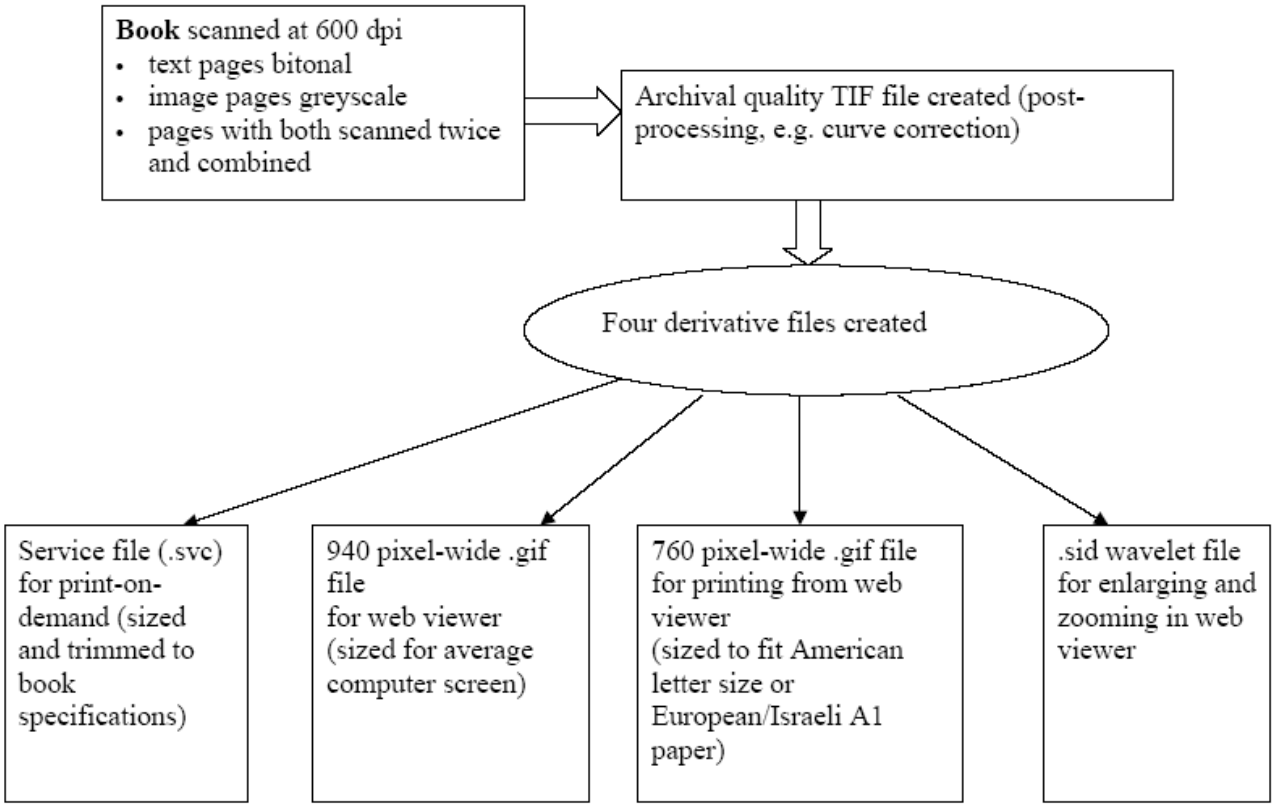
Total= 67 + 2, # not given = 29

The broad geographic distribution of these books is important. Most of these books are held in special collections or archives and are non-circulating. It is highly doubtful that they would be lent through interlibrary lending services. By being geographically dispersed, many people who have an interest in seeing these books physically are able to gain access. While digital initiatives, which we will talk about shortly, provide enormously broadened access to content and images, the physical items will always be of interest at a minimum as artifact.

The way that these collections are managed has a lot to do with how a library or archive defines yizkor books. Besides the books described already, there are other collections of literature that are sometimes also referred to as yizkor books, and which serve some of the same purposes, certainly for genealogists, historians, anthropologists, etc., and which certainly also fit within the rubric of grey literature. Among these are German-language books published by non-Jews about the destroyed Jewish communities of the German-speaking countries; individual survivor memoirs; general town histories by a single author where the dominant part of the book is not about individuals; books of memories collected through oral interviews by school groups in Israel; and Hasidic groups' histories of their rabbinical dynasties, which are usually associated with a particular town. While all of this material is collected by

some of the institutions mentioned, it may be managed as one large collection or as several distinct ones, depending on the collection development guidelines of each library. The New York Public Library collects all of the above but does not count them as part of the yizkor collection; they define yizkor strictly as the Jewish communal response. These variations in definition account for the wide ranging numbers found on the JewishGen website. The existence of multiple books for many of the towns, and multiple languages for many of the books add to the complexity of managing such collections.

Ironically, while the books were published as a means for perpetuating memory, they were published at a time when permanent paper was not used much, and as a result, the books themselves are in danger of 'vanishing' due to their poor production quality. Because of the poor production quality, the brittleness of the paper, and the very few copies in existence, these books are generally kept in special collections or archive type environments for maximal protection.



New Audiences, Digitization and Open Access

It was mentioned earlier that the intended audience for these materials was very small, and that their purpose for publication was personal and community use, both liturgical and for the general perpetuation of memory. These books, however, like much grey literature, have found new and broader audiences. The largest single group of users is by far, family genealogists. This group uses yizkor books to track individual family members, their spouses, children, occupations, and other personal details that can sometimes be gleaned from either the necrologies or the text. When such personal details are not available, family historians sometimes read the sections of the book detailing general information about the town, simply to derive understanding or pleasure from this description of their ancestral towns.

A second group of users are scholars whose uses of the books are as varied as the academic world. Linguists of Yiddish use the books to identify local terminology and turns of phrase. This is much less the case in the other languages in which yizkor books are written, all of which are more highly standardized and better documented than is Yiddish. Where a historian of a specific region might look at only a handful of books closely, those that come from towns in the region they study, a folklorist might compare descriptions of one particular custom in dozens or even hundreds of books.

A third group of users are novelists, playwrights, filmmakers, and visual artists for whom the detailed descriptions, photographs, and artwork in the books provide imaginative material, particularly where authentic, time- and place-specific models are needed for their creative endeavors.

These newfound audiences, coupled with the desire to have these materials handled as little as possible, combined to make yizkor books ideal candidates for digital access. One such endeavor arose from a community group, JewishGen, which has for many years been the digital home page for Jewish Genealogy, the single most important resource in the field and an incredible volunteer undertaking. The electronic access to records and family information made possible through this non-profit group with a total of one paid employee is remarkable. One of JewishGen's projects is an ambitious translation project, in which donations are used to pay translators. Given the cost of translation, and the difficulty raising

funds for an entire book, chapters as they are completed are posted immediately on their web site. To date 38 Yizkor Books have been completely translated and about a hundred others have had a chapter or more translated. Although this project is aimed primarily at family historians, there is no reason why they can't be put to other uses. Helen Winkler, a Canadian folk dance instructor, has combed JewishGen's online translations to find every description of dances and dancing customs in them. Winkler does not read the languages of the yizkor books and scans the JewishGen translations as they become available. On the resulting bibliography, which she has made freely available on the web, she links to JewishGen and promotes the translation project (5).

Worthwhile as JewishGen's translation project is, the original materials remain in danger of crumbling, and required a major digitization project to facilitate their use. Criteria set out by New York Public Library for it's digitization projects included (1) that the material be fully catalogued; (2) that it be high-demand material; (3) that the originals be materials that should not be handled frequently, but still in good enough condition to not be damaged by careful handling during the digitization process; (4) that the primary user groups be ones which are either accustomed to using digital resources or are motivated to learn.

The New York Public Library received an offer from a partner organization, the National Yiddish Book Center (6), which would raise the funds needed for the scanning of the books, in exchange for which they would have the right to sell reprints-on-demand. NYPL retained the rights to web-mount the images, but had to raise their own funds to do so. This is one of the few collections where having availability online would not detract from the marketability, in fact, it was felt that marketability would be enhanced. Most people have no interest in owning a yizkor book unless their family is mentioned in it. The online version will let people find out if their family is mentioned, and if so, a link is provided to the seller, the National Yiddish Book Center, so their sales are actually increased. On the other hand, scholars will want to be able to browse large numbers of yizkor books looking for the kinds of mentions they are interested in: folklore, linguistic patterns, shtetl institutions, etc. These users will want to print out only those pages that match their research interests and will never buy the whole book for a particular town. These users never were potential buyers of the reprints, so no sales are considered as 'lost' by such an initiative.

In taking on this project, it was felt that the collection was small and discrete and would allow the staff to 'get their feet wet' doing text-heavy digitization. It was quickly learned that this was actually a large digitization project, with 600 books averaging over 400 pages each, but the commitment had been made.

The digitizing project is near completion, with over 600 books currently available through the NYPL website (7). Once in the site, there is an alphabetical list of the towns for which books exist. Because the language of the texts is often not English, there can be multiple spellings of many of the towns to account for differing transliterations. For example, someone was looking for a town which they spelled as 'Choroyev' but gave some possible other spellings as 'Choloyev', 'Cholojow', and Choloyew'. One of the spellings worked, and a book for the town, from which the person assumed there were no survivors, was found. Once a book of interest is located, one enters the viewing software and is able to navigate to other books without returning to the alphabetical list. If a book is not yet loaded online, only the bibliographic record is displayed. Once one is inside of a book entry, there are links to the bibliographic record in the NYPL catalog; to a free download for an enlargement plug-in; and to the National Yiddish Book Center for purchasing – all using external pop-up windows and without taking the user out of the viewer. The website is highly contextual, with background information about the yizkor tradition and links to other major organizations involved in the project. The catalog records contain the current town spelling and country in the subject fields, so that users can confirm that they are viewing records for the town that they are interested in, as many towns had similar names. There are also links to as many common spellings as was hoped would be useful to users who may not know the current conventional spelling of their ancestral town.

The NYPL digital project used whole image scanning techniques rather than text capture, because optical character recognition software is simply not yet at an advanced enough level to properly decode Hebrew and Yiddish, or even Roman-alphabet languages that use diacritics such as Polish and Romanian. Additional issues with this collection are the varieties of fonts and production values, for example - one book is reproduced from a manuscript written on a Hebrew dot-matrix typewriter, and will be very difficult for a non-human eye to understand. The multiple languages per volume issue also makes it difficult to automate any part of the optical character scanning process. Thus, even when the software is able to handle diacritics and difficult fonts, the cost of making word-searching available for these books will still be considerably more than for single-language texts. Consistent with current best-practice in book digitization, every page, including blank ones are imaged. There are at least two images for each page, one .gif image at 960 pixels for viewing and one .gif image at 760 pixels for printing. Additionally, there is a third file made, a MrSID file (multi-resolution seamless image database; a.k.a .sid) for pages that have visual interest. A .sid file is a proprietary wavelet file (produced by LizardTech (8)) which allows for instantaneous resizing and zooming on complex images. With a .sid file the viewer can actually enlarge to see individual pixels in the image. While the user may download the viewing software for free, the producer of the images must buy the software for creating and displaying these files. The images are

shot at 600 dpi, high enough resolution so that when OCR software is ready, the books will not have to be re-scanned. The text-only pages are bi-tonal and the image pages are shot using grey-screen. Pages with a mix of the two are shot twice and then combined (flattened). For each page imaged, there are multiple files produced – a .tif file, which is an archival quality master, that has been processed for curve correction, de-skewing, de-speckling, etc; an .svc file, the service file which is sized to fit a standard sized page and trimmed according to the practices of the partner organization, the NYBC, to produce the reprints; and the aforementioned files for viewing and for printing. As a result, the .tif image is often times of superior quality to the original, and the 760 pixel .gif image fits onto a single sheet of printing paper.

The fact that these books exist in multiple languages, some reading from left to right and others from right to left, presents some interesting problems not encountered in many other digitizing projects. Because of the lack of adequate OCR software, keyword searching is not, at this point, an option. Also, JewishGen can now speed up their translation project, which may alleviate the need for OCR by creating English-language born-digital text that can easily be searched. By using the full-text images and viewer that NYPL has mounted, the translators can be anywhere in the world to gain access to the text that they are working with. One recent upgrade, that of a drop-down image-by-image navigation bar, has sped up navigation through longer books, though the image numbers cannot be keyed to specific page numbers, due to multiple paginations in each book. Other upgrading plans include the addition of books from the collection of Yad Vashem, in Israel, and the addition of unpublished yizkor books that have been loaded by individuals to the Internet. At present, NYPL provides links to these sites.

Now that the images of these books are on the web, access to them is 'open' though it is still protected. In many cases, copyright is still held by the families or the societies. Users are able to view, read, and print the content, but files may not be copied and mounted onto other websites, even personal ones. There are several reasons for this, but mainly the complexity of the collection and the extreme need for context, and respect for the intellectual integrity of the material limits this type of use. Anyone, however is free to link to the NYPL site.

This year marks the 60th anniversary of the liberation of the death camps of Europe. The number of survivors of this tragedy who are still alive today, and who were adults at the time, is dwindling. Those who were children at the time are also aging, and may have limited memories. The past 60 years have been an era when oral tradition and the accounting of first hand witnesses and survivors have kept the truth of this horror alive. As we lose these people as live historical resources to counter harmful revisionist histories that have arisen, additional resources are needed. The wide and open access to yizkor book collections, as described in this paper, is an important step in the furthering of Holocaust era research and in the perpetuation of memory.

References

1. Madsen, Catherine, In the Bond of Life: Yizkor Books of the Second World War, PaknTreger n46, p.20-25, 2004.

2. Hall, Kathatrina, The Politics of Memory: Memory and the Dynamic of Empathetic Identification within Historical Accounts of National Socialism and the Holocaust, J. Holocaust Education 8(3), p 41-70, 1999.

3. Hebraic Collections, An Illustrated Guide, Library of Congress [http://www.loc.gov/rr/amed/guide/hs-intro.html], accessed October 3, 2005.

4. JewishGen, [http://www.jewishgen.org/yizkor/yizlibs.html], updated 18 Mar 2005, accessed June -October, 2005

5. Winkler, Helen (compiler), Jewish Music and Dance as recorded in Yizkor books, [http://www.klezmershack.com/articles/winkler/yizkorlinks.html], revised March 4, 2005, accessed October 26th, 2005.

6. National Yiddish Book Center, [http://yiddishbookcenter.org/], accessed June-September 2005.

7. New York Public Library, Dorot Jewish Division, Yizkor Books [http://www.nypl.org/research/chss/jws/yizkorbooks_intro.cfmtp], accessed June - September, 2005.

8. Information on creation of MrSID files – see [http://www.lizardtech.com]

Recommended Reading

Guzik, Estelle, Genealogical Resources in the New York Metropolitan Area, New York, NY : Compiled and published by the Jewish Genealogical Society, 404 pp. 1989. ISBN: 0962186309.

Kugelmass, Jack; Boyarin, Jonathan (editors and translators) and Zachary Baker (Geographical index and bibliography), From a Ruined Garden: The Memorial Books of Polish Jewry, 2nd expanded edition, Bloomington: Indiana University Press, 351 pp. 1998. ISBN: 0253333946 or ISBN: 0253211875.

Mokotoff, Gary; Sack, Sallyann Amdur; and Alexander Sharon. Where Once We Walked: A Guide to the Jewish Communities Destroyed in the Holocaust, revised edition, Bergenfield, NJ: Avotaynu, 704 pp. 2002. ISBN: 1886223157

Author Information

Faith Jones heads the Literature and Languages unit of the Mid-Manhattan Library, The New York Public Library's central circulating library. From 2000 until early 2005 she was a librarian and Yiddish bibliographer in the Dorot Jewish Division of The New York Public Library, in which capacity she managed the Yizkor Book Digitization Project. She is currently working in a three-person collective to translate the complete poetic works of Yiddish poet Celia Dropkin.
Email: fjones@nypl.org

Gretta Siegel is the Science Librarian at Portland State University in Portland, Oregon and, for a while, was the ad-hoc Judaic Studies Librarian at PSU as a new program was under development. In this capacity, she became aware of a large local collection of Yizkor Books, which piqued her interest. Having found little published information about them, she embarked on this writing project. She has been working in various areas of grey literature for over 20 years. Email: siegelg@pdx.edu

Building a Digital Commons for Cyber Security Resources

Patricia Erwin
Dartmouth College; Institute for Information Infrastructure Protection, I3P, USA

Abstract
The Institute for Information Infrastructure Protection (I3P) is a consortium that includes academic institutions, federally-funded national laboratories, and non-profit organizations in the United States. The I3P brings experts together to identify and help mitigate threats aimed at the U.S. information infrastructure. Funded by the U.S. Department of Homeland Security (DHS), and the National Institute for Standards and Technology (NIST), the Consortium functions as a virtual national lab, with the ability to organize teams and workgroups to address research and policy-related aspects of the vulnerabilities inherent in the information infrastructure. A core value and long-term goal of the I3P is information sharing. In 2003, this led to the development of the I3P Knowledge Base project. With funding from NIST, the I3P Informatics Services Team was charged with creating a digital commons of cyber security information, tools, and resources for researchers.

Much of the early (2003-2005) information the I3P Knowledge Base provided was temporal, such as the cyber security events calendar and the funding opportunities alert service, or information aimed at the development of the Consortium. As the team began work on the digital library, a host of new challenges were presented. Cyber security, or information infrastructure, is a relatively new area of research, with much of the information about the topic falling into the category of 'grey literature.' While research has been published along the more traditional publication channels, there is valuable research information contained in technical bulletins, presentations, and workshop notes scattered throughout our members' offices and informally organized archives. The information is in a variety of formats, and presents a host of preservation, ownership, and access issues. There also is no fully developed or widely used taxonomy for understanding and categorizing cyber security information. The development of the taxonomy will be a second thrust of the project.

Perhaps the most interesting challenge of this project is how best to balance the need to make available research information in cyber security against certain security and financial risks posed by organizing and making available the information. This presents for the research librarian not only technical and administrative challenges, but also ethical questions.

The goal of my paper is to present a case study, outlining the challenges, the major stakeholders and their roles in creating both challenges and solutions, and the role librarians have played in this project. I believe that the topic of the paper, and the conference focus on open access to grey literature, are an excellent vehicle for this discussion. In examining this case study a number of larger social and political issues will be touched upon. While each of these topics are worthy of more in-depth discussion, I will focus this paper on the building of a cyber security digital commons.

Introduction

Cyber security and information infrastructure protection have emerged as key international issues. The Institute for Information Infrastructure Protection (I3P) was established to promote research to address the potential for cyber threats. The Informatics Services Team has been charged with developing a core set of tools and services aimed at providing access to cyber security information. Central to our mission is creating a digital library, composed of many information types. Much of the information that will be contained in the digital library falls under the definition of grey literature. This paper will outline the proposed development of the Digital Library, the challenges we face in building it, and the very prominent role grey literature contributes to the lasting value of the I3P Digital Library.

The goal of my paper is to present a case study, outlining the challenges, the major stakeholders and their roles in creating both challenges and solutions, and the role librarians have played in this project. I believe that the topic of the paper, and the conference focus on open access to grey literature, are an excellent vehicle for this discussion. In examining this case study a number of larger social and political issues will be touched upon.

Definitions
It is useful in thinking about the building of a cyber security digital commons to understand the vocabulary of the discipline. Additionally, one might ask 'what is a digital commons', or even a 'digital library'. This section provides an explanation of how specific terms used throughout this paper.

Information Infrastructure Protection
Infrastructure prior to the 1980s was thought of in terms of physical public works systems, such as our water or gas supplies. As more of these systems were computerized, it was recognized that the computer systems that now manage our water and gas utilities were vulnerable to exploitation. The term 'information infrastructure protection' refers to the identification, exposure and protection mechanisms put in place to guard against attacks against eight critical U.S. infrastructures[i] whose processes are controlled by computerized systems.

Cyber Security
Cyber security is a somewhat more problematic term. The National Telecommunications and Information Administration (NTIA) defines it as the protection of information against unauthorized disclosure, transfer, modification, or destruction, whether accidental or intentional. Although it is assumed that the NTIA is not identifying all information as needing protection, there is a subtle message that information 1) needs to be protected, and 2) to do something with information you need authorization.

Digital Commons
When we at the I3P speak of the Digital Commons, we are referring a set of tools and services that are available for public use. The term aptly describes a commitment the team feels to providing cyber security information to the broad public. Certainly, the concept of the 'commons' is not a new one. It was communally owned land from which all citizens could draw benefit. In that same vein the Cyber Security Digital Commons is intended as a virtual place the public is encouraged to use. Thinking about our project in this manner, has been a driving force behind the planning for and development of features of interest to not only researchers, but also the broader public.

Digital Library
There presently exists no comprehensive source for research information in the broad area of information infrastructure protection or cyber security. Individual organizations have in many cases done an excellent job of capturing their published information in these research areas. Academic or research libraries have also acquired and provided access to the published research for members of their academic communities. The I3P Digital Library will provide users with information about the broader body of knowledge in cyber security regardless of its format, location, or ownership.
The I3P Digital Library seeks to create a unique service that 1) captures unpublished and published information in the broad area of information infrastructure protection and cyber security; 2) provides a comprehensive index to information; and 3) offers research tools and services to researchers and the general public.

Cyber Security Taxonomy
The purpose of the cyber security taxonomy is to create a system for naming and organizing digital objects into subject-specific groups. There presently exists no universally adopted cyber security subject-specific taxonomy. The taxonomy produced by the I3P will be a set of subject terms that can be used with our meta-records to describe a digital resource.

Overview of the I3P Consortium

The Institute for Information Infrastructure Protection (I3P) is a consortium of leading national cyber security institutions, including academic research centers, government laboratories and non-profit organizations. It was founded in September 2001 to help meet a well-documented need for improved research and development (R&D) to protect the Nation's information infrastructure against catastrophic failures. The institute's main role is to coordinate a national cyber security R&D program and help build bridges between academia, industry and government. The I3P continues to work toward identifying and addressing critical research problems in information infrastructure protection and opening information channels between researchers, policymakers, and infrastructure operators.

The I3P chooses to avoid a static operating plan. Instead, its guiding principle is to study the gap areas in information infrastructure research – the vulnerabilities that are not being addressed by other research entities. This approach also produces data and knowledge that informs opinion and helps to build a community of researchers that can lead efforts to recognize threats and develop remedies. In 2003 the I3P released the *Cyber Security Research and Development Agenda* which identified eight underserved

research areas. As part of its community building process, the I3P also reaches out to government and industry to foster collaboration and information sharing.

The I3P Consortium has assumed a major role in helping to untangle such cyber security issues as infrastructure interdependencies, or systems of systems, and it is beginning an investigation of cyber security related policy, risk, and economic issues. The I3P is managed by Dartmouth College, with administrative offices in Hanover, New Hampshire.

The Digital Commons of Cyber Security

The Digital Commons of Cyber Security seek to be the electronic conduit through which users learn, collaborate, and create new knowledge in the broad area of information infrastructure protection. Our unique approach is the emphasis on scope, making it the first place the cyber security community [and others] turn to for what is happening in the world of information infrastructure protection.

In addressing our mission we have developed a spectrum of information services including an international cyber security events calendar, a funding opportunities alert service, Security in the News – our daily news aggregation services, a directory of international cyber security organizations, and the Digital Library.

There presently exists no comprehensive source for research information in the broad area of information infrastructure protection or cyber security. Individual organizations have in many cases done an excellent job of capturing their published information in these research areas. Academic or research libraries have also captured and provided access to the published research for members of their academic communities. We see the I3P Digital Library as providing the user with information about the broader body of knowledge in cyber security regardless of its format, location, or ownership.

Cyber Security Digital Library

Following, is a brief description of each what will be contained in the first iteration of the Digital Library:

- **Update Alert:** A page will be devoted to listing newly described and/or received information resources. This page will be automatically updated weekly.
- **Search the Digital Library:** A user will be able to search meta-records about information resources of a wide variety of fields, including: Author, Title, Subject, Keywords, Owner, Information Type, and conference, database, or journal name (if the object is included in a larger work)
- **Browse the Digital Library:** A user will have the opportunity to browse lists of objects, organized by Author, Title, Subject, Owner, Database (in which object is included)
- **Searchable Glossary:** A user may search a cyber security term and be provided with one or more definitions from reputable sources.
- **Building the I3P Cyber Security Taxonomy:** This is an overview page of progress made in developing the cyber security taxonomy. Consortium members will have an opportunity to review and comment on progress made.
- **Suggest a New Resource:** Users will have the opportunity to submit materials, citations, and other resources for inclusion in the Digital Library.
- **Personalized Research Page:** A researcher, from a Consortium Member institution, may set up an individualized citation collection page. We also offer access to bibliographic management software.
- **Rights Management:** Most of the information objects described by meta-records are not owned by Dartmouth College or the I3P. It is important for users to realize that the I3P Digital Library provides a "front door" into research materials, but ownership of items rests elsewhere. The I3P will provide a statement of ownership on each meta-record. The user will need to contact the owner for access to the information. This page also states the I3P policy regarding the access to and use of I3P information.
- **Request Research Assistance:** As a service to our members, the Informatics Team will provide literature searches on research topics of Consortium interest.

Content Development and Management

Content development and management have been addressed through a strong collection development policy. It clearly articulates what information resources are collected and why. A digital library's collection can be somewhat different from a more traditional library collection. Often, individual resources are not owned by the digital library, but are accessed from an off-site collection. The digital library may also serve as a pointing mechanism to resources that require purchase of access by the user. The I3P Digital Library incorporates three types of access, links directly to resources held in the I3P repository, links to digital resources held elsewhere, and information about resources for which access must be purchased.

In identifying content for the Digital Library staff use a set of criteria that includes an examination of the origins of the proposed content, quality of the resource, the type of resource, and how its' potential for long-term preservation.

Emphasis on Grey Literature

Much of our work in developing the Cyber Security Digital Library will be identifying, cataloging, and making accessible grey literature. Early on in the project we realized that this would be our unique contribution to body of knowledge in cyber security. We have identified a list of twelve resource types we anticipate will fall under the rubric of grey literature. We are casting a wide net in our attempt to capture as much quality information as possible. In pursuit of grey literature we will be searching for such diverse resource types as trip reports, research notes, blogs, and industry data.

Taxonomy development and Resource Identification

The development of a cyber security taxonomy will serve two purposes. We are presently using a 'hunter-gatherer' approach to finding individual resources in cyber security. While very effective, it has illustrated the point that there presently exists no universally adopted cyber security subject-specific taxonomy. We might need to search for one resource under a list of similar names or topics.
The purpose of the cyber security taxonomy is to create a system for naming and organizing digital resources into subject-specific groups. The taxonomy produced by the I3P will be a set of subject terms that can be used with our meta-records to describe a resource. The taxonomy developed will include input from researchers and domain experts, contain preferred and cross-referenced terms, and be expandable.

A further enhancement project that may be implemented is the concept of a user-defined taxonomy. Briefly, the user-defined taxonomy would allow searching on terms that the user found most useful. We know, for instance, that a researcher looks for information using terms that may be unfamiliar to the general public. While both groups may be seeking the same information, the route to retrieving that information might be very different. This idea of user-defined subject searching in cyber security is one the team is interested in exploring further.

Challenges to Building the Cyber Security Digital Library

While there are many challenges to building a digital library, the inclusion of grey literature in that library has presented by far the most interesting and in some cases perplexing set of considerations. Some basic tenets of librarianship, such as the good in providing access to information, have been tested both by the subject matter we are working with, and the underlying assumption that not all cyber security research information should be made publicly available. This is also a relatively new body of knowledge. The foundational resources or early writings were not necessarily ever vetted through a publication process, or even captured as formal documents. Our logistical challenge will be to assume the role of cyber-truffle hounds, ferreting out those pockets of cyber security information.

The overarching challenge is embedded in the composition of the I3P Consortium. Academic research centers, government laboratories and non-profit organizations view access to information in very different ways. In some cases this is cultural, but far greater difference is in how U.S. Federal government regulations address the access to information issue. The challenges faced by the I3P in building a collection of grey literature fall into three broad categories: social; legal; and logistical.

Social Challenges
As noted previously, the progenitor of cyber security information is the nexus between computer science and public policy.

Many of those working in cyber security operate in a culture of informality and sharing. This is not to imply a less than serious academic approach to the discipline, but rather a certain outlook that is reflected in naming conventions- think "honeypot' or 'phishing', and the growth of open-source software. This approach is relevant to the capturing of grey literature in that many non-academics are as likely to publish their ideas on their personal website or blog, as oppose to publishing papers through the more traditional publishing model. In attempting to capture and make accessible this information staff have had to be far more creative and dogged in their hunt for information.

Along with the culture of informality is an assumption among computer scientists that machines can manage their electronic information, as oppose to a librarian's approach to information management. While it is true that many of the older, established approaches librarians have used to manage information are simply 'past their prime', the tasks of evaluating information, developing standard taxonomies, and providing a standard systematic process for handling the lifecycle of information are still best done by humans with some degree of special training. This thinking can be a huge obstacle to overcome for any group attempting to manage computer science related resources.

The final social challenge involves the incentives an organization might have for not keeping certain types of information. Librarians are well aware of the link between preservation, access and censorship. If information is not preserved or cataloged then it is essentially non-existent to the user. As discussed later in this paper there may be strong incentives for not keeping information, and therefore making sure that the information never becomes either intentionally or unintentionally available.

Legal Challenges

Much of the grey literature we will be cataloging has been collected through informal means, that is, it has been stored in someone's office, file cabinet, or hard-drive. Several of the legal challenges we face are direct results of not having access to information about the provenance of a document. The chain of authenticity may have been lost, ownership of the information may be obscured, and those agreements that once dictated how the information would be used may be long lost or separated from the information they were intended to protect.

It is very hard to predict that a document, memo, or even note will prove to be a key piece of research information twenty years from now. Intellectual property is regularly produced without a date, author, or ownership stamp. A major challenge as we build the digital library will be to verify that that information we receive is actually owned by the provider, and therefore they can legitimately grant us use of the information.

Additionally, much business information is protected through the use of nondisclosure agreements (NDA). Many NDAs protect any information that is not publicly available. These agreements outline terms of use, such as who may have access to the information, how the information must be labeled, and how long the NDA is in effect. Unfortunately, the agreement may over time be separated from the information. Consider the implications of publishing a data-rich report on intrusion detection in a business or financial organization. While the data might be old, the simple process of making the information publicly available could be disastrous to the business. At the same time this type of data is notoriously hard to come by and extremely valuable to researchers.

Export control laws in the United States provide a matrix of guidelines controlling information about technologies that may be provided to others outside the United States. Certainly, these laws are relevant when thinking about the technology transfer of research finding. Again, in developing the Cyber Security Digital Library my team may not always know if a grey literature resource was subject to export control, and therefore having imposed restrictions on access. In handling grey literature this is of particular concern.

The final legal challenge was alluded to earlier in this paper. There may be very good reasons why not holding on to non-publicly released information, no matter the research value of the information, may be in an organization's best interest. The U.S. Freedom of Information Act (FOIA) allows citizens to request disclosure of records held in government organizations. While we generally think of FOIA as relating to federal information, a FOIA request may be filed with any government organization, including a state university, national research lab, or government institute. FOIA has a number of exceptions aimed at protecting the national security, trade or commercial secrets, and personal privacy, but the fact that FOIA exists can make those who hold information nervous about maintaining it long-term.

Logistical Challenges

If the social and legal challenges to this project were not daunting enough, the logistical challenges present another wall between users and cyber security grey literature. Given a body of knowledge that is not that old, with much of the early information produced as grey literature, the task of identifying those

pockets of information, and gaining access to them for cataloging and preservation purposes has been formidable. Our approach is to form a 'roving band of catalogers' who will go where the information is stored. We will start with I3P Consortium member institutions and branch out from there.

Conclusion

In looking at this project, it is obvious that the challenges are many and varied. This is balanced against what we on the project see as a strong public service component to the project. As society becomes more dependent on computing systems for managing our communications and critical infrastructures, the public needs trusted services that provide quality information in a manner easily understood.

Researchers, in looking at the body of knowledge known as cyber security, must have access to not only resources published through standard methods, but also those pockets of information and historical data that might not be easily found. Serving as a conduit to these resources is a primary motivator for the team.

Finally, there is that link between preservation and censorship that we can not ignore. The team sees a component of their mission as preserving, and therefore providing for potential accessibility, information that may be of lasting value to the understanding the political, social, and technical aspects of information security.

Endnotes

[i] *Critical Infrastructures and Key Assets*. CRS Report to Congress, October 1, 2004. http://www.fas.org/sgp/crs/RL32631.pdf.

William J. Rapaport. *Philosophy of Computer Science*. Department of Computer Science and Engineering, June 2005. http://www.cse.buffalo.edu/tech-reports/2005-16.pdf

Abby Smith. *Authenticity in Perspective*. Council on Library and Information Resources, 1992. http://www.clir.org/pubs/reports/pub92/smith.html

Author Information

Patricia Erwin is Senior Assistant Director for Informatics Services at the Institute for Information Infrastructure Protection (Dartmouth College, Hanover, New Hampshire). She has held information management related positions at the Dartmouth College Library, the University of Virginia Library System, and in private industry. She regularly teaches graduate and undergraduate courses in knowledge management, organizational learning, and critical thinking. Ms Erwin holds a Masters Degree in Library and Information Science from Indiana University, and a Masters Degree in Internet Strategies Management from Marlboro College. Her primary interests are in how organizations leverage what they know to push a disciple or area of knowledge forward, the role management of information plays in the research process, and how organizational structure enhances or impedes the flow of information.
Email: perwin@thei3p.org

Uniform requirements for Grey literature? Proposal for the adoption of a "Nancy style"

Paola De Castro and Sandra Salinetti
Istituto Superiore di Sanità, ISS, Italy

Background

In the last international conferences on Grey Literature (GL), the Internet "sea change" suffered by GL was deeply analysed under its different nuances, coming to the conclusion that the characteristics initially associated to this production have been completely upset. The main advantages of the Net regard information retrieval, document usability and availability against an increased level of responsibility for its authors and issuing organizations.[1]

This 7th International Conference on GL is devoted to open access, the movement exploring the most effective and affordable strategies for serving the interests of research, researchers, institutions and societies that support research, as stated in the famous Budapest meeting in 2001, the starting point for the development of the Budapest Open Access Initiative (BOAI). The purpose of that meeting was to accelerate progress in the international effort "to make research articles in all academic fields freely available on the Internet".[2] Since then, the BOAI has been developing successfully all over the world and involves different kinds of publications from pre-print to journal articles.

In this context, GL has now all rights to be included among the useful documents to be shared by the global scientific community and represents the greatest challenge in the information transfer process. Yet, in many cases, it will benefit from a proper re-styling to be exploited at best: when a document containing unique and precious information meets a formally correct production in absence of profit implications, success is guaranteed. This International Conference helps us to find the way.

GL: no more a supporting but a leading role

Journal articles and books have always been playing the role of main actors in the scientific publication arena. Information transmitted by these means is accurate, refereed, correctly promoted, traceable and available for all those who have the privilege of working in a research institute or can otherwise afford the cost of purchasing it. On the contrary, up to now, GL has been generally considered of minor importance, mainly because it was difficult to retrieve, poor in formal quality, produced in limited number of copies, and addressed only to small groups or "invisible colleges": it was a Cinderella that is now going to become the leading actor on the information stage.

Since the beginning of the 80s, after the York Seminar[3], with the production and diffusion of a specific standard for technical reports, the ISO 5966/82[4], the formal requirements for a correct presentation of GL started to be applied by issuing organizations, thus contributing to improve the quality of documents and their retrieval through bibliographic databases.

In the 90s thanks to the wide use and development of Desk Top Publishing (DTP) programs and later the diffusion of the Internet, GL production underwent a further impulse towards a better quality and availability. The interest for GL has been widely increasing at all levels: the "grey" attribute has no longer negative implications, but remains only to identify a genre that is now completely different from its first samples produced before the 80s (technical reports of very poor editorial quality but high information level).
In this framework, now that the ISO 5966 is out of date, we believe that it is important to give all GL producers the possibility to dress up their Cinderellas with a flick of a magic wand. How? Producing a "Nancy style" from seven mice and a pumpkin, meeting at GL7!

---

[1] De Castro P, Salinetti S. Quality of grey literature in the open access era: privilege and responsibility. *Publishing Research Quarterly* 2004:20(1): 4-12.
[2] http://www.soros.org/openaccess/
[3] Alberani V, De Castro P. Grey literature from the York Seminar (UK) of 1978 to the year 2000. *Inspesl* 2001;35(4):236-47.
[4] International Organization for Standardization. *Presentation of scientific and technical reports*. (ISO 5966/82). Geneva: ISO; 1982.

Thus our Cinderella will finally take off her old ash rags and put on a proper wedding dress to show her best look for a royal marriage! Cinderella will become a princess and her poor origin will be hardly traceable. On the contrary, the freshness of her new look associated with her precious old virtues (unique and original information which cannot be found elsewhere) will represent the best chance to conquer the heart of the entire realm of science! Let's avoid breaking this spell and let everybody stop to admire her elegance, her beauty and inner richness.

"Nancy style", the magic of Cinderella: six horses, a coachman and a coach to be introduced at Court

The production of a valuable and formally correct document is strictly associated with the existence and application of reference standards that should be commonly followed by all possible GL authors and producers.
In the biomedical field, everybody knows what the *Uniform requirements for manuscripts submitted to biomedical journals*, better known as "Vancouver style", are.[5] A short story of this "style" will help understanding why we are proposing a "Nancy style" for GL.
Vancouver was the place were a small group of editors of medical journals met in 1978 to reach an agreement for a common format of manuscripts submitted to their journals. Since then, these guidelines, which were initially based only on the format of references to be included in the publication, have been developing in scope and also the number of journals using them is widely increasing.
Today, the border line between grey and open literature becomes less and less clear above all in terms of online availability, therefore it seems particularly important to call the attention of all GL authors and producers on the necessity to draw up a reference tool for writing and editing their products that now are mainly distributed through the Internet. This is particularly relevant in the absence of a proper standard for the presentation of scientific and technical reports (the best sample of GL) since the valuable but not up-dated ISO 5966, produced in 1982, was withdrawn in the year 2000, and not yet replaced by a new standard.
Following the red thread of both the relevant editorial issues contained in the last edition of "Vancouver style" (October 2005, http://www.icmje.org/) and the basic principles of the ISO 5966, a proposal is presented and discussed to reach an initial consensus on what might lead to the adoption of *Uniform requirements for the production of institutional reports: how to write and distribute grey literature.* This is particularly useful now that open access is widely shared also by the GL community.
Seven mice, magically transformed into six horses and a wonder-coachman, led Cinderella to the castle. Our mice to reach the "castle" of consensus are represented by the following seven items:

1. *Ethical considerations*
   They regard responsibilities of the authors of reports and their issuing organizations mainly in terms of scientific quality and financial implications. They also take into account the appropriateness and relevance of peer review and possible conflicts of interest, privacy and confidentiality.

2. *Publishing and editorial issues*
   They include considerations associated with intellectual property (copyright) and the new responsibilities arising from electronic publishing.

3. *Report structure*
   It is the most powerful mouse allowing reaching the desired goal, i.e. the clear understanding of the report content. A well designed structure includes Front matter, Body of report and End matter – each part containing different kinds of information which may be compulsory (i.e., bibliographic information in the Front matter), recommended (i.e., structure of the Core of report), or optional (i.e., non textual material or Appendixes) (Figure 1).

[5] International Committee of Medical Journal Editors. *Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication*. ICMJE: 2005. Available from http://www.icmje.org/. Last visited: 31/10/2005.

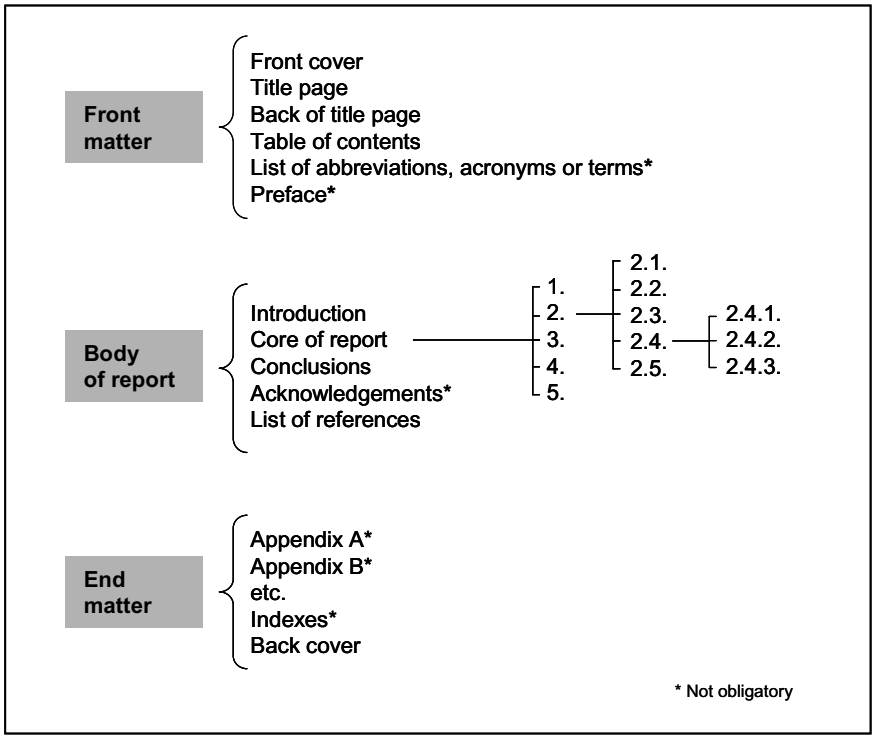


Figure 1. Recommended structure of a report

4. *References*
   They are the sources of information used in the report and shall be correctly placed and cited in the reference list to permit their retrieval.

5. *Non textual material*
   It is represented by tables and figures which play a significant part in the understanding of concepts explained in the text adding value to the information given. It is not compulsory, but, when included, it shall be properly placed and arranged.

6. *Revision editing*
   It is a careful reading and final consideration of the text before its release; it has the objective to ensure that the technical content of a report is clear, complete, accurate and understandable to the intended audience. There are different levels of revision mainly depending on urgency in diffusion, availability of editorial staff and budget.

7. *Instructions to authors*
   They represent our coachman leading the horses to the right direction. All issuing organizations should provide guidelines for a correct production of reports including rules and suggestions for all the items described above.

And now, still using the Cinderella metaphor, what does the pumpkin stand for?
The pumpkin is the content of our document, that is its most important attribute, always prevailing on all the others! A formally correct report – with a smart layout and well balanced structure – is absolutely useless without any good substance! Needless to say, the first responsibility of any author or issuing organization regards the content of the documents produced, which may only be improved and made clearer or better understandable thanks to a formally correct presentation. Yet, a good pumpkin cannot reach the castle to join the party of the noble (prestigious journal articles and monographs) if a flick of the "Nancy wand" does not magically change it into a coach, driven by well-trained horses and lead by a coachman. All guests allowed at Court must follow a proper etiquette!
The draft "Nancy style" will be discussed in the Round Table on "Quality Assessment of Grey Literature" of GL7. It will be integrated on the basis of the different suggestions of the experts in the field, and each comment will be most welcome to reach a consensus for the best practice to be followed.

The draft is divided into 5 parts:

1. Statement of purpose;
2. Ethical considerations;
3. Publishing and editorial issues;
4. Report preparation;
5. General information on the *Uniform requirements*.

The Report preparation, including a section on Report structure (see Figure 1), is the most practical part of "Nancy Style" and considers both the layout and formal requirements for each element of the report. A careful evaluation of all the items will help authors to understand how to prepare and organize a document and why they should respect these guidelines with the objective to facilitate communication.
The *Uniform requirements* are envisaged as a useful tool for all institutions and authors producing GL in order to provide added value to their documents. They may also be used as a guide for training courses to be organized inside institutions producing GL.
The *Uniform requirements* are not for profit and, after approval, they may be reproduced for educational purposes without regard for copyright. Their free distribution and translation will be encouraged provided that the official source will always be correctly cited.
The *Uniform requirements* will be periodically updated and a list of GL producers adopting them will be available from the official authority which will be established in Nancy, if consensus is reached.

A happy end: Cinderella becomes a Princess

GL producers have been eyewitnesses of the "sea change" introduced by the DTP and the Internet and have tried to adjust old rules with new requirements, facing the difficulty of authors who are often unaware of the challenges staying behind an autonomous production of documents.
The approval of "Nancy style" will provide a useful tool to guarantee the best quality of GL as it represents the consensus reached by the most distinguished experts in the field taking in due consideration both existing recommendations and standards, as well as years of experience in the production and diffusion of technical reports.
Finally, GL might be hardly distinguishable from open literature, since production techniques and distribution strategies are not so different for the two genres and open literature, in turn, is now disguising itself in grey when entering institutional repositories and open archives.
If correctly spread and applied, "Nancy style" will forever change the poor and grey Cinderella into a Princess who will be, as in fact is, mostly appreciated by the entire community. The term "grey" will no longer be associated to poor, second class literature, but to a fresh and new production, still keeping the originality of its contents and remaining free from restrictions and limits imposed by the most strict market rules of open literature.

Thanks to "Nancy style", the spell of Cinderella will not end at midnight!

Author Information

Paola De Castro is head of the publishing unit of the Istituto Superiore di Sanità (Italian National Institute of Health) and member the editorial committees of the publications edited by the Institute (a quarterly official journal, a monthly newsletter, different series of technical reports and information leaflets). She has published many articles on the information transfer process both at national and international level; delivers courses on scientific writing for the Italian National Health Service operators; is member of the Italian Library Association (AIB), the European Association of Science Editors (EASE) and the EAHIL (European Association of Health Information and Libraries). Email: paola.decastro@iss.it

Sandra Salinetti
Publishing Activities deals with the production of documents produced by the Istituto Superiore di Sanità, assisting authors in the preparation and editing of books and papers and reviewing the texts to comply with national and international standards and make them available through the Internet. She participates in a training project on scientific writing to improve authors' personal skills in the production of scientific documents with special reference to the presentation of technical reports.

Grey Literature in Public Administration: An Example of a Specific Quality Assessment System

Markus Weber
Competence Centre for Evaluation, CCE
The Swiss Federal Office of Public Health, SFOPH, Switzerland

Abstract
In the current discussion about "evidence" of effectiveness in health promotion and prevention, there is a call for including knowledge about the management processes and context of intervention implementation. Evaluation traditionally takes these aspects into account in judging the value and effectiveness of interventions. But evaluation reports are not often integrated into the evidence building process as they do not match the quality criteria of "published" research. However, we argue that this does not necessarily mean that their scientific quality is inferior.
This paper looks at a specific system of quality assurance and assessment procedures for managing evaluation studies as a basis for the discussion on how to broaden the concept of "evidence" to include information gathered through evaluation studies.
The Competence Centre for Evaluation (CCE) of the Swiss Federal Office of Public Health (SFOPH) commissions external evaluation studies of public health interventions.
By introducing and using a quality assurance system the CCE wants to achieve two main objectives. Firstly, the evaluation studies need to be of sound scientific quality. Secondly, they need to be useful and practicable, i.e. they need to produce conclusions that can be understood by the target group of the study and recommendations that can be implemented. The two main tools for assessing the quality of a report are described as well as how they are embedded within a wider quality assurance system.
Our meta-evaluations (evaluation of the evaluation) take into account the Evaluation Standards of the Swiss Evaluation Society (standards of good practice for conducting evaluations, www.seval.ch). Four quality dimensions of an evaluation are mentioned: Propriety, Accuracy, Utility and Feasibility (each with 3 to 10 standards). They refer to the process as well as the product of an evaluation (the report).
Wider scope for the discussion: In addition to including "quality assured" evaluations, which other "grey" material could/should we include as "evidence" of effectiveness (e.g. policy papers, guidelines, good practices papers, expert opinion etc.)? Such grey literature provides a lot of information on implementation processes, management and context, which is important for understanding about why and how interventions are "effective". What kind of criteria could be developed to assess such knowledge? Or do they already exist? Could/Should this type of evidence be graded according to classical concepts of "rating evidence"?

1. Introduction

Current debate suggests that the notion of „evidence" (of effectiveness) in health promotion and prevention interventions needs to be broader than "results" only, and especially those based on the RCT 'gold standard' (randomized controlled trials)[1]. In essence, there is a call for also including knowledge about implementation and management processes, as well as the context of interventions. Evaluation traditionally takes such issues into account when judging the value and effectiveness of specific interventions. But evaluation reports are not taken into account in the traditional evidence building process: as "grey literature" they do not meet the traditional "evidence" inclusion criteria, even if their methodology is of high quality. In the first place we have to assure that evaluations do in fact meet scientific quality criteria. This paper presents the quality control measures used by the Competence Centre for Evaluation (CCE), Swiss Federal Office of Public Health (SFOPH) towards this end.
The CCE is responsible for commissioning and managing all the SFOPH's external evaluations of public health measures - mostly of health promotion and prevention programmes and projects[2]. It is an internal service that, on the one side, has to assure the scientific quality, ethical conduct and trustworthiness and, on the other, the usefulness of the evaluation studies. Studies are mandated to external, neutral private and university research institutes.
Whilst our experience is concerned with health promotion and prevention evaluations, I am sure that much of it holds true for research and grey literature in general; knowledge gathered must be useful, and practically oriented.
The paper is structured as follows: In section 2, I try to explain the principles and background of our quality assurance system as well as those of the Swiss Evaluation Society's (SEVAL) Evaluation Standards. In the discussion (section 3) I refer back to the "evidence" discussion.

2. The CCE's quality assurance system

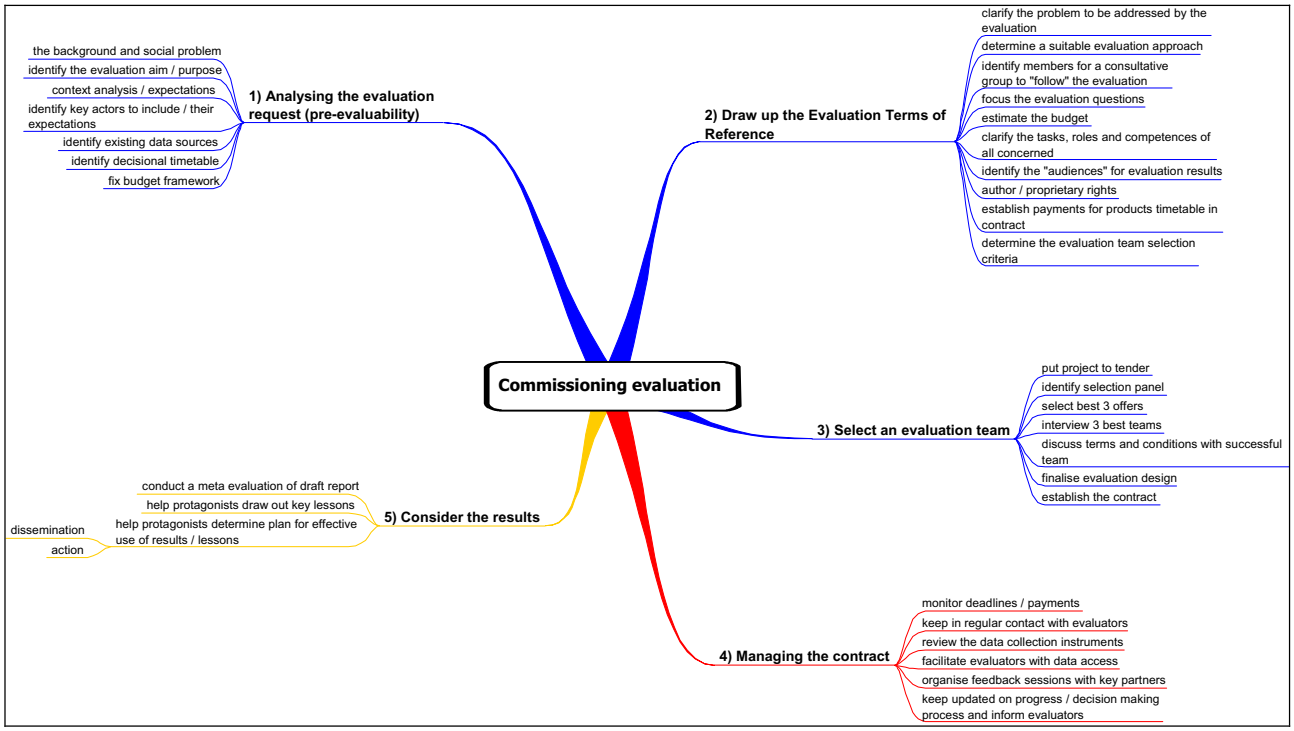
2.1. Objectives

By introducing and using a quality assurance system we want to achieve two main objectives. Firstly, we need to assure that the studies are conducted according to **sound scientific and ethical standards**. Secondly, the products of these studies need to be **useful and practicable**, i.e. the studies need to produce conclusions that can be understood by the target group of the study and recommendations that can be implemented.
Scientific quality is a necessary but not sufficient condition for successful evaluation studies. How can we get evaluation reports that first of all trigger implementation and learning processes in the area evaluated, but also beyond in other areas? We need reports that allow for easy understanding and utilisation by politicians, civil servants, practitioners in the field (medical doctors, social workers, prevention specialists, etc) and the research community. In other words, information about management and processes of the interventions has to be included in the analysis and reporting.

2.2. Description

Assessing the first draft of a final evaluation report is one of the last steps in a comprehensive package of quality assurance procedures used by the CCE to assure quality. It would not be very sensible to just come in at the end of a study and judge the quality of a report; rather it has to be steered from the beginning. We have standardised processes, guidelines, models and checklists that are used to guide the process from A to Z, i.e. from the first request for a study to actual commissioning, accompanying the study throughout, assessing the report (meta evaluation) and discussing and supporting a work plan for the utilisation/implementation of the study results. The quality processes and ethical conduct of the evaluation study as well as the quality of stakeholder involvement are the main pillars of quality assurance: the final report is the "output" of such. It has to include not only its findings, conclusions and recommendations, but equally a description of the processes involved to arrive at such results. *Figure 1* shows the 5 main steps of the evaluation process from a commissioner's point of view, as well as the many sub tasks that have to be considered within each step. Many of these are supported by CCE's checklists, models, etc.

Figure 1: 5 Steps of commissioning an evaluation



We clarify and repeat several times throughout the process exactly what we want and how we will assess it. All the relevant information about the object to be evaluated, the questions to be asked etc. are described in the Evaluation's Terms of Reference that is used to "Call for Tenders". And everything is again clarified in the contracts, kick-off meetings and other discussions between the external evaluators and the CCE.
This system has been successfully used for several years (SFOPH 1997, Läubli Loud 2004) and more recently was adjusted to take into account the SEVAL (quality) Evaluation Standards (Widmer et al. 2000[3]) which are described below.

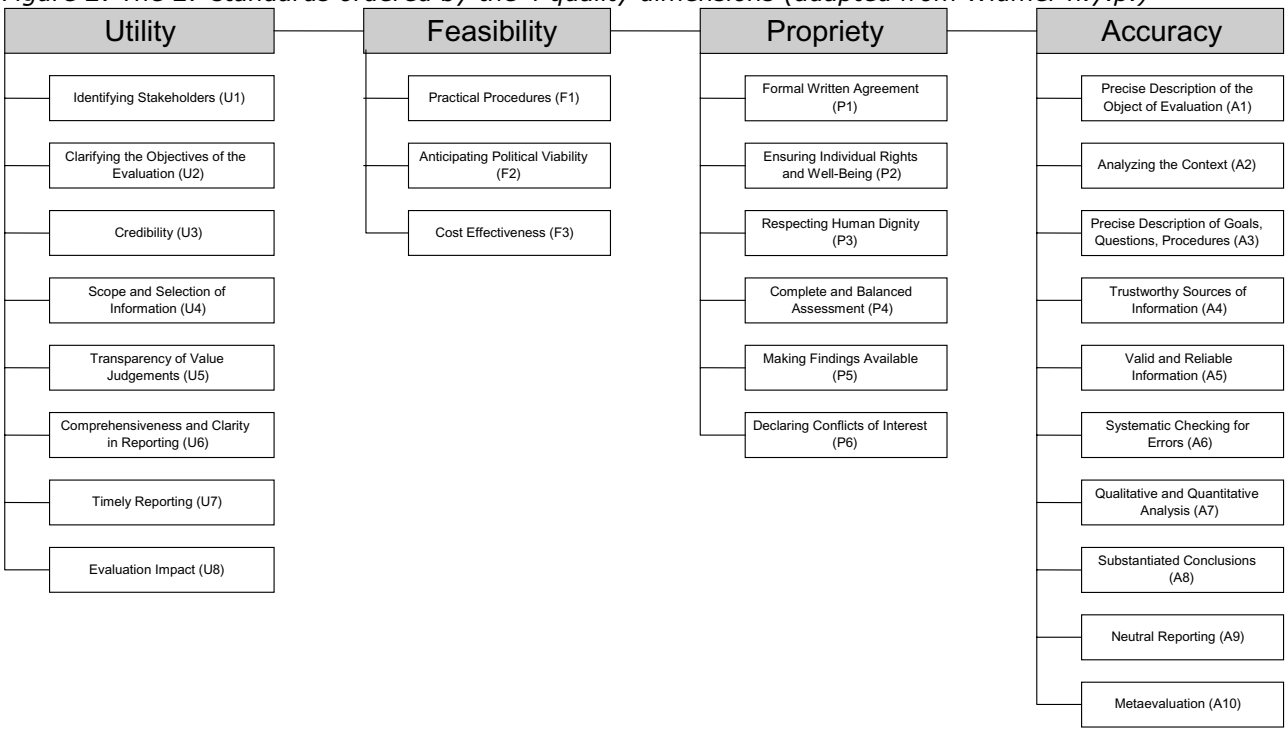
However, the CCE's quality assurance system is giving us only the formal structures (procedures and tools) which should support quality. In order to be really successful it has to be embedded in a so called "evaluation culture", both within the organisation and within the evaluator community. All people involved have to share a common understanding of evaluation, of what it is, of what it can do and of what it cannot do. The development and approval of the evaluation standards by the Swiss Evaluation Society (SEVAL), the integration of this topic into conferences, basic and continuing education, and the existence and utilisation of similar tools can be seen as indicators for the rise of such an "evaluation culture"[4].
Broken down to the level of an individual evaluation study, this means that the evaluators have to know what we expect (and we have to know what we can expect) and they have to be able to do it.
In the following paragraphs, I first introduce the SEVAL Standards, then explain how they are integrated into two specific tools developed and used by the CCE for (a) assessing a first draft of a final evaluation report and (b) for feeding these comments back to the authors, followed by some comments on how the product of these efforts can be characterised.

2.2.1. The SEVAL Standards

The SEVAL Standards are standards of good practices for conducting evaluations. They refer to the process as well as the product of an evaluation. The 27 standards are grouped into four quality dimensions: Utility, Feasibility, Propriety and Accuracy (cf. Figure *2*). The objective of each dimension is as follows (Widmer n.y.p.):

- The **8 Utility standards (U)** guarantee that an evaluation is oriented to the information needs of the intended users of the evaluation
- The **3 Feasibility standards (F)** ensure that an evaluation is conducted in a realistic, well-considered, diplomatic and cost-conscious manner
- The **6 Propriety standards (P)** ensure that an evaluation is carried out in a legal and ethical manner and that the welfare of the stakeholders is given due attention
- The **10 Accuracy standards (A)** ensure that an evaluation produces and disseminates valid and usable information

Figure 2: The 27 standards ordered by the 4 quality dimensions (adapted from Widmer n.y.p.)



The SEVAL Standards define the demands placed on an evaluation but do not specify the instruments to be used. Overall, they share the same concerns and objectives as those defined by the CCE: Sound scientific quality and ethical conduct (especially through the accuracy and propriety standards) and production of practical knowledge (utility and feasibility standards). The standards are categorised according to the quality dimensions. But they are not all equally relevant to every evaluation (e.g. subject to which methodology was applied) and certainly not to every phase of an evaluation (from initial planning to utilisation). Those who use the SEVAL Standards need to relate them to their specific evaluation needs and develop their own tools accordingly. General guidelines and checklists based on the SEVAL Standards are currently under development. The target audience is evaluation commissioners

within the Swiss federal administration (Widmer n.y.p., only in German for the moment). The CCE has long before developed tools which are based on the SEVAL Standards. In practice, they are not much different from each other, but our tools have been developed earlier on the basis of our experience and are precisely adapted to our needs. CCE team members were also involved in the development of both the SEVAL Evaluation Standards as well as the guidelines and checklists under development.

2.2.2. Two tools for assessing an evaluation (process and output/report)

The evaluation report is the document that sets out how the evaluation was conducted, how the data was analysed to arrive at which conclusions, etc. This means that the assessment of the report is at the same time the assessment of the whole evaluation process. CCE conducts a meta evaluation of both the written report and the evaluation process; having managed the contract throughout, it is regularly informed about how the evaluation was conducted. Under meta evaluation we understand the scientific and ethical quality control of an evaluation (cf. SFOPH 2005).
For assessing the quality of a first draft of an evaluation report we use a tool in the form of a checklist. The checklist assesses whether SEVAL Standards have been followed fully, partly or not at all. (or not applicable as the case may be). As mentioned before, not all standards are equally relevant at this stage and for every evaluation. So reviewers need to weigh them accordingly. Reviewers are also asked for general comments on strengths and weaknesses of the report, on the feasibility of the conclusions and on general operational, strategic and political lessons that can be learned from this study.
Every report is reviewed by at least two of our team members (we are 6 social scientists). For certain cases, external experts are asked to review the report as well.
The ratings and comments are then "translated" onto the second tool that we use at this stage. It is a form used to provide synthesised feedback to the authors/evaluators about how we rate the quality of their work and what we would like them to change. The form refers to the SEVAL Evaluation standards and the CCE's comments in relation to such. For example:

Positive: The underlying reasoning and points of view upon which an interpretation of evaluation results rests are described in such a manner that the bases for the value judgments are clear. (As given in the U5 standard: Transparency of Value Judgments)

Negative: The underlying reasoning and points of view upon which your interpretation of the evaluation results rests are not always clearly described. The basis of your value judgments is not transparent enough. E.g. page 23 …

Based on assessments per individual criteria, the CCE concludes with an appreciation about the overall quality (excellent, very good, good, etc.). (Detailed feedback about unclear formulation, grammatical or typing errors and other comments on specific paragraphs in the draft report etc. are also provided.). The meta evaluation is returned to the evaluators together with a timetable for completion (production of the revised report, the executive summary and/or other products e.g. oral presentations to stakeholders, etc.).

2.3. The product

These procedures and tools help us to achieve our two main objectives (sound scientific quality and production of useful and used knowledge). Scientific quality and professional ethical conduct is assessed through a strict review of the final product (evaluation report) - the last of the quality assurance procedures. However, the CCE also uses its own checklist to regularly monitor the evaluation process throughout. The production of useful, practical and used knowledge is supported by a comprehensive quality assurance system from start to end (to include the interpretation of the results and their transfer into action).
Generally speaking the evaluation products (executive summary, full technical report, vulgarised version, etc.) are put on the internet for public access. Their being "Grey Literature" therefore has several advantages: easy access, quick access and utilisation by main target group (area being evaluated) and others, adaptability, etc.
The orders for the printed version our guidelines (SFOPH 1997) and the access statistics to our internet site, show that we get many requests, from a lot of different fields and countries, from universities, other public administrations and from private companies and other people.

3. Discussion and Conclusions

The system described in this paper may seem very comprehensive and needing a lot of human resources. Being quite time consuming in its development it is making live easier for us now and recently allowed to compensate for human resource cuts of almost 50%. In the last restructuring of our institution we got additional tasks to be done with equal resources.
By using these quality assurance procedures we want to assess and steer evaluation studies towards producing quality information of practical use. The Swiss Federal Office of Public Health is mentioned in several international and national studies as a good and successful example of how to handle evaluation in public administration (Jacob and Varone 2002, Fornerod 2001, Widmer et al. 2001).

Moreover, our evaluations discuss "effectiveness" in relation to the processes and procedures used to implement the intervention as well as the key contextual influences. They therefore provide useful "evidence" about not only "what", but also why, how and for whom the intervention is considered "effective". Nevertheless, as "grey literature" the products of our commissioned evaluations would not be used in developing "evidence".
Traditionally, evidence on "effectiveness" is usually only discussed in terms of "what works" and the evidence is developed from peer-reviewed, published research only. Systematic reviews of published research that are conducted according to strict criteria, are the golden standard. But such "evidence" gives us little insight into which conditions best support interventions, particularly those that are multi faceted and implemented at multi levels within the community. As Saan (2005: 7) put it, literature reviews should include knowledge about "a) what works b) how it works and c) under what conditions".
Another disadvantage of the "traditional" approach used in systematic reviews lies in the fact that "studies are selected on the basis of their research quality, and not of the quality (*and practicability*), of the health promotion interventions" (Aro et al. 2005: 12). In other words, a broader understanding of evidence is needed, which in turn requires a broader scope of data. Most of such data is not available in published research, but in the "grey literature" such as policy papers, guidelines, 'models of good practices', expert opinions, etc. This type of data should be integrated into the systematic review process. (Molleman/Bouwens 2005: 8 and Aro et al. 2005: 12).
But is the quality of such data of high enough standard to be included as 'scientific evidence'? Quality criteria and protocols that go beyond the published research area are very much in their infancy at the moment (Aro et al. 2005: 13).
We have presented the CCE's means of trying to assure good scientific, ethical and useful quality evaluations - our quality assurance systems and standards. Yet the comprehensive information provided in such evaluations is not necessarily published – it therefore remains outside the boundaries of data used to constitute scientific evidence. *Should this be the case? And if not, how would we rate/grade the information provided? Could, or indeed should such information be "rated" according to classical concepts of "rating evidence" – according to which criteria? Are existing criteria appropriate?*

Deciding what else could/should be classified as "good" evidence in an attempt to build bridges between the "quality of research" and the "quality of interventions" could have consequences on the acceptance of health promotion as a whole. The evidence on the effectiveness of health promotion and prevention interventions is often debated and remains controversial. The inability of health promotion to always demonstrate "effectiveness" evidence by means of the established methodology could well weaken its status within the health system and indeed with policy-makers who often take the (wrongly) held position, that a lack of evidence of effectiveness equates to evidence of no effect (Speller et al. 2005).

4. Acknowledgements

I sincerely thank Marlène Läubli Loud for her contribution to this paper. She gave some initial input and valuable feedback to earlier versions. Furthermore, as head of the CCE she encouraged me to participate in the GL7 conference and to write a paper on the quality assessment of grey literature in the first place.

Notes

[1] A good overview on this debate can be found in Promotion & Education (2005) and some basic issues in Nutley et al. (2003)

[2] A description of the CCE (history, role, etc.) and published tools and reports can be found on www.health-evaluation.admin.ch and Läubli Loud (2004)

[3] Approved in 2001 by the Swiss Evaluation Society SEVAL, www.seval.ch

[4] A more detailed description on the background of an "evaluation culture" and on the current status quo of its development in Switzerland can be found in Läubli Loud (2004)

References

Aro, A.A., S. Van den Broucke, S. Räty (2005), Toward European consensus tools for reviewing the evidence and enhancing the quality of health promotion practice. - In: Promotion & Education Suppl. 1, 2005: 10-14. The challenge of getting evidence into practice: current debates and future strategies, IUHPE International Union for Health Promotion and Education, ISSN 1025-3823

Fornerod, S (2001), A quoi et à qui servent les évaluations ?, Mémoire de master, IDHEAP, Lausanne

Jacob, S., F. Varone (2002), L'institutionnalisation de l'évaluation des politiques publiques : expériences étrangères et scénarios pour la Belgique. Université catholique de Louvain, Louvain-la-Neuve

Läubli Loud, Marlène (2004), Setting Standards and Providing Guidelines. The Means Toward What End? - In: Evaluation, Vol 10 (2): 237-245, Sage, London, ISSN 1356-3890

Molleman, G.R.M., J.G.M Bouwens (2005), Building the evidence base: from tool development to agenda-setting and defining a joint programme for health promotion in Europe. – In: Promotion & Education Suppl. 1, 2005: 8-9. The challenge of getting evidence into practice: current debates and future strategies, IUHPE International Union for Health Promotion and Education, ISSN 1025-3823

Nutley, S., I. Walter, H.T.O. Davies (2003), From Knowing to Doing. A Framework for Understanding the Evidence-Into-Practice Agenda. – In: Evaluation, 9 (2): 125-148, Sage, London 2003, ISSN 1356-3890

Promotion & Education, Suppl 1, 2005. The challenge of getting evidence into practice: current debates and future strategies, IUHPE International Union for Health Promotion and Education, ISSN 1025-3823

Saan, H (2005). The road to evidence: the European path. – In: Promotion & Education Suppl. 1, 2005: 6-7. The challenge of getting evidence into practice: current debates and future strategies, IUHPE International Union for Health Promotion and Education, ISSN 1025-3823

SFOPH, Swiss Federal Office of Public Health (1997), Guidelines for Health Programme & Project Evaluation Planning. SFOPH, Berne.
www.health-evaluation.admin.ch (-> tools -> guidelines) [site visited 24.11.2005]

SFOPH, Swiss Federal Office of Public Health (2005), Glossary of Evaluation Terms, SFOPH, Berne
www.health-evaluation.admin.ch (-> tools -> glossary of terms) [site visited 24.11.2005]

Speller, V., E. Wimbush and A. Morgan (2005), Evidence-based health promotion practice: how to make it work - In: Promotion & Education Suppl. 1, 2005: 15-20. The challenge of getting evidence into practice: current debates and future strategies, IUHPE International Union for Health Promotion and Education, ISSN 1025-3823

Widmer, T., Ch. Landert, N. Bachmann (2000), Evaluation Standards of the Swiss Evaluation Society (SEVAL Standards). SEVAL, Berne.
www.seval.ch (-> standards) [site visited 24.11.2005]

Widmer T., E. Rüegg, P. Neuenschwander (2001), Stand und Aussichten der Evaluation beim Bund, University of Zurich, Zurich.

Widmer, T. (not yet published), Leidfaden für Wirksamkeitsüberprüfungen beim Bund. Guidelines, University of Zurich, Zurich

Author Information

Markus Weber studied Sociology, English Linguistics and Social and Preventive Medicine at Zurich and Lausanne University in Switzerland. For many years he worked for institutions active in harm reduction measures (street work, shelter homes, etc.). On finishing his studies he started work at the Swiss Federal Office of Public Health in the Competence Centre for Evaluation (CCE) where he is responsible for designing, mandating, assessing and valorising external evaluation studies on the Office's public measures (prevention, control, information) – www.health-evaluation.admin.ch.
Email: markus.weber@bag.admin.ch

Public funded research and Open Access: Perspectives and Policies

Cees de Blaaij
Library of Zeeland, ZEBI, Netherlands

Abstract:
There are several arguments for promoting the necessity of Open Access (OA). Public funded research can be considered as a common good. From that point of view the accessibility of scientific information is crucial as a political instrument in strengthening a democratic society and to improve the knowledge driven society by efficient and effective distribution of scientific information
An important question is what are governments doing to promote OA? Do they consider OA as a priority on their political agenda? There is also the issue of accessibility of public funded research to improve existing systems of knowledge sharing among scientists.
This article analyses and reviews these issues relative to the present situation in the USA and several European countries (UK, France Germany). Ongoing initiatives will be addressed that strengthen the OA movement in general. The last issue to be addressed is the impact of open access journals. An analysis will show that the "business-model" of OA is not a blind alley.

The issue

Advances in digital technology have radically reshaped the landscape of scientific publishing. Parties involved are commercial publishers, scientific communities and institutions, governments, libraries and the public in general. The phenomenon of Open Access with various publishing models is hotly debated as a viable alternative to the traditional "subscriber pays" model. The final outcome of the debate is not certain. It is within the setting of the issues of rising serials costs, industry concentration, and advances in information technology and the emergence of other publishing policies for the distribution of online scholarly research that the solution for an open and affordable access to scientific information has to be found for the good of society as a whole.

A key issue in the OA debate is the *public* accessibility of the results of *public* funded research. One of the most heard arguments to support this is the taxpayer's argument. For instance the Alliance for taxpayers[1] in the US support the principle that American taxpayers are entitled to open access on the Internet to the peer-reviewed scientific articles on research funded by the U.S. Government because they have paid for it. This argument has some validity but it cannot be the defining argument.[2] It even can be misleading. It seems to suggest that OA is required for taxpayers and not the world. It can locate the individual benefit or it can locate the benefit for the public interest. Arguments for OA should be multi focused. Also on a political level as democracies and networked societies evolve and citizens can be better informed or at least want to have the opportunity to access scientific information like medical literature.
An economic argument for OA is that free distribution of information is the essential tool for economic development and material well being in our age and in this way stimulates innovation. Jan Velterop has correctly observed "that freely accessible research optimizes the scientific process as well as its 'translation' into societal benefits."[3]
An important step to realize this was taken last year when the OECD issued a *Declaration on access to research data from public funding* on behalf of OECD Committee for Scientific and Technological Policy at Ministerial Level[4] as an incentive to develop international and national policies for free access to public funded research.

Old wine, new bottles?

From a historical point of view the principle of OA is not a new phenomenon in itself, on the contrary. Already Greek philosophers debated their views in public lectures. Another early example of OA was when Martin Luther nailed his handwritten 95 theses to the door of the Castle Church of Wittenberg in 1517 as a means to publish his theological views. In the Middle Ages and the Renaissance, scholars publicized their ideas and discoveries by writing letters to one another. Later in the seventeenth century as scholars

[1] The Alliance for Taxpayer Access, Source: http://www.taxpayeraccess.org
[2] For an extensive discussion about the taxpayers argument SPARC Open Access Newsletter, issue #65
[3] Jan Velterop, Public funding, public knowledge, publication, in: Serials. The journal for the Serials Community, vol. 16, nr.2, p. 169-174.
[4] Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communique, source http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html

shifted their standard medium of communication from writing to print. In the 1660s, the first scholarly journals collected the latest letters and printed them for the convenience of a "mass" academic audience. Actually we are talking about old wine in new bottles. One can see OA as a comeback of an old tradition in a new format. Different is the present setting in a knowledge based economy.
It is obvious that the distribution of scientific information including grey literature into the public domain needs to be efficiently and effectively organized to create surplus value.

Koyaanisqatsi

Unfortunately the market for distributing scientific information has become a disturbed one in the last decade caused by two factors. The first cause is the ongoing concentration of economic power of commercial publishers by merger transactions which destabilizes the competitive market for scientific publishing.[5] This industry concentration heavily influenced the price of serials negatively. Secondly the introduction of new intellectual property laws to accommodate legal security of the networked world. The outcome was that newly introduced copyright legislation in the EU and US was more in favour of protecting the interests of publishers and not enough consideration for the interests of the public domain particularly in regard to the fair use principle.[6]
One can characterize this whole situation as one in which the Hopi Indians speak in their language about *Koyaanisqatsi*[7], a concept meaning "life out of balance". On one side copyrights have been tightened on the other side access rights for scientists have not been redefined in the light of the serials crisis[8] or as way to legally improve scientific communication on the internet. Instead journal titles are cancelled by librarians and as a result "it has become increasingly clear that this crisis extends past the library, into our classrooms and laboratories. Not only are whole lines of scholarship in danger of disappearing, but professionals in industry, government, and education are finding that the information that does remain available is too expensive to access."[9]
One remedy to regain the balance is to improve the access and availability of publicly funded research in the public domain.

Significance of publicly funded research

The significance of public funded research can be quantified. It has as a large part in the overall production of scientific information compared in a worldwide perspective. But in general governments funding is lower than what business enterprises are willing to spend to research and development

Table 1 shows some indication on governmental R&D spending.[10] Some conclusions:

- The government sector accounts for 30,5% of funding in OECD countries. The business sector also performs most R&D. Business accounted for 62%. Its contribution to the overall R&D effort increased in the second half of the 1990s and has slightly decreased since. According to the latest available data, it accounted for about two-thirds of total R&D expenditure in the OECD area in 2003.
- The governments' role in R&D funding differs sharply across the three main OECD regions continues. It funds only 17% of R&D in Japan and 31.2% in the US, and 34. 2% in the EU.
- In the United Kingdom, Canada, the United States and Ireland, the share of government funding of R&D has increased moderately since 2000.
- Government remains the major source of R&D funding in almost a third of OECD countries
- Foreign funding of R&D continues to be an important source of financing in many OECD countries.

[5] A U.K. Office of Fair Trading investigation in 2002 concluded that "there is evidence to suggest that the market for STM (scientific, technical, and medical) journals may not be working well," but declined to recommend any market intervention at that time because "it remains to be seen whether market forces, perhaps enhanced by the use of new technology, will remedy the problems that may exist.", in: Office of Fair Trading. The market for scientific, technical and medical journals, London, September 9, 2002, p. 1.
[6] Cees de Blaaij, Two worlds: about bars and stars in scientific information publishing. An analysis of open source ideology as a means of self-controlled publishing, in: An international journal on Grey Literature, vol. 1, no. 1, 2005, p. 20
[7] Koyaanisqatsi (1983), a documentary movie. The film questions whether, in our haste for technological advancement, humankind have progressed out of balance with nature.
[8] Examples: strict conditions associated with access, which limit uses such as archiving the journal content locally, creating teaching materials, incorporating published information into databases, posting articles on institutional web sites.
[9] Mike Sosteric, Freedom from the Press: Alternative Academic Publication Strategies and the True Potentials of Information Technology, Tech. Source, Apr. 1999, available at http://ts.mivu.org
[10] OECD Science, Technology and Industry Scoreboard 2005 – Towards a knowledge-based economy, R&D expenditure by source of financing 2003 as a percentage of the national total, OECD, R&D database, May 2005

Table 1: R&D Expenditure. By source of financing. Percentage share in national total, 2003 or latest year available

	Business enterprises	Other (other national sources + abroad)	Government
Mexico (2001)	29,8	11,1	59,1
Poland	30,3	7,0	62,7
Hungary	30,7	11,3	58,0
Portugal (2001)	31,5	7,5	61,0
Greece (2001)	33,0	20,4	46,6
New Zealand (2001)	37,1	16,6	46,3
Turkey (2002)	41,3	8,2	50,6
Italy (1996)	43,0	6,2	50,8
United Kingdom	43,9	24,8	31,3
Austria	43,9	21,3	34,7
Slovak Republic	45,1	4,1	50,8
Iceland (2001)	46,2	19,8	34,0
Australia (2002)	46,4	9,2	44,4
Canada	47,5	18,0	34,5
Spain	48,4	11,6	40,1
Norway	49,2	8,9	41,9
Netherlands (2002)	50,0	12,9	37,1
Czech Republic	51,4	6,7	41,8
France (2002)	52,1	9,5	38,4
EU-25 (2002)	**54,5**	**10,6**	**34,8**
EU-15 (2002)	**55,1**	**10,8**	**34,2**
Denmark (2001)	61,4	10,4	28,2
Total OECD	**61,6**	**7,9**	**30,5**
United States	63,1	5,7	31,2
Ireland (2002)	63,4	8,6	28,0
Belgium (2001)	64,3	14,3	21,4
Sweden	65,0	11,6	23,5
Germany	66,1	2,7	31,1
Switzerland (2000)	69,1	7,7	23,2
Finland	70,0	4,3	25,7
Korea	74,0	2,1	23,9
Japan	74,5	7,8	17,7
Luxembourg (2000)	90,7	1,6	7,7

The statistical figures from OECD[11] in the government and higher education sectors are often based on estimates by national authorities and methods of evaluations are changed periodically. Certain national characteristics may have a strong influence on R&D performance by the government and higher

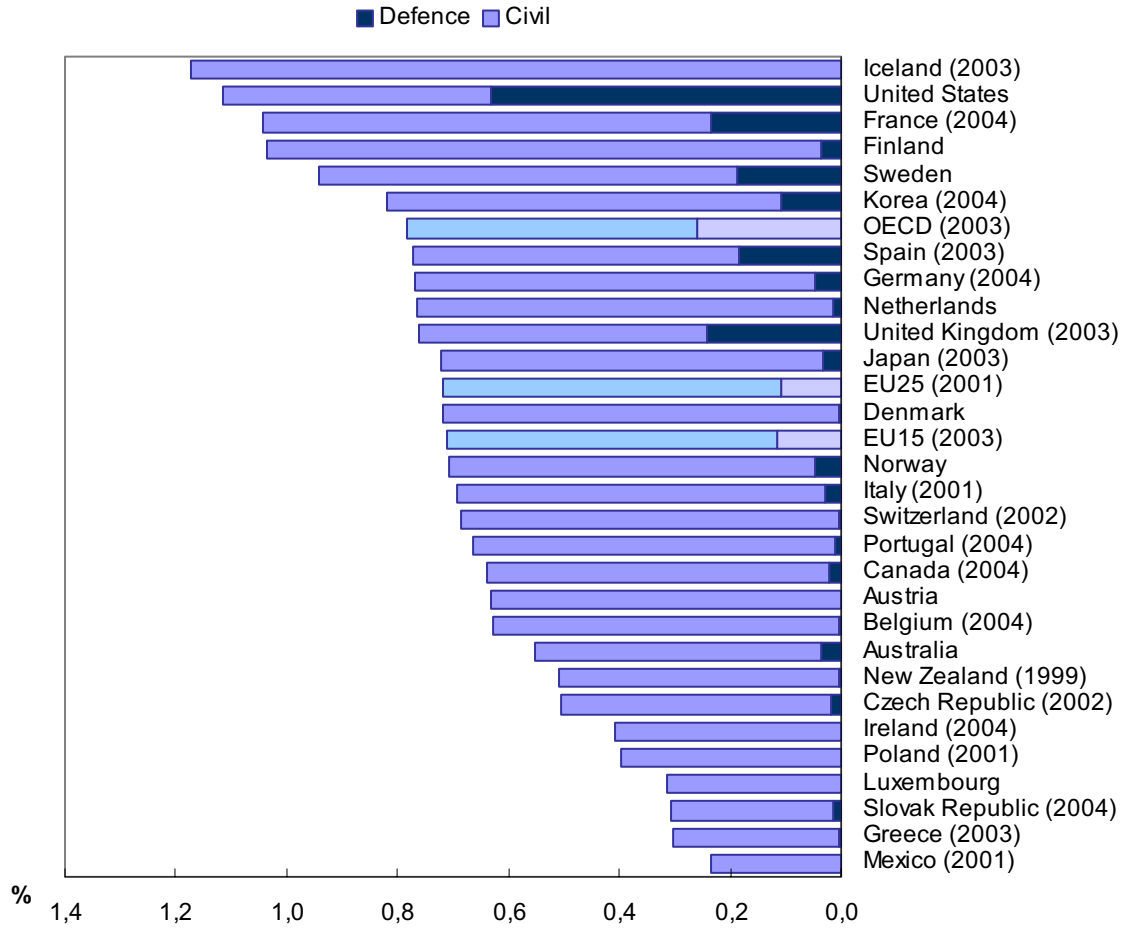
[11] OECD, Main Science and Technology Indicators database, May 2005 and OECD, Research and Development Statistics, May 2000.

education sectors. US figures for these sectors are underestimated. Public-sector R&D covers only federal government activities, not those of individual state and local governments; and since 1985 figures for researchers exclude military personnel in the government sector. In the higher education sector, R&D in the humanities is often not included.

R&D spending: civil vs. defence

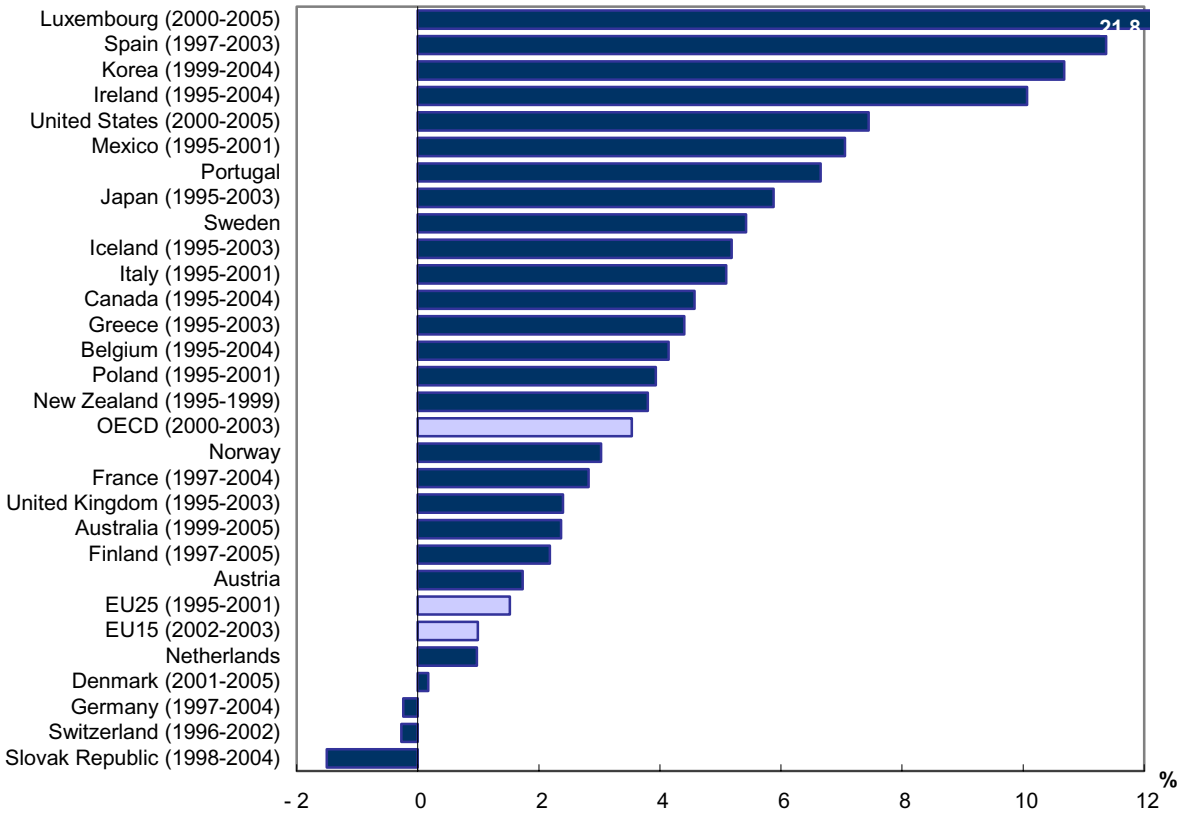
Another interesting figure is the ratio between spending on R &D for civil purposes and defence. Since the beginning of the early 1990s, the US government defence R&D budget has increased as a share of GDP and reached 0.63% in 2005. This is more than two and a half times the ratio for the United Kingdom and France, which have the second- and third-highest ratios in 2003 (about 0.24% of GDP).
In the United States, almost 57% of governmental R&D expenses are devoted to defence R&D in 2005. The United Kingdom is second with almost one-third. Spain, France and Sweden were the only other OECD countries for which the share of defence R&D exceeded one-fifth. Three-quarters of the growth in government expenditure in the United States between 2001 and 2005 can be attributed to defence R&D.

Table 2: Defence and civil R & D budgets as a percentage of GDP[12]



[12] See also note 4, Defence and civil R&D budgets – GBAORD as percentage of GDP, 2005 or latest available year, OECD, R&D database, May 205

Table 3: Change in government R & D budgets



Since 2000, government R&D budgets have grown on average by 3.5% (in real terms) in the OECD area. Growth has been modest in the EU25, averaging 1.5% a year since 1995, compared to 6% in Japan and more than 7% in the United States. The Slovak Republic, Switzerland and Germany have all experienced slightly negative growth since the mid-1990s.

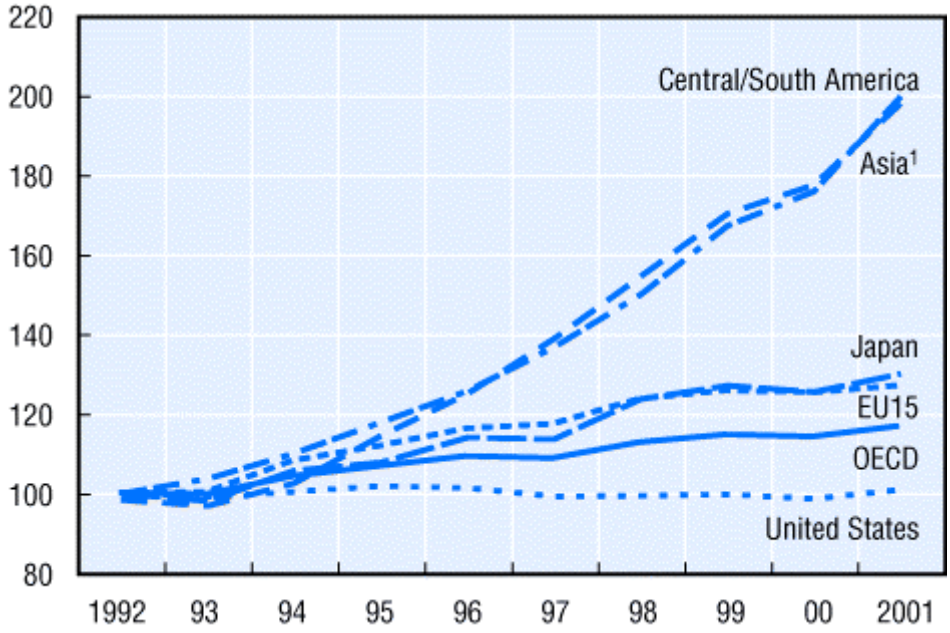
Output in scientific information

The output of scientific articles is measured by the OECD on basis of data from the Institute of Scientific Information.[13] This indicates scientific production. The output of articles is heavily concentrated in a few countries. In 2001, 82% of world scientific articles were released in the OECD area and about 65% in G7 countries.
The United States are the leader, with over 200 000 articles. The geographical concentration of output is very similar to that of R&D expenditures. The production of scientific articles is usually greater in countries where R&D intensity is higher. In Switzerland and Finland, output exceeded 1100 articles per million population in 2001. The intensity is highest in the Nordic and the English-speaking countries, and is also high in the European Union (557). On the other hand, output of scientific literature remains low in Korea and Japan compared to their R&D efforts. A statistical bias towards English-speaking countries may be part of the reason.

Over the past ten years, the intensity of article output increased in almost all OECD countries. The development of scientific activity and increasing co-operation among researchers stimulated the increase in Science &Technology publications. But while output kept growing rapidly in Western Europe and Japan, the number of articles stabilized in the United States and even started declining in the United Kingdom and Canada.

[13] Source Thomson Scientific: http://www.isinet.com

Table 4: Growth of scientific articles by area, 1992-2001



Life sciences dominate the OECD data. Physical sciences are the main field of publication in Eastern Europe, Korea and Portugal, as in the emerging Asian economies.

The major producers of scientific articles, Switzerland and the United States, are also the most cited. Both have a strong reputation worldwide in biomedical research and physics. In fields such as Earth and space sciences for Chile, mathematics for Slovenia or psychology for Argentina, emerging countries also achieve world recognition.

"Raison d'etre" for Open Access

OA is one solution for handling the output of the knowledge society but several commercial scientific publishers have repeatedly complained about OA as a real threat to scientific publishing in general.[14] Mainly they criticize OA as being destructive for the quality of scientific publishing because of a lack of peer review. The other side of the story is that they followed the practice of exploiting scientific information as a commodity for a maximized profit. As a result journal prices have been escalating.[15] Many research libraries weren't able to subscribe to journals anymore or had to cut in monograph purchasing in order to continue their journal subscriptions. The "eating disorder" from some of the commercial scientific publishers – some would call it greed - did consume the budgets of their *clientele.* Another unpleasant consequence was that monograph publishing became more difficult because publishers were unwillingly to produce special titles for a limited market and so the costs per monograph are also rising.[16]

The notion that commercial publishers can make big profits on the scientific results of public funded research triggered the awareness that scientific information has common good characteristics, is non-exclusive and is not competitive when it is consumed. It makes the economic nature of scientific information very different from other consumer products.
Scientific knowledge is not a scarce item, it comes in abundance when it is shared. This speeded the realization of OA journals and institutional repositories of scholarly research with free access by universities and research centers.[17]

[14] Responses to questions posed by the Science and Technology Committee by Elsevier, from: http://www.elsevier.com/authored_news/corporate/images/UK_STC_FINAL_SUBMISSION.pdf
[15] For a price overview: http://www-us.ebsco.com/home/printsubs/priceoverview.pdf
[16] As library budgets are squeezed by expensive journals, the market for books is reduced. Today many scholarly books sell only 200-400 copies compared with 1500 copies a decade ago, source: http://www.ala.org/ala/acrl/acrlissues/scholarlycomm/scholarlycommunicationtoolkit/faculty/facultyeconomics.htm
[17] Raym Crow, The case for Institutional Repositories: A SPARC (Scholarly Publishing and Academic Resources Coalition) Positioning Paper, from: http://www.arl.org/sparc/IR/ir.html

Growth of OA

From the perspective of growth clearly the OA movement is pushing the market. There is a steady growth of the number of journals in the Directory of Open Access Journals. In February 2005 it stood at 1,463; double that of a year ago. In November 2005 1908 journals were available.[18] Also the number of institutional repositories[19] with OA policy is steadily rising.

Table 5: Growth of Open Access Journals & Repositories

	Feb. 2005	Nov 2005
Directory of Open Access Number Journals	1,463	1,908
OAIster (Number Institutions)	405	523

The number of organizations that have signed the Berlin Declaration has gone up to 134.[20]

Governments and public policy towards scientific information

The present situation makes it clear that the serial crisis is a crisis of public policy. One reason for the continuation of this crisis has been the fact that central governments have been very hesitating or even reluctant to promote the interests of the public domain. But alas some ambivalence in the position of governments cannot be denied. As Larry Lessig has keenly noticed: 'the real problem lies with governments, which are too often disciplined by a market that is more interested in private rather than general welfare. The market that controls today's policy makers is keen to keep them from grasping obvious truths that would substantially add to the general good[21]."
An affirmative and striking example has been the dismissive reaction of the UK government[22] in 2004 to the report of the Science and Technology Committee concerning scientific publishing.[23]
The committee had endorsed principles of open access and offered particular strategies. The UK government declined the recommendations that it would promote and enact OA. The Committee's answer was that the "Government has clearly decided against the author-pays model ahead of the further investigation that it was urged to pursue. This approach prejudges the issue."[24]
One can see this as a victory for Reed Elsevier the world largest publisher. It is also a relief for a lot of scientific societies which are dependent on income derived from subscriptions by member and libraries. If the recommendations would have been accepted it would have put Britain into a leadership role in the OA movement.

Till up to now very few countries have carried out a national policy on OA although several OECD countries signed the Declaration on access to research data from public funding.
The conclusion must be that more governmental support for OA in a direct of indirect way is essential in finding an appropriate balance in access and intellectual property rights especially in regard to public funded research. The commercial publishing market will not solve these problems despite the fact that they are experimenting with OA-like tactics but in way not to endanger their investments.

So from the perspective of public interest and the taxpayer's point of view governments should feel warm-hearted to the promotion of Open Access.

[18] Directory of Open Access Journals, source: http://www.doaj.org/
[19] Country Update on Institutional Repositories, source: http://www.surf.nl/download/country-update2005.pdf and OAIster (collection of freely available, previously difficult-to-access, academically-oriented digital resources from University of Michigan Digital Library Production Service.
[20] Number of signatures, source http://www.zim.mpg.de/openaccess-berlin/signatories.html
[21] Cit. "Do you Floss" from Larry Lessig, in: London Review of Books, 18 Aug 2005. Professor of Law at Stanford Law School and founder of its Center for Internet and Society.
[22] UK Government's Response to the Committee's report, source: http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/1200/120006.htm#a1
[23] UK House of Commons Science and Technology Committee Report: Scientific Publications: Free for All, source: http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm
[24] See note 6.

Initiatives on national levels

So what are governments doing to promote open access to public funded research and do they consider OA as a priority on their political agenda?

Australia
The Australian Research Information Infrastructure Committee (ARIIC) prepared a statement outlining its support for open access in publishing.[25] They support institutional repositories

Denmark
A policy is encouraged by the Danish Electronic Library (DEFF) committee on electronic publishing.[26]

Finland
The Open Access Scientific Publishing Committee has published a report. The committee was appointed to put forward recommendations for the promotion of open access to scientific and scholarly publications in Finland. The aim of the recommendations is not to change the traditional standards used for evaluating the quality of scholarly publications, but to improve access to and the availability of the publications. Institutional repositories are promoted.[27]

France
The four major research institutions – who have signed the Berlin Declaration – CNRS, Inserm, INRA, INRIA declared their intention to move towards institutional repositories.[28] The central government has not yet ventilated her opinion.

Germany
The German Rectors' Conference is an advisory body to the federal government and the state governments. In their recommendation they make a direct connection between the serials crisis and the making of institutional repositories.[29] A lot of German research institutions have signed the Berlin Declaration also.

Italy
On November 4th-5th 2004 thirty-two Italian universities (more than 40% of the Italian universities) gathered in Messina, Sicily (Italy) to sign the "Messina Declaration", and committed to sign the Berlin Declaration.[30] A large number of Italian Universities have signed the Berlin Declaration. Institutional repositories are being implemented. A special Working Group is trying to get support for a national Open Access policy.

Norway
The Norwegian council for Higher Education has recommended its institution to establish Institutional Repositories[31] Norway has not yet implemented the Berlin Declaration but are planning to do so not just endorse it.

[25] ARIIC's Statement on Open Access, source: http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/australian_research_information_infrastructure_committee/documents/open_access_pdf.htm
[26] DEF's Action Plan 2001-2002, source: http://www.deflink.dk/eng/omdef/action_plans.asp
[27] Report of the Open Access Scientific Publishing Committee, source: http://www.minedu.fi/julkaisut/tiede/2005/tr08/kuvailu.html#DESCRIPTION
[28] Vers un accès libre aux résultats de la recherche…Le CNRS, l'Inserm, l'INRA et l'INRIA créent des archives institutionnelles pour les chercheurs, source: http://www2.cnrs.fr/presse/communique/640.htm
[29] Zur Neuausrichting des Informations- und Publikationssystems der deutschen Hochschulen, source http://www.hrk.de/de/download/dateien/Empfehlung_Bibliothek.pdf
[30] Gli atenei italiani per l'Open Access: verso l'accesso aperto alla letteratura di ricerca , source: http://www.aepic.it/
[31] Brief report on the Norwegian Council for Higher Education conference on Open Online Access to Research which took place a few days ago in Oslo, source: http://www.ub.uib.no/avdeling/fdok/fdok/Referansegruppe/harnad.htm

Sweden
There is no national policy, but the foundations already are available in Sweden, as the Lund University Libraries [32] has one of the largest and most active university eprint archives and programs for self-archiving university research output[33]. It also hosts the Directory of Open Access Journals.

The Netherlands
No national policy is available but in practical sense a number of academic institutions and Dutch universities cooperate to set up a national infrastructure for electronic publishing and established a repository called DARE (Digital Academic Repositories).[34]

United States
The Unites States do not have a national policy but there are many OA initiatives. An important step was taken by the National Institutes of Health for access to medical information. Other initiatives like the Public Library of Science and Biomedcentral are important for the access of journals.
In the US there are about 250 research universities. One conclusion from a survey from the Coalition for Networked Information (CNI) in the spring of 2005[35] is that it seems very clear that institutional repositories are becoming well-established as campus infrastructure components. Interesting is it to note that there is a growing interest among research funding agencies in data management, curation and archiving that is not necessarily closely coupled to the open access debates.
Another conclusion is that research libraries in the US are leading the way in policy formulation and operational deployment for institutional repositories.
The most noticeable difference in approach to OA between the US and Europe is the level of policy deployment: Less centralized and more oriented towards the needs of individual organizations.

Impact of open access

It is difficult to estimate the size and the nature of the impact of OA because of a lack of satisfying analysis. Another problem for assessing the impact of OA is the nature of OA.
OA publishing consists of different types of publishing not only journals but also individual articles on basis of author self-archiving or institutional archiving.

Some conclusions:
- The currently available journals are often more in the lower ranking journals
- Three of the four subject groups: Life Sciences, Medicine and Chemistry have one or more OA journals among the top 9% of the category
- High ranking OA journals are not evenly distributed. They are most frequent in the Physics, Engineering & Mathematics subjects.
- The number of OA journals that are covered in the Thomson ISI database continues to grow through the creation of new titles as well as traditional journals which choose an OA distribution model.

Open Access is well established in the subject areas of mathematics, physics and engineering because these scientific areas already had a long history of pre-print servers. These servers contained articles. It is likely that researchers with their experiences adopted OA journals more easily than researchers in the social sciences and humanities.

Other issues
It is curious but that the field of humanities has a very limited part in the Open Access movement. In this area the OA is moving slowest. Are there any reasons for this one may ask?

[32] http://www.lub.lu.se/headoffice/staff/larsbj.html
[33] http://lu-research.lub.lu.se/information.html http://eprints.lub.lu.se/
[34]DARE, source http://www.darenet.nl/en/page/news.list/show
[35] Institutional Repository Deployment in the United States as of Early 2005, source: Dlib Magazine, Sept. 2005, vol. 11, no. 9, source: http://www.dlib.org/dlib/september05/lynch/09lynch.html

Here are some clues:

- Journal prices are lower. This reduces the urge to use OA a business model. According to the 2002 Library Journal pricing survey, the average subscription prices for journals in STM fields were 10-20 times higher than the average prices in the humanities.[36]
- Humanities research receives overall less funding than STM research.
- On average, humanities journals reject articles more (70-90%) than STM journals (20-40%). So the cost of peer review per accepted article is higher in the humanities, lower in the STM fields.
- There is more public demand for OA to research on STM issues than humanities
- Preprint archives are very common in the natural sciences, rare in the humanities. The circulation and flow of scientific knowledge in STM is much faster due to the up-to-date demand from researchers.
- Demand for journal articles in the humanities drops off more slowly after publication than demand for articles in the STM fields. Humanities journals will be anxious to keep their income and most of them will not be inclined to put to their articles online according to OA. Of course there are some exemptions as the Directory of OA journals shows.
- Humanities are more book orientated than STM

This doesn't mean the OA future for humanities is not bright. It is probably also the law of the restraining lead. So why the rush…..

Author Information

Cees de Blaaij studied Social and Economic History at the University of Nijmegen and Library Science at the University of Amsterdam. He worked for Ernst & Young, management consultants, and the Institute for Information Law (University Amsterdam) as information professional. At the moment he is working as an academic librarian and coordinator digital services for the Public and Academic Library of Zealand, Netherlands. He took part in several GL conferences. He published several articles on issues concerning copyright in the digital environment and accessibility of grey literature on the Internet.
Email: cdbl@zebi.nl

[36] Tables for Periodicals Price Survey 2002, source: http://www.libraryjournal.com/index.asp?layout=article&articleid=CA209908#t2

Access and Document Supply: a comparative study of grey literature

Chérifa Boukacem
University of Lyon, France

Joachim Schöpfel
Institut de l'Information Scientifique et Technique, INIST
Centre National de Recherche Scientifique, CNRS France

Abstract
The report addresses the different aspects of the accessibility and dissemination of grey literature in the digital age where the de-materialization of documents has led to a new paradigm that has superseded the intrinsic characteristics of printed material.

Based on the added value of grey literature for academic institutions, the report attempts to provide an analysis of the ongoing transformations, especially concerning the way in which research and development in the area of grey literature have become part of the open access movement.

In this context, we will analyse some of the major public supply services for the dissemination of grey literature: their typology, their strategic approach, and the special conditions and characteristics of their service. What are their projects with regard to grey literature and the open access movement? What is the impact of these projects on document supply, acquisition policy and the information system?

For the study, we selected five public institutions: the British Library (UK), the CISTI (Canada), INIST (France), KISTI (Korea) and the TIB Hannover (Germany). We excluded networks and corporate profit-based suppliers.

Introduction

One characteristic of grey literature is that it is "often difficult to (…) acquire through normal bookselling channels" (British Library)[1] because it lies "outside of commercial circuits of publication and dissemination" (AFNOR)[2]. For those professionals and scientists who are desirous of gaining more information, the search for and acquisition of this kind of "underground literature" can be a time-consuming, sometimes expensive and even frustrating experience (see for example Hartley 2004 for conference proceedings).

Because of the importance of grey resources for scientific research and teaching, all major public document suppliers invest in collections and delivery services for theses, conference proceedings, reports and unpublished working papers. Some of them offer additional help for the identification and localising of these often poorly recorded documents.

These special collections and their bibliographic controls induce significant costs to libraries, and "grey supply" is often more expensive than the traditional delivery of items from "white" publications.

Today, libraries and document supply services are confronted with the open archive (OA) movement and have had to adapt to it. The following paper gives an overview of the grey holdings and delivery services of five major document suppliers – the *British Library* (BL), the *Canadian Institute of Scientific and Technical Information* (CISTI), the French *Institut de l'Information Scientifique et Technique* (INIST), the *Korean Institute of Scientific and Technical Information* (KISTI) and the *German National Library of Science and Technology* of the University of Hannover (TIB) – and describes their reactions and projects in the OA environment. This overview is completed by an analysis of the position of the international lobby organisation of information services, the ICSTI.

Some remarks on OA and grey literature

The origins of the short history of open access are known: free peer-reviewed journals like Harnad's *Psycholoquy* or Guédon's *Surfaces* and article repositories such as Ginsparg's *ArXiv* or Harnad's *CogPrints*

[1] <http://www.bl.uk/>
[2] <http://www.afnor.fr/portail.asp>

have started to challenge the traditional model of scientific communication since the early 90s. These and other initiatives were the scientists' and academics' reaction to the pricing policy implemented by the publishers of journals and reflected their intention to re-appropriate the channel of the communication of and to improve the dissemination of research results with the new information and communication technologies (see Crawford et al., 1996). The general impact on intellectual property and the "value chain" has been well documented by Roosendaal (2004); its effect on document delivery has already been conceptualised (see Rowse, 2003 or Brown, 2003).

We won't add another review to the rather abundant literature on the theory, history, structure and software of the open access movement. The reader will find comprehensive introductions to the European and more specifically to the French initiatives on the Southampton[3], URFIST Paris[4], CNRS[5] or Max-Planck[6] websites or through textbooks such as Grüttemeier & Mahon (2002) and Aubry & Janik (2005).

Yet, the conceptualisation of the relationship between OA and grey literature is just at the beginning stages (see for example Sondergaard et al., 2003, Banks, 2004 or Stock and Schöpfel, 2005). Do the "green" or "gold" roads (Guédon, 2004) transform white into grey documents because they leave the traditional commercial control and dissemination channels? Or, on the contrary, do open archives promote the shift from grey to white documents by increasing their availability and access? It is obvious that the boundary between white and grey is shifting, and that we need more empirical evidence on grey literature and open access systems in order to update our understanding of what is grey.

Compared to article repositories, open archives for grey literature are more recent and less numerous. One of the first, most important initiatives (if not the first) was the 1996 creation of the *Networked Digital Library of Theses and Dissertations* (NDLTD) by Virginia Tech. Other totally or partly grey realisations are the French-Canadian *Cyberthèses* developed by the Montreal and Lyon 2 universities (electronic theses and dissertations, ETDs), *Grisemine* from the University Lille 1 (see Claerebout 2004), *Archimer* from the French Research Institute for Exploitation of the Sea (IFREMER), the French national portal in mathematics *Mathdoc* (preprints, course material, proceedings and ETDs together with white resources and archives) and the open archives of the CNRS Centre of Direct Scientific Communication at Lyon (CCSD), such like ArchiveSIC and MemSIC (publications, dissertations and working papers in information and communication sciences) or TEL (ETDs, see Paillassard et al., 2005).

In spite of their great variety, all these archives and repositories have in common the increased value awarded to grey resources and the improvement of their availability and dissemination. Nevertheless, the majority of them are small-scale projects compared to the existing print holdings of grey literature. Another characteristic of these archives: up to now, reports, proceedings and other non-traditional types of grey literature (educational materials, working papers etc.) have been rather under-represented, compared to dissertations and preprints (cf. Banks 2006).

But what exactly is the impact of the OA movement on the former suppliers of theses, reports and proceedings? How are they reacting in response to the development of institutional repositories; how are they involved especially in the domain of grey literature? There is very little evidence on this topic (for the North American context and the role of the physics preprints archive, ArXiv.org, see Jackson, 2004), so we decided to undertake an empirical study based on information available on the Web or published elsewhere and on interviews with the leading managers of the cited institutions.

Questions and methodology

The underlying questions are about the existing holdings of grey resources, supply conditions and pricing, involvement in digital and in more specific OA projects, and on approaches and attitudes towards the OA movement. Five major institutions were selected for this study:

The *British Library* (BL)[7]
The *Canadian Institute of Scientific and Technical Information* (CISTI)[8]
The French *Institut de l'Information Scientifique et Technique* (INIST)[9]
The *Korean Institute of Scientific and Technical Information* (KISTI)[10]

[3] http://www.eprint.org
[4] http://www.ccr.jussieu.fr/urfist/ArchiveOuverte/OA.html#ou
[5] http://openaccess.inist.fr
[6] http://www.zim.mpg.de
[7] http://www.bl.uk.com
[8] http://icist.cnrc.gc.ca
[9] http://www.inist.fr
[10] http://www. kisti.re.kr

The *German National Library of Science and Technology* of the University of Hannover (TIB)[11]

These "traditional suppliers" have in common a public mission to collect, preserve, archive and disseminate scientific information through a non-profit ILL and a document delivery service that is based on a mixed economic model with their income supplied both by public funding and their customers' fees. In a certain way, they all share the "belief that collecting world knowledge and making it available for investigation generates new ideas and advances society" (The British Library, 2005).

ILL and document supply networks such as Subito, IMPALA or AutoDoc, were excluded because they have neither their own holdings nor a specific acquisition policy. For similar reasons, we excluded corporate, profit-based suppliers such as Infotrieve.

First we gathered information about collections and supply conditions – services, prices, activity reports etc. – from the web sites or from other sources. The information was organized in the following plan:

	Holdings	**Services**	**Requests**	**Prices**
BL				
CISTI				
INIST				
KISTI				
TIB				

Based on this information, we contacted the leading supply managers of the five institutions and proposed a survey on the following topics:

1. Figures on grey document supply and ILL in 2004.
2. This area of their activities compared to the overall supply and ILL (%).
3. The recent evolution compared to the previous system.
4. Their projects in the area of grey literature (also together with other partners).
5. Their projects regarding the open access movement (also together with other partners).
6. The impact on the collection of grey literature.
7. The impact on document supply (service offer, pricing).
8. The impact on the bibliographic control of grey literature (cataloguing, record data).
9. The impact on the information system.

The original verbal contact taken up with the supply managers was followed by a written version of the survey and a semi-directive interview that potentially included other areas as well. The synthesis and interpretation of these responses will be communicated to the various institutions for comments and validation.

We'll try to get additional information from the International Council for Scientific and Technical Information (ICSTI), a forum that "seeks to reduce or eliminate barriers to effective transfer of information by (…) enhancing access to and delivery of information to business, academia, government and the public".[12]

Our underlying assumption is that even if the impact of increased access to digital resources on document supply is relatively well documented, little is known about the specific effect on grey literature, especially in the sector of open access. Our hypotheses are:

- The portion of grey literature in overall supply activity is rather small.
- The "grey supply" follows the general trends of the overall activity.

[11] http://www.tib.uni-hannover.de/
[12] http://www.icsti.org/about.html

- Due to their public mission, all institutions are strongly interested and more and more involved in open access projects.
- One part of these OA projects is related to traditional grey literature.
- This increased involvement has or will have various effects on the way they function and their activities, especially on the development of collections and supply services.
- This involvement may also impact bibliographic control (the shift from conventional cataloguing to metadata harvesting) and the information system (the shift from a central library management and/or supply system to a more distributed and OAI compatible structure).

During the 7th International Conference on Grey Literature held at Nancy, we presented the background, methodology and some preliminary results. The definitive results will be published in 2006.

In the following, we communicate the first (and incomplete) findings on the holdings, services and OA projects of the five institutions, without any results from the still ongoing survey. It is a work in progress, mentioned to inform others about our research project without any anticipation of future findings and conclusions.

The British Library

The British Library is the world's largest supplier of scientific information (Brindley, 2005). Its grey holdings – mainly dissertations, reports and conference proceedings – comprise about 4.9 million items from a wide variety of sources (UK theses, NASA and NTIS reports, proceedings from most of the international scientific events, but also maps, manuscripts, audio and music documents, etc.)[13]. The British Library was the most important national input centre in the European EAGLE network, providing more than 50% of the overall SIGLE records.

Nevertheless, since 1998/1999 the British Library has to cope with a significant decrease in the number of requests. Even if the exact number of requests is no longer published in the annual reports, the decline can be estimated at 5-10% per year (from 4 million in 1998/1999 to 2.4 million in 2004/2005).

In spite of this, the decrease in revenue from document supply does not affect its acquisition policy "which (is) seen as part of the public good" (Brown, 2003). The fundamental role of information collecting and preserving has been validated and reinforced by the report "Redefining the Library: The British Library's strategy 2005-2008"[14] that places the end-user at the heart of its overall consideration. Their central credo is to "enrich the user's experience"; their main objective: the creation of a digital library. In this new environment, document delivery is no longer a "stand-alone business" or "separate entity" but "one part of an integrated range of information services" (Brindley, 2005).

The British Library contributes to OA projects mainly through the Research Support Libraries Programme (RSLP), which indicates improved relationships with the UK higher education community (Law, 2004, Ceeney, 2004); one pilot project was the MAGiC cooperation with Cranfield University on scientific and technical reports (Needham, 2002). Other projects at the edge of grey and OA are the selective archiving of websites, the archival sound recordings (ASR) project and the digitalisation of selected collections in the context of the UK Digital Preservation Coalition (Millar, 2001). Some "Co-operation and Partnership Programmes" (Smith, 2001) are focused on the preserving, digitalisation and disclosure of national heritage resources and non-traditional items (legal materials, company reports, official publications etc.).

CISTI

The CISTI has a long history going back to 1929 (Krym and VanBuskirk, 2001). As one of the largest collections in North America, its mission is "to support the research and innovation communities by managing and disseminating high-value scientific, technical and medical information products and related services".[15] The CISTI has become, over the years, the major provider of scientific information in Canada, making up about 36% of the Canadian market in the year 2000.

Apart from scientific periodicals (50,000+ titles), the CISTI collections include monographs, conference proceedings, technical reports and dissertations. For all types of documents the CISTI outlines a specific acquisition plan, except those dissertations that are collected only in response to document supply requests[16].

[13] http://www.bl.uk/services/document/dsc.html
[14] http://www.bl.uk/about/strategy.html
[15] CISTI Strategic Plan 2000-2005
[16] http://cisti-icist.nrc-cnrc.gc.ca/irm/policy_f.html

Since 2002, the CISTI has reported a decrease in revenue from its traditional activities, which is linked to the development of "concurrent" alternative access to scientific information – access to online journals ("big deals"), the availability of free resources on the Web, and open (free) journals and archives. This decrease in income from traditional document supply, considered to be a long-lasting reality, is at the origin of the redefinition of its economic model, especially of the charges for resource acquisition, and of its strategic development.

The strategic plan 2005 – 2010[17] proposes a partnership strategy where CISTI creates, hosts and maintains an open access institutional repository for the publication of Canadian scientists and academics.

Another open access partnership is with the press of the CNRC that, since 1929, has published journals, monographs, and other types of grey literature. The online version of the 15 peer-reviewed CNRC journals is available to the Canadian scientific community for free through the CISTI web site. For other countries, the same access is based on a subscription that contributes to the economic viability of the CNRC press.

These projects are part of the "depository services program" (DSP) designed "to ensure that Canadians have ready and equal access to federal government information", made possible by funding from the Treasury Board of Canada (VanBuskirk and Krym, 2003).

"This integrated, partnership-based approach to STM information will result in a Canadian science "infostructure" and a viable national STM publishing infrastructure. The competitive advantages offered by this investment will lead to better research and a more innovative economy"[18].

The overall position of the OAI seems rather prudent – based on a realistic knowledge of national scientific production and the information requirements of their customers, the CISTI acknowledges that the number of supply requests will continue to decrease but that this is mainly linked to the online access of electronic periodicals and not to institutional repositories or other open archives.

INIST

Founded in 1988, the Institute for Scientific and Technical Information is a service unit of the French National Centre for Scientific Research (CNRS). Its mission is to collect, analyse and disseminate the results and findings of worldwide research in STM, social sciences and humanities. The supply of copies of scientific and technical documents is part of its traditional activities, whether or not the documents are held at INIST (Schöpfel, 2003).

The legal background to this service is based on the French law on intellectual property as well as on an agreement with the French copyright agency (CFC).

The INIST grey collections are rather significant: 70,000 conference proceedings, 150,000 dissertations and theses and 70,000 scientific and technical reports. As the national centre of the European EAGLE network, INIST is a central part in the French landscape of the collection and dissemination of grey public research documents.

The CNRS and INIST have been particularly active in several ways in the European and French OA movement (Romary, 2005, Schöpfel, 2005, Grüttemeier, to be published). Some of these activities are:

- The signature of the Berlin Declaration on Open Access 2003.
- Institutional Membership of BiomedCentral (French host of the BMC archives).
- Contribution to a national common OA policy and the development of a central national repository for French public research, together with other public research organisations (INSERM, INRIA, INRA).
- The development of OA platforms for digital periodicals ("I-Revue") and reports ("LARA", see Stock et al., 2006).
- The organization of conferences and seminars on open access to scientific information and the creation of a Web site dedicated to OA issues with international news, reference texts and review articles.[19]

[17] http://cisti-icist.nrc-cnrc.gc.ca/about/stratplan05_f.pdf
[18] http://cisti-icist.nrc-cnrc.gc.ca/about/stratplan_summary_e.html
[19] http://www.inist.fr/openaccess

The general evolution of the INIST supply activity is well documented (see Schöpfel, 2003, Bador et al., to be published). Since 2001, the downward trend is about 10% per year. INIST offers print and electronic delivery (ARIEL); secured document supply with digital rights management is under preparation.

Document supply and resource acquisition are not directly linked. INIST's "historical" objective is to supply all requested documents, may they be part of its collections or not. In 2005, more than 20% of the items supplied were from other (French and foreign) libraries, and this percentage is steadily increasing. At the heart of the supply service is the identification and localisation of the requested items. As a result, the INIST supply service could become more and more independent of its own collections which would be "replaced" by remote resources (online resources, holdings from other libraries).

KISTI

The South Korean science and technology information centre, a non-profit affiliated organization of the Korean Research Council of Public Science and Technology under the control of the Prime Minister's Office (Choi 2003), provides a nationwide information service to support industrial and technological development in Korea by facilitating the access to and utilization of information resources.

The KISTI collection contains a wide variety of resources, journals, monographs, R&D reports, conference proceedings, patent specifications etc. The document delivery service supplied more than 300,000 items in 2001, from the KISTI holdings or from other libraries (11%), especially foreign document service bodies such as the British Library (Shin and Oh, 2002).

KISTI has developed a Scientific Information Integrated System (SIIS), a portal system that gives access to databases and online resources.[20] Grey literature is part of these digital information resources: access is provided to research and development reports (200,000 items), master and doctoral dissertations (700,000 items), Korean patents (2,100,000 items) and foreign patents (14,000,000 items), the science and technology factual database (1,400,000 items) and scientific materials and instruments (22,000 items) (Choi, 2003).

The OAI is discussed in Korea but up to now it seems unclear if it would become a national standard for the long-term preserving of and access to digital information (Choi 2003). Nevertheless, KISTI has been developing a multilingual electronic theses and dissertations (ETD) system since 1989 as a national digital library in South Korea and as part of the international effort to construct a networked digital library of theses and dissertations (Zhang, Lee and You, 2001). This database supports full-text searches in English, Korean, Chinese and Japanese, and provides access at individual page level. In September 2001, it contained already 23,368 ETDs; about 25% of the visits came from more than 20 foreign countries.

TIB Hannover

As one of the most important libraries in Germany since its foundation in 1959, the German National Library of Science and Technology of the University of Hannover (TIB) defines its task as the following: "to comprehensively acquire and archive literature from around the world pertaining to engineering and the natural sciences… Holdings (…) are available (…) through our document delivery service".[21] The library "places a particular emphasis" on the acquisition of grey literature. The TIB holdings comprise around 8.2 million items (print, microforms, CD-ROMs), a figure that includes conference proceedings, research reports, patents, norms and standards, and dissertations.

The TIB participates in a large number of projects, with a key focus on electronic publications and the development of a digital library. Some current projects are on grey literature and/or open access:[22]

- CODATA: contribution to the publication and citation of scientific primary data through attribution of DOI/URNs.
- ELAN Niedersachsen: development of e-learning services (metadata model, content integration into online access systems, cross-linking with standard services, long-term preservation in an OA environment).

[20] http://science.knowledge.go.kr
[21] http://www.tib-hannover.de/en/about_us/overview
[22] http://www.tib-hannover.de/en/about_us/projects/?p=on

- Probado: development of a service (preservation, handling, search, delivery) for complex, non-textual data and documents (e-learning, music, 3-dimensional computer graphics).

Concluding remarks

This communication is limited to an overview of the focus of the study and of a global description of the general mission, holdings, acquisition policy and OA projects of the five institutions. Some preliminary results:

- All libraries place special emphasis on grey literature and have important grey collections, especially of conference proceedings, reports and dissertations.
- All structures are more or less involved in OA projects. Nevertheless, the significance and range of these projects depend on their specific mission and role in the national context. At least two structures – CISTI and INIST – participate actively in the development of national institutional repositories.
- Some of these projects are directly linked to grey literature, especially to ETDs and preprints but also to courseware, primary data and reports.

These first results are based on an analysis of web-based and published information. They will be amended and completed in the next months by the feedback on our survey on document supply and OA projects.

References

Aubry, C. and Janik, J. (2005), *Les archives ouvertes: Enjeux et pratiques*. ADBS, Paris.

Bador, P., Boukacem-Zeghmouri, C., Lafouge, T., Prost, H. and Schöpfel, J. (to be published), "Analysis of the downward trend in document delivery in pharmaceutics: a case study from INIST in France". *Interlending & Document Supply*.

Banks, M. (2004), "Connections between open access publishing and access to gray literature". *Journal of the Medical Library Association*, vol. 92, no. 2, pp. 164-166.

Banks, M. (2006), "Towards a Continuum of Scholarship: The eventual Collapse of the Distinction Between Grey and non-Grey Literature?", in Farace, D. and Frantzen, J. (Ed.), *GL7 Conference Proceedings. Seventh International Conference on Grey Literature: Open Access to Grey Resources.* Nancy, 5-6 December 2005. TextRelease, Amsterdam.

Brindley, L. (2005), "The British Library: its origins, development and future". *Interlending & Document Supply*, vol. 33, no. 2, pp. 76-80.

The British Library (2005), *32nd Annual Report and Accounts 2004/2005*. The Stationary Office, London.

Brown, D. (2003), "Is this the end of the 'article economy'? A strategic review of document delivery". *Interlending & Document Supply*, vol. 31, no. 4, pp. 253-263.

Ceeney, N. (2004), "The British Library – delivering the world's knowledge: some recent developments". *Interlending & Document Supply*, vol. 32, no. 4, pp. 234-238.

Choi, W. (2003), "The Development of Digital Libraries in South Korea". *Libri*, vol. 53, pp. 130-141.

Claerebout, M.F. (2004), "Grisemine, a digital library of grey university literature", in Farace, D. and Frantzen, J. (Ed.), *GL5 Conference Proceedings. Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information.* Amsterdam, 4-5 December 2003. TextRelease, Amsterdam, pp. 27-31.

Crawford, S. Y., Hurd, J. M. and Weller, A. C. (1996), *From print to electronic: the transformation of scientific communication*. Silver Spring, ASIS Monograph Series.

Grüttemeier, H. (to be published), "The Way to Open Access – French Strategies to Move Forward". *Tushu Qingbao Gongzuo (Library and Information Service)*.

Grüttemeier, H. and Mahon, B. (2003), *Open Access to Scientific and Technical Information. State of Art and Future Trends*. ICSTI/INIST/INSERM Seminar, 23-24 January 2003, Paris, France. IOS Press, Amsterdam.

Guédon, J.C. (2004), "The 'green' and 'gold' roads to Open Access: the case for mixing and matching". *Serials Review*, vol. 30, no. 4, pp. 315-328.

Hartley, J. (2004), "On requesting conference papers electronically". *Journal of Information Science*, vol. 30, no. 5, pp. 475-479.

Jackson, M.E. (2004), "Document access in the United States: revisited 1978-2004". *Interlending & Document Supply*, vol. 32, no. 3, pp. 164-168.

Krym, N. and VanBuskirk, M. (2001), "Resource-sharing roles and responsibilities for CISTI: change is the constant". *Interlending & Document Supply*, vol. 29, no. 1, pp. 11-16.

Law, D. (2004), "Tilting at windmills: BLDSC and the UK higher education community". *Interlending & Document Supply*, vol. 32, no. 2, pp. 85-89.

Millar, I. (2001), "Nouvelles orientations pour la British Library". *Bulletin des Bibliothèques de France*, vol. 46, no. 6, pp. 86-92.
Molgat, L. (2005), "Transforming Document Delivery in the e-Content Environment". World Library and Information Congress: 71th IFLA General Conference and Council. August 14th-18th 2005, Oslo, Norway.
Needham, P. (2002), "MAGiC: Shining a new light on a grey area". *Serials*, vol. 15, no. 3, pp. 201-206.
Paillassard, P., Schöpfel, J. and Stock, C. (2005), "How to get a French doctoral thesis, especially when you aren't French". *Publishing Research Quarterly*, vol. 21, no. 1, pp. 73-93.
Pintat, R. (2003), "Archives ouvertes: une alternative à la subordination des bibliothécaires aux éditeurs numériques?" *Bulletin des Bibliothèques de France*, vol. 48, no. 2, pp. 90- 94.
Romary, L. (2005), "The way to open resources". *Berlin 3 Open Access: Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, University of Southampton, Southampton, 28 February-1 March, available at: http://www.eprints.org/berlin3/
Roosendaal, H. E. (2004), "Driving change in the research and HE information market". *Learned Publishing*, vol. 17, no. 1, pp. 1-6.
Rowse, M. (2003), "Individual article supply: some strategic directions". *Interlending & Document Supply*, vol. 31, no. 2, pp. 86-93.
Schöpfel, J. (2003), "INIST-CNRS in France: 'a model of efficiency'". *Interlending & Document Supply*, vol. 31, no. 2, pp. 94-103.
Schöpfel, J. (2005), "Between open access and copyright: document supply in France". *Interlending & Document Supply,* vol. 33, no. 3, pp. 158-161.
Shin, E.J. and Oh, K.M. (2002), "Interlending and document supply developments in South Korea". *Interlending & Document Supply,* vol. 30, no. 3, pp. 136-138.
Smith, G. (2001), "Document delivery as an aspect of co-operation: the British Library's Co-operation and Partnership Programme". *Interlending & Document Supply,* vol. 29, no. 3, pp. 114-119.
Stock C. and Schöpfel J. (2004), "Grey Literature in an Open Context: From Certainty to New Challenges", in Farace, D. and Frantzen, J. (Ed.), *GL5 Conference Proceedings. Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information.* Amsterdam, 4-5 December 2003. TextRelease, Amsterdam, pp. 199-202.
Stock, C., Rocklin, E. and Cordier, A., "LARA – open access to scientific and technical reports". Farace, D. and Frantzen, J. (Ed.), *GL7 Conference Proceedings. Seventh International Conference on Grey Literature: Open Access to Grey Resources.* Nancy, 5-6 December 2005. TextRelease, Amsterdam.
Sondergaard, T. F., Andersen, J. and Hjorland, B. (2003), "Documents and the communication of scientific and scholarly information. Revising and updating the UNISIST model". *Journal of Documentation*, vol. 59, no. 3, pp. 278-320.
VanBuskirk, M. and Krym, N. (2003), "Resource-sharing roles and responsibilities for CISTI: for better or for ILL?" *Interlending & Document Supply*, vol. 31, no. 3, pp. 169-173.
Zhang, Y., Lee, K. and You, B.J. (2001), "Usage patterns of an electronic theses and dissertations system". *Online Information Review*, vol. 25, no. 6, pp. 370-378.

Acknowledgments to H. Prost for providing statistical data on INIST document supply and to N. Hull for a careful review of the English version.

Author Information

Chérifa Boukacem graduated with a Master's Degree in library sciences from the University of Algiers (DZ) in 1996. She obtained her Ph.D. from the University of Lyon-2 in 2004 with a comparative study on French academic interlibrary loan and access to digital resources. From 2001 to 2003, she worked at the department of scientific and technical information of the University of Lyon-1, she teaches since 2000 library and information science to students and professionals, is member of the French research group in library sciences, DOCSI/GRESI, and co-edits a digital serials, Konex. Her main interests are on e-publishing and multimedia, library software, statistics and scientometrics and professional counselling. Actually, she is working on the transformation of the digital library and on a longitudinal study of document supply, impact factor and the switch from print to electronic resources.
Email: boukacemc@yahoo.fr

Joachim Schöpfel graduated from the University of Hamburg in 1984. A research assistant and lecturer at the University of Hamburg, Dept. of Developmental and Educational Psychology from 1985 to 1990. He obtained his Ph.D. from the same university in 1992. He is presently head of the library and document delivery department at the French Institute of Scientific and Technical Information (INIST-CNRS). He taught Culture and Society (1992-2001) and since 2001, teaches Documentation at the University of Nancy. He is member of the UK Serials Group, associate editor of TGJ, a member of the Editorial Board for Interlending & Document Supply, and a member of the former EAGLE, European Association for Grey Literature Exploitation. Email: schopfel@inist.fr

Access to Grey Content: An Analysis of Grey Literature based on Citation and Survey Data: A Follow-up Study

Dominic J. Farace
GreyNet, Netherlands

Jerry Frantzen
Boer & Croon, Netherlands

Joachim Schöpfel and Christiane Stock
INIST-CNRS, France

Albert K. Boekhorst
UvA, Netherlands; UP, South Afirca

Introduction

Grey literature, an area of interest to special librarians and information professionals, can be traced back a half-century. However, grey literature as a specialized field in information studies is less than a decade old. At GL'97 in Luxembourg, grey literature was redefined "as information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishers (i.e. where publishing is not the primary activity of the producing body)." The subject area was broadened and the need for continuing research and instruction pursued. The results of an online survey carried out in 2004 compared with survey results a decade prior indicate two changes: (1) a move to more specialization in the field of grey literature and (2) a move to more balance in activities related to research and teaching as compared with the processing and distribution of grey literature. It is not that the activities of processing and distribution are today of less concern, but technological advances and the Internet may have made them less labour intensive. The burden that grey literature poised to human resources and budgets appears to have been reduced enough that the benefits of its content is discovered. And this discovery of a wealth of knowledge and information is the onset to further research and instruction in the field.

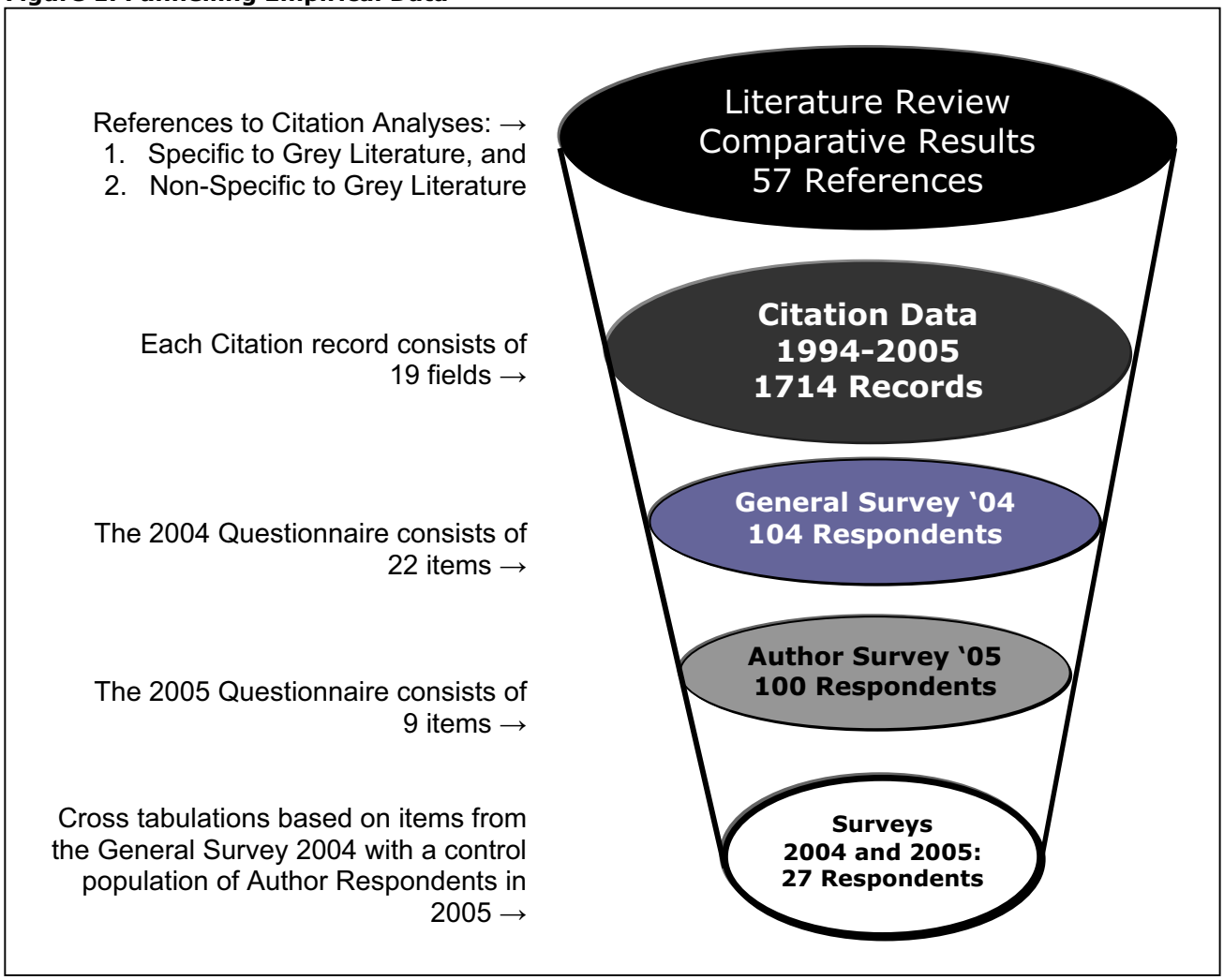
Research Goal

The idea behind this study is that - by using the same pool of authors - survey data linked to citation data will allow for a clearer demonstration of the impact of their research, where only part of the impact is covered by citation analysis alone. Hopefully, the new combined results will provide a better profile of these meta-authors, who are also the source of GreyNet's knowledge and information base. This could lead to the subsequent development of information policies and services that are more in line with the needs of authors and researchers, whereby their results would become even more accessible well beyond the grey circuit.

Research Plan

This research is a follow-up to two projects carried out in 2004. One was a citation analysis based on the published papers in the GL Conference Proceedings and the other was a general survey, which dealt with the response of information professionals to key issues and topics in the field of grey literature. In this study, we seek not only to update and integrate the data from the citation analysis but at the same time to introduce the instrument of an author survey in order to better assess the work and expectations of those who are actually doing research and authoring papers on the topic of grey literature. These are referred to as the meta-authors of grey literature.

Figure 1. Funnelling Empirical Data



I. Results of Citation Analysis 1994-2005

It is not the intention here to repeat the findings of last year's research [(22)], which was a more textual account and analysis, but rather suffice to provide another format in order to present the cumulative results of the research. To this end, tables and their explanations are emphasized. Further, we find that once the citation database had been updated with the records from the GL6 Conference Proceedings (2005), new trends and developments can be identified. And, it is these that could have a marked influence on access to grey content issuing from the conference series.

Table 1.
General Citation Data

Conf No.	No. of papers	Papers without citations	No. of papers with citations	Total No. of citations	Average No. of citations per paper
1	37	9	28	345	12.3
2	25	4	21	247	11.8
3	29	10	19	275	14.5
4	28	4	24	250	10.4
5	20	3	17	227	13.4
6	24	0	24	370	15.4
Total:	**163**	**30**	**133**	**1714**	**13.0**

Unlike previous conferences in the series, all of the conference papers in the GL6 Proceedings without exception contained references. These same proceedings claim the highest number of citations (370) irrespective of the number of conference papers. And, these proceedings maintain the highest average number of citations (15.4) per conference paper.

Table 2.
Standard or Hyperlink Citations

Conf No.	Citations total	Standard citations	Hyperlinked citations	Explanative notes	Name & address
1	345	322	0	15	8
2	247	243	2	2	0
3	275	207	63	4	1
4	250	160	76	14	0
5	227	155	67	5	0
6	370	195	162	12	1
Total:	**1714**	**1282**	**370**	**52**	**10**

The GL6 Conference Proceedings not only had the highest average number of hyperlinked citations compared with standard citations from previous conferences in the series, but also the total number nearly doubled compared with that of the year prior. However, this increase in hyperlinked citations was not at the cost of standard citations, because the overall number of citations per conference paper had increased. Noticeably, what did not increase was the quality of the hyperlinked citations. A considerable number of which only show a URL without further description of the source. This may bear out what Chu[20] infers by hyperlinked citations being different from standard citations in that they point more to resources rather than they support or refute academic research.
Further what we find in the search of the citation database, but which is not shown in the table above, is the increase of citations to grey publications compared to commercial publications. This distinction is based on document type and can again be explained by the increase in hyperlinked citations, which referred mostly to WebPages [(1)] and Web papers available through non-commercial publishers. It is then important to researchers and librarians than an accurate link between publications and their references are made. [(2)]

Table 3.
Serial Citations

Conf No.	No. of papers with citations	Citations total	No. of serial citations	Maximum Citations per paper	Minimum Citations per paper
1	28	345	0	69	1
2	21	247	23	73	1
3	19	275	17	62	1
4	24	250	13	27	1
5	17	227	26	31	2
6	24	370	58	35	1
Total :	**133**	**1714**	**137**		

Serial citations (i.e. citations to previous conference papers in the GL Series) have not only doubled in total number every year for the past three conferences but also show an average increase of more than 5% for each of the same past three consecutive conferences - from 5% in the 4th to 11.5% in the 5th to 15.7% in the 6th. This may indicate not only more access to previous conference papers and/or proceedings in the GL-Series but also further use and application of research results originating from within this Conference Series.

Table 4.
Self-Citations

Conf no	No. of papers	Citations total	No. of Self-citations	No. of Non-Self citations	Not applicable*
1	28	345	42	231	72
2	21	247	15	189	43
3	19	275	19	175	81
4	24	250	18	134	98
5	17	227	40	128	59
6	24	370	37	221	112
Total:	**133**	**1714**	**171**	**1078**	**465**

The number of self-citations in the GL6 Proceedings (10%) appears to be declining to the level of the first 4 conferences, which together averaged 8%. These conferences were prior to the 2000-2003 break in the series. After its relaunch with the 5th conference, the meta-authors may have had to rely in that conference year on their own findings to substantiate arguments, since they did not have ready access to previous conference papers nor to GreyNet, the Grey Literature Network Service, which was also dormant in that same four-year period.

Table 5.
Age of Citations

Conf no	Citations total	No. of citations in Year of Conf.	No. Minus 1 Year	No. Minus 2 Years	No. Minus 3 Years	Earliest year of cited work
1	345	83	37	22	13	1949
2	247	60	71	15	26	1944
3	275	86	41	18	22	1945
4	250	89	32	26	13	1886
5	227	64	28	21	16	1949
6	370	177	28	19	11	1896
Total	**1714**	**559**	**237**	**121**	**101**	

Another significant figure from the citation data of GL6 is the sharp increase in the number of citations dated the same year of the conference, which was 47,8%. Looking at the overall average of the first 5 conferences in the GL Series, the average was 28,4%. Once again, this increase not only illustrates a trend in research to cite current work but is also influenced by the ratio of hyperlinked citations that carry the date in which the conference is held.
While it is beyond the scope of this current paper, future analysis of the citation data once the GL7 records have been entered in the database may reveal further trends and distinctions between standard and hyperlinked citations as they impact and influence work by meta-authors in the field of grey literature.

II. Results of the Author Survey 2005

If we now turn to the results of the author survey carried out in 2005, a brief word on the population of the respondents show that they are all past or present authors in the GL-Conference Series. The total population of these meta-authors (i.e. informational professionals working in the field of grey literature and doing research and authoring papers) since the start of the GL conference series in 1993 is roughly 230. Initially, there were 103 respondents to this online survey. However, three of them withdrew their content submission to GL7 and subsequently were deleted from the survey bringing the total number of respondents to an even hundred.

Table 6.
Continent where the Author lives and works

	Frequency	Percent
North America	35	35,0
Europe	52	52,0
Asia	8	8,0
Other	5	5,0
Total	100	100,0

Across the board, whether looking at citation data such as cited works and citing authors or whether looking at the respondents to the general survey in 2004 or this Author Survey in 2005, North America and Europe account for 85% to 90% of global activity in the field of grey literature.

Table 7.
Citation Style for Grey Literature would be of benefit for the author's work?

	Frequency	Percent
Yes	55	55,0
No	21	21,0
Depends	11	11,0
NA	13	13,0
Total	100	100,0

While more than 50% of the authors respond with a simple yes to this open question, another 11% would be inclined to such guidelines as long as it would not complicate and duplicate their work at hand. Such guidelines should be in general use supported by a global community and in place for multiple types of grey literature.

Table 8.
Commercial publisher accepted one or more of their works?

	Frequency	Percent
Yes	49	49,0
No	38	38,0
Depends	5	5,0
NA	8	8,0
Total	100	100,0

Nearly 50% of the authors had one or more of their manuscripts accepted by a commercial publisher. Another 5% is unclear - depending on crossover situations - where a grey publisher was taken over by a commercial publisher or where a publication has moved into the realm of OAI.

Table 9.
Author has published on other topics than Grey Literature?

	Frequency	Percent
Information Science	59	59,0
Other subjects	25	25,0
NA	16	16,0
Total	100	100,0

While 84% of the authors published on other topics than grey literature, 59% of them remained within the field of information science. The other 25% published in a variety of different fields in the natural sciences, social sciences, and humanities.

Table 10.
Author's view on Open Access

	Frequency	Percent
Positive (unqualified)	66	66,0
Positive (qualified)	29	29,0
NA	5	5,0
Total	100	100,0

Only 5% of the authors did not respond to this question, while 95% are favourable to Open Access. A near two-thirds sufficed with a simple statement, while 29% provided more lengthy and qualified arguments for their position. Their positions ranged the full gamut from views held and published by the Wellcome Trust[*] to those of The Royal Society[**].

III. Comparative Results of Survey and Citation Data

In the first two parts of this paper, we looked separately at results of citation data and survey data. It is our intention in this final part of the paper to present some comparative results, as they appear from cross-tabulations of the 2004 and 2005 Surveys and the extent to which other combined data lend themselves to empirical observation.
A selection of 5 items from the 2004 Survey was made in an effort to determine if differences exist in the responses between those informational professionals simply working in the field of grey literature and the meta-authors (i.e. those who are both working in this field of information and who are also doing research and authoring publications on the topic of grey literature). The meta-authors in this research totalled twenty-seven. They in fact were the population of respondents, who completed both the 2004 General Survey and the 2005 Author Survey. Once the five items for cross-tabulation were chosen, it was then necessary to delete the 27 respondents from the 2004 Survey population so that the results of the one group would not influence the outcome of the other. This resulted in two groups of respondents having survey populations of 77 and 27 respectively. A check shows that the total number of respondents for each item is 104. However, due to some changes in the recoding of responses, minor discrepancies appear in the subtotals. Since these are only indicative results, we choose to share the findings as they are calculated in the tables below.

Table 11.
The average net-user should at least recognise the term 'grey literature'

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**0**	**7**
	,0%	25,9%
NA	**8**	**0**
	10,4%	,0%
No	**29**	**7**
	37,7%	25,9%
Yes	**40**	**13**
	51,9%	48,1%
Total	**77**	**27**
	100,0%	100,0%

On this open-ended item, we see almost an inverse relationship between the meta-authors and the non-meta authors regarding the percentage of no-answers (NA) and qualified statements (Depends). While there was little variation in the percentage that agreed, there was almost a 12-percentage point difference in those who disagreed. The Meta-authors are more convinced that the term grey literature belongs in the vocabulary of net-users.

[*] http://www.wellcome.ac.uk/print/wtd002766_print.html
[**] http://www.royalsoc.ac.uk/page.asp?id=3882

Table 12.
Grey Literature should be free to access

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**13**	**7**
	16,9%	25,9%
NA	**7**	**1**
	9,1%	3,7%
No	**12**	**2**
	15,6%	7,4%
Yes	**45**	**17**
	58,4%	63,0%
Total	**77**	**27**
	100,0%	100,0%

While the majority of both groups of respondents favour free access to grey literature, the meta-authors lead by 5-percentage points in their agreement on this particular survey item.

Table 13.
Grey Literature should be free of charge

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**17**	**10**
	22,1%	37,0%
NA	**6**	**2**
	7,8%	7,4%
No	**14**	**4**
	18,2%	14,8%
Yes	**40**	**11**
	51,9%	40,7%
Total	**77**	**27**
	100,0%	100,0%

Over half of the Non-meta authors feel that grey literature should be free of charge, while the meta-authors were not in the majority on this standpoint. Instead, they chose more often to qualify their response allowing for differences in the sector in which grey literature is produced, the size of the corporate author and/or producing body, financial position, etc.

Table 14.
Grey Literature itself constitutes a field in information studies

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**4**	**4**
	5,2%	14,8%
NA	**7**	**1**
	9,1%	3,7%
No	**14**	**2**
	18,2%	7,4%
Yes	**52**	**20**
	67,5%	74,1%
Total	**77**	**27**
	100,0%	100,0%

While the majority of both groups of respondents agree that grey literature constitutes a specialized field in information studies, it is not surprising to find that the meta-authors carry almost a 7-percentage point lead on this survey item.

Table 15.
The Luxembourg Convention on Grey Literature still holds

	Information Professionals	
	Non-Meta Author	Meta-author
Depends	**7**	**2**
	9,1%	7,4%
NA	**4**	**2**
	5,2%	7,4%
No	**10**	**4**
	13,0%	14,8%
Yes	**56**	**19**
	72,7%	70,4%
Total	**77**	**27**
	100,0%	100,0%

Not only did both groups of respondents overwhelming favour the current definition of grey literature known as the 'Luxembourg Convention' but this survey item also shows the least amount of variation in percentages between the two groups.

IV. Summary of Findings and Conclusion

In order to be clear on the results, which are based on different types of data applied in this study, separate subheadings are used below.

Based on Citation Data:

- Hyperlink citations are rapidly gaining ground on standard citations
- Hyperlink citations tend to increase the total number of citations in a conference paper
- Hyperlink citations are also increasing the number of references to grey literature
- Self-citations are decreasing, while serial citations are increasing
- Nine of the top-ten cited authors are also meta-authors in the GL Conference Series

Based on Survey Data:

- Nearly half of the meta-authors also make use of commercial publishers
- More than three quarters of the meta-authors also publish on other topics than GL
- Without reservation, nearly two-thirds of the meta-authors favour OAI
- However, nearly one-third of the meta-authors provide qualified statements on OAI, thus requiring GreyNet to further analyse these responses before rendering a position statement[***].

Based on Comparative Data:

Differences of opinion were uncovered between meta-authors and non-meta authors:

- Meta-authors were significantly more inclined to qualify their statements
- Non-meta authors were significantly more inclined not to respond to a given question
- Analysis of the top-5 types of grey literature resulting from both the citation data and the survey data show that four of them are the same, namely: conference papers, journal articles, reports, and WebPages.
- However, significant differences appear in the meta-authors' production and use of these types of grey literature. Reports are first to be produced, while fourth in line to be cited.

In close, the literature review uncovered a wealth of citation formats available for grey literature. However, the differences in formats and uses of hyperlinked and standard citations require further research and development. Perhaps the 'Nancy Style' proposed at GL7 will offer a framework for this and other best practices in the field of grey literature.

[***] GreyNet's position on OAI is scheduled for publication in the Editor's Note of The Grey Journal, TGJ volume 2, number 1, Spring 2006. – ISSN 1574-1796.

References and Sources

This is a follow-up study; references from the earlier study are not repeated here. They can be found in Ref. 22, below.

Bibliographic References Specific to Grey literature

1. Davis, Philip M. (2002), The effect of the Web on undergraduate citation behavior: a 2000 update. - In: Collection & Research Libraries, v. 63, n. 1, pp. 53-60.
2. Morrisey, Locke J. (2002), Bibliometric and bibliographic analysis in an era of electronic scholarly communication. - In: Science & Technology Libraries, v. 22, n. 3-4, pp. 149-160.
3. Kraus, Joseph R. (2002), Citation patterns of advanced undergraduate students in Biology, 2000-2002. - In: Science & Technology Libraries, v. 22, n. 3/4, pp. 161-179.
4. Brown, Cecelia (2003), The role of electronic preprints in chemical communication: analysis of citation, usage, and acceptance in the journal literature. - In: Journal of the American Society for Information Science and Technology, v. 54, n. 5, pp. 362-371.
5. Pelzer, Nancy L. and William H. Wiese (2003), Bibliometric study of grey literature in core veterinary medical journals. - In: Journal of the Medical Library Association, v.91, n.4, pp. 434-441.
6. Hasbrouck, La Mar, Joanna Tallano, Jon Mark Hirshon, and Andrew L. Dannenberg (2003), Use of epidemiology in clinical medical publications, 1983-1999: a citation analysis. - In: American Journal of Epidemiology, v.157, n.5, pp. 399-408.
7. Garg, K.C. (2003), An overview of cross-national, national, and institutional assessment as reflected in the international journal Scientometrics. - In: Scientometrics, v. 56, n. 2, pp. 169-199.
8. Hawkins, Donald T., Signe E. Larson, and Bari Q. Caton (2003), Information science abstracts: tracking the literature of information science. Part 2: A new taxonomy for information science /. In: Journal of the American Society for Information Science and Technology, v. 54, n. 8, pp. 771-781.
9. Sondergaard, Trine, Jack Andersen, and Birger Hjorland (2003), Documents and the communication of scientific and scholarly information. - In: Journal of Documentation, v. 59, n. 3, pp. 278-320.
10. Valderrama Zurian, J.C., R. Aleixandre, and M. Castellano (2004), Citation-count analysis in Addiction 2001. - In: Addiction, v. 99, n. 3, pp. 387-388.
11. Waugh, C. Keith and Margie Ruppel (2004), Citation analysis of dissertation, thesis, and research paper references in workforce education and development. - In: The Journal of Academic Librarianship, v. 30, n. 4, pp. 276-284.
12. Schaffer, Thomas (2004), Psychology citations revisited: behavioural research in the age of electronic resources. - In: The Journal of Academic Librarianship, v. 30, n. 5, pp. 354-360.
13. Ullah, Mohd. Furqan, S.S. Kanwar, and Pradeep Kumar (2004), A quantitative analysis of citations of research reports published by National Institute of Hydrology, Rorkee. - In: Annals of Library and Information Studies, v. 51, n. 3, pp. 108-115.
14. Sharif, Muhammad Ashraf and Khalid Mahmood (2004), How economists cite literature: citation analysis of two core Pakistani economic journals. - In: Collection Building, v. 23, n. 4, pp. 172-176.
15. Valderrama, Rafael Juan Carlos, José Maria Desantes, and Antonio J. Torregrosa (2004), Identification of information sources and citation patterns in the field of reciprocating internal combustion engines. - In: Scientometrics, v. 59, n. 3, pp. 321-336.
16. Beile, Penny M., David N. Boote, and Elizabeth K. Killingsworth, E.K. (2004), A microscope or a mirror? : A question of study validity regarding the use of dissertation citation analysis for evaluating research collections. - In: The Journal of Academic Librarianship, v. 30, n. 5, pp. 347-353.
17. Bremholm, Tony L. (2004), Challenges and opportunities for bibliometrics in the electronic environment: the case of the Proceedings of the Oklahoma Academy of Science. - In: Science & Technology Libraries, v. 25, n.1-2, pp. 87-107.
18. Chen, Chaomei and Diana Hicks (2004), Tracing knowledge diffusion / In: Scientometrics, v. 59, n. 2, pp. 199-211.
19. Swygart-Hobaugh, Amanda J. (2004), A citation analysis of the quantitative/qualitative methods debate's reflection in sociology research: implications for library collection development. - In: Library Collections, Acquisitions, & Technical Services, v. 28, n. 2, pp. 180-195.
20. Chu, Heting (2005), Taxonomy of inlinked Web entities: what does it imply for webometric research? - In: Library & Information Science Research, v. 27, n. 1, pp. 8-27.
21. Bollen, Johan, Herbert van de Sompel, Joan A. Smith, and Rick Luce (2005), Toward alternative metrics of journal impact: a comparison of download and citation data. - In: Information Processing & Management, v. 41, n. 6, pp. 1419-1440.
22. Schöpfel, J., C. Stock D.J. Farace and J. Frantzen (2005), Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit. - In: GL6 Conference Proceedings : Sixth International Conference on Grey Literature : Work on Grey in Progress, 6-7 December 2004. – Amsterdam : TextRelease, 2005, p. 55-63. – ISBN 90-77484-04-3.

Bibliographic References Non-specific to Grey literature

23. Davis Herring, Susan (2002), Use of electronic resources in scholarly electronic journals: a citation analysis. - In: College & Research Libraries, v. 63, n. 4, pp. 334-340.
24. Kotiaho, Janne S. (2002), Ethical considerations in citing scientific literature and using citation analysis in evaluation of research performance. - In: Journal of Information Ethics, v. 11, n. 2, pp. 10-16.
25. Kostoff, Ronald N. (2002), Citation analysis of research performer quality. - In: Scientometrics, v. 53, n. 1, pp. 49-71.
26. Zhao, Dangzhi and Elisabeth Logan (2002), Citation analysis using scientific publications on the Web as data sources: a case study in the XML research area. - In: Scientometrics, v. 53, n. 4, pp. 449-472.
27. Cronin, Blaise and Debora Shaw (2002), Identity-creators and image-makers: using citation analysis and thick description to put authors in their place. - In: Scientometrics, v. 53, n. 1, pp. 31-49.
28. Koehler, Wallace (2002), Nightmares in citation analysis. - In: Reference and User Services Quarterly, v. 42, n. 1, pp. 41-42.

Sources used for the Collection of Citation Data

29. Farace, D.J. and J. Frantzen (1994-….), Conference proceedings on grey literature: GL-conference series. – ISSN 1386-2316
Amsterdam : GreyNet/TransAtlantic, 1994-2000; Numbers 1-4
Amsterdam : GreyNet/TextRelease, 2003-2005; Numbers 5-6

Author Information

Dominic J. Farace is Director of TextRelease, an Amsterdam based information bureau specializing in grey literature and networked information. He is a native Louisianan and holds two degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked for the past twenty-nine years. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (Swidoc/KNAW), he founded GreyNet, Grey Literature Network Service, in 1993 and has since been responsible for the international GL-Conference Series. In this capacity, he serves as Program and Conference Director as well as managing editor of the conference proceedings. Since 2004, he is a Guest Lecturer on Grey Literature in the Masters Program at the University of Amsterdam, Faculty of Humanities. He is Guest Editor for PRQ, Publishing Research Quarterly and the editor of The Grey Journal. Email: info@greynet.org

Jerry Frantzen graduated in 1999 from the College of Amsterdam in Library and Information Science. He is presently employed with Boer and Croon Strategy and Management Group in Amsterdam and is technical editor of The Grey Journal

Joachim Schöpfel graduated from the University of Hamburg in 1984. A research assistant and lecturer at the University of Hamburg, Department of Developmental and Educational Psychology, from 1985 to 1990, he obtained his Ph.D. from the same university in 1992. He is presently head of the library and document delivery department at the French Institute of Scientific and Technical Information and teaches Culture and Society (1992-2001) and Documentation (from 2001 on) at the University of Nancy. He is member of the UK Serials Group, associate editor of TGJ, a member of the Editorial Board for Interlending & Document Supply, and a member of the former EAGLE, European Association for Grey Literature Exploitation.

Christiane Stock graduated from the University of Freiburg in 1984. She joined INIST-CNRS the French Institute of Scientific and Technical Information in 1989. Member of the Technical Committee for the SIGLE database since 1993, she also set up the national agency for the ISRN (International Standard Report Number). Today she is the head of the monographs and grey literature section at INIST.

Albert K. Boekhorst (1943) studied sociology at the University of Amsterdam. Since 1987 he is working at the dept of Information Science at the same university. He is visiting professor at the University of Pretoria (South Africa) and also teaches at the University of Peking (China) and at the Pedagogical University of Tallinn (Estonia). He teaches courses on the role and infrastructure of information and knowledge in organisations. In his research he concentrates on theoretical aspects of the role of information in societies in general and the access to information more specifically. Information literacy is a special field of interest in this context. More information on his work is to be found at www.hum.uva.nl/akb. Comments to boekhorst@uva.nl

Appendices

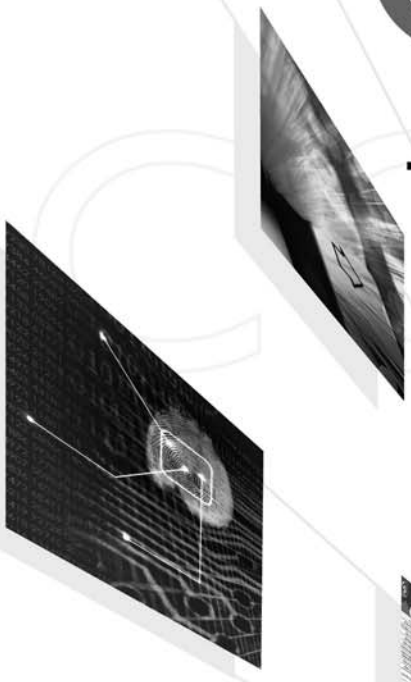
List of Participating Organizations

American University of Antigua, AUA	West Indies
Battelle Memorial Institute	United States
Boer & Croon, Strategy and Management Group	Netherlands
British Library, BL	United Kingdom
Bundelkhand University	India
Centre National de Recherche Scientifique, CNRS	France
CCLRC - Rutherford Appleton Laboratory	United Kingdom
Comissariat à l'Energie Atomique, CEA	France
Dartmouth College; I3P	United States
EBSCO Information Services	United States
European Organization for Nuclear Research, CERN	Switzerland
Grey Literature Network Service, GreyNet	Netherlands
Information and Library Network Centre, INFLIBNET	India
Institut National de Recherche Pédagogique, INRP	France
Institut National des Sciences Appliquées de Lyon, INSA	France
Institut de l'Information Scientifique et Technique, INIST	France
International Atomic Energy Agency, IAEA	Austria
International Information Associates, IIa	United States
Istituto di Ricerche sulla Popolazione e le Politiche Sociali, IRPPS-CNR	Italy
Istituto di Scienza e Tecnologie dell'Informazione, ISTI-CNR	Italy
Istituto Superiore di Sanità, ISS	Italy
Japan Science and Technology Agency, JST	Japan
Korea Institute of Scientific and Technical Information, KISTI	South Korea
Library of Zeeland, ZEBI	Netherlands
LORIA, Campus Scientifique	France

New York Academy of Medicine, NYAM	United States
New York Public Library, NYPL	United States
New York University School of Medicine, NYU	United States
Office for Official Publications of the European Communities	Luxembourg
Office of Scientific and Technical Information, OSTI	United States
Organisation for Economic Co-Operation and Development, OECD	France
Portland State University, PSU	United States
Pusan National University	South Korea
Regional library of Science and Technology, RLST	Iran
Saarland University and State Library	Germany
Swiss Federal Office of Public Health, SFOPH	Switzerland
Technische Informationsbibliothek, TIB Hannover	Germany
TextRelease, Professional Information Services	Netherlands
United States Department of Energy, DOE	United States
Université de Nancy 2	France
Université Henri Poincaré, UHP	France
Université Paris 8	France
University of Amsterdam, UvA	Netherlands
University of Bergen, UiB	Norway
University of California, Irvine Libraries, UCI	United States
University of Granada	Spain
University of Lyon	France
University of Murcia	Spain
University of Pretoria, UP	South Africa
Virginia Reid Moore Marine Research Library – CMA	United States

CORDIS

The Community Information Service on European Research and Innovation activities



CORDIS is a unique service for strengthening European competitiveness. An interactive platform, CORDIS assists its users to access European funding and encourages technology and knowledge transfer between EU Member states, accession, candidate and associated countries.

http://cordis.europa.eu.int

Research and Innovation
Forschung und Innovation
Recherche et innovation
Ricerca e innovazione
Teadusuuringud ja innovatsioon
Tutkimus ja innovaatio
Kutatás és innováció
Rannsóknir og nýsköpun
Pētniecība un inovācija
Tyrimai ir inovacijos
Riċerka u innovazzjoni
Onderzoek en innovatie
Forskning og utvikling
badania i innowacje
Investigação e Inovação
Výzkum a inovace
Raziskave in inovacije
Forskning och utveckling

CORDIS
1990 — 2005

Publications Office

Index to Authors