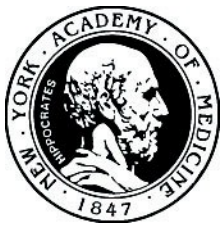
Eleventh International Conference on Grey Literature

The Grey Mosaic, Piecing it All Together

Sponsors:

Conference Proceedings

ISSN 1386-2316



GL11 Program and Conference Bureau

TextRelease

Javastraat 194-HS, 1095 CP **A**msterdam, **T**he **N**etherlands
www.textrelease.com ▪ **conference@textrelease.com**
Tel/Fax +31 (0) 20-331.2420

CIP

GL11 Conference Proceedings
Eleventh International Conference on Grey Literature : The Grey Mosaic, Piecing It All Together, 14-15 December 2009 / compiled by Dominic J. Farace and Jerry Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, February 2010. – 137 p. ; 30 cm. – Author Index. – (GL-Conference series, ISSN 1386-2316 ; No. 11)
The British Library, INIST-CNRS, NYAM, and the National Technical Library (NTK) are corporate authors and associate members of GreyNet. These conference proceedings contain the full text of some eighteen papers presented during the two days of Plenary and Parallel Sessions. The papers appear in the same order as in the conference program. Included is a List of Participating Organizations, and Sponsored Advertisements.
ISBN: 978-90-77484-13-5

Foreword

The Grey Mosaic, *Piecing It All Together*

Over the past 15 years, Grey Literature has developed from a millennium movement to a well defined field in information studies. This process has been considerably influenced and shaped by results of research issuing from the International Conference Series on Grey Literature. The Eleventh Conference in this series endeavors to piece together traditional features inherent to grey literature with more recent elements both technology and policy driven. In any field of science and technology, information and knowledge aggregated in research must be made available beyond the limits of any one specific information community and should be openly accessible to net citizens. This principle related to grey literature requires that its uses and applications in diverse subject areas benefit information society as a whole. Furthermore, net citizens must acquire a sense of safeguard and security that the grey resources on which they come to rely have passed some level of corporate governance. In piecing the various components and facets of grey literature together, GL11 depicts a virtual image of the Grey Mosaic.

On behalf of the Conference Host and Sponsors, the Program Committee Members and Chairpersons, I would like to thank the authors and co-authors for their content contributions to these proceedings. Likewise, I welcome those reading these conference proceedings to voice their comments and/or recommendations either directly to the authors or to GreyNet International.

Dr. Dominic J. Farace
Grey Literature Network Service

Amsterdam,
March 2010

GL11 Conference Sponsors

FLICC/FEDLINK, USA
Library of Congress

BL, United Kingdom
The British Library

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

DTIC, USA
The Defense Technical Information Center

EBSCO, USA
EBSCO Information Services

IIA, USA
Information International Associates, Inc.

GL11 Conference Sponsors

JST, Japan
Japan Science and Technology Agency

LCC, USA
The Law Library of Congress

NLM, USA
The National Library of Medicine

NYAM, USA
The New York Academy of Medicine

OSTI, USA
Office of Scientific and Technical Information

University of Ljubljana,
Slovenia

GL11 Program Committee

Roberta Shaffer and Robin Harvey ***(Co-Chair)***
FLICC/FEDLINK
USA

Elizabeth Newbold
BL, British Library
United Kingdom

Daniela Luzi
CNR, National Research Council
Italy

Anne Asserson
EuroCRIS, Current Research Information Systems
Norway

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

Christiane Stock
INIST-CNRS, Institut de l'Information Scientifique et Technique;
Centre National de Recherche Scientifique
France

Dr. Lawrence Marcus
LC, Library of Congress
USA

Janice Kaplan
NYAM, The New York Academy of Medicine
USA

Deborah E. Cutler
OSTI, Office of Scientific and Technical Information
USA

Julia Gelfand
UCI, University of California, Irvine
USA

Dr. Joachim Schöpfel
University of Lille 3
France

Table of Contents

Foreword 3
Conference Sponsors 4-5
Program Committee Members 6
Conference Program 8-9

Program

Opening Session 11
Session: Impact of Grey Literature on Net Citizens 19
Session: Grey Literature Repositories 47
Session: Open Access to Grey Resources 68
Parallel Presentations 92

Advertisements

FLICC/FEDLINK, Federal Library and Information Center Committee 10
NYAM, The New York Academy of Medicine 18
IIA, Information International Associates, Inc. 27
ERIC, Education Resources Information Center 64
Refdoc.fr, CNRS-INIST 73
J-STAGE, Japan Science and Technology Agency 110
NTK, National Technical Library, Prague 135
EBSCO Information Services 136

Appendices

Author Information 130-132
List of Participating Organizations 133-134
Index to Authors 137
GL11 Publication Order Form 138

OPENING SESSION

Chair, *Roberta Shaffer, Federal Library and Information Center Committee, FLICC/FEDLINK, USA*

Keynote Address	**Peter R. Young, Chief Asian Division, Library of Congress, USA**	**11**
Opening Paper	**The Grey System for Monitoring Self-Funded Research** *Aleksandr V. Starovoitov and Yuri M. Bogdanov, CITIS;* *Aleksandr M. Bastrykin, and Leonid P. Pavlov, VNTIC, Russia*	**13**

IMPACT OF GREY LITERATURE ON NET CITIZENS

Chair, *Michelle Springer, Office of Strategic Initiatives; Library of Congress, United States*

Raising Awareness of Grey Literature in an Academic Community Using the Cognitive Behavioral Theory **19**
Yongtao Lin, Tom Baker Cancer Centre and Marcus Vaska, Health Sciences Library, University of Calgary, Canada

The Integration of Grey Literature and Primary Research Data in Open Source Analysis: The Jordan Property Regime **28**
June Crowe and Bonnie C. Carroll, Information International Associates, Inc., United States

The secrecy of grey unveiled: Grey literature and the freedom of information **37**
Cees de Blaaij, Library of Zeeland, The Netherlands

GREY LITERATURE REPOSITORIES

Chair, *Dr. Lawrence Marcus, Library of Congress, LC, United States*

Documenting an Environmental Disaster: The River Valley Collection at Marion Public Library **47**
Ian Fairclough, George Mason University, United States

Mosaic: Shades of Grey **54**
Keith G. Jeffery, Science and Technology Facilities Council, United Kingdom
Anne Asserson, University of Bergen, Norway

Are legal texts Grey Literature? Towards an understanding of grey literature that invites the preservation of authentic and genuine originals **59**
Michael Lines, University of Victoria, Canada

A Multi-Institutional Approach to Technical Report Literature: Development of the Technical Report Archive and Image Library (TRAIL) **65**
Maliaca Oxnam, University of Arizona Libraries, United States

OPEN ACCESS TO GREY RESOURCES

Chair, *Dianne McCutcheon, National Library of Medicine, NLM, United States*

Digitizing Grey Matter from the Antarctic Bibliography Collection **68**
Tina Gheen and Sue Olmsted, National Science Foundation Library, NSF, United States

Usage of grey literature in open archives: state of the art and empirical results **74**
Joachim Schöpfel and University of Lille 3;Hélène Prost, INIST-CNRS, France

From CNR Annual report to an Institutional repository: Which successful strategies? **82**
Rosa Di Cesare, Daniela Luzi, Roberta Ruggieri and Loredana Cerbara, National Research Council, Institute of Research on Population and Social Policies, Italy

PARALLEL SESSIONS

OpenAIRE Project - Open Access Infrastructure for Research in Europe 92
Stefania Biagioni, Donatella Castelli, Paolo Manghi;
Institute of Information Science and Technologies, ISTI-CNR, Italy

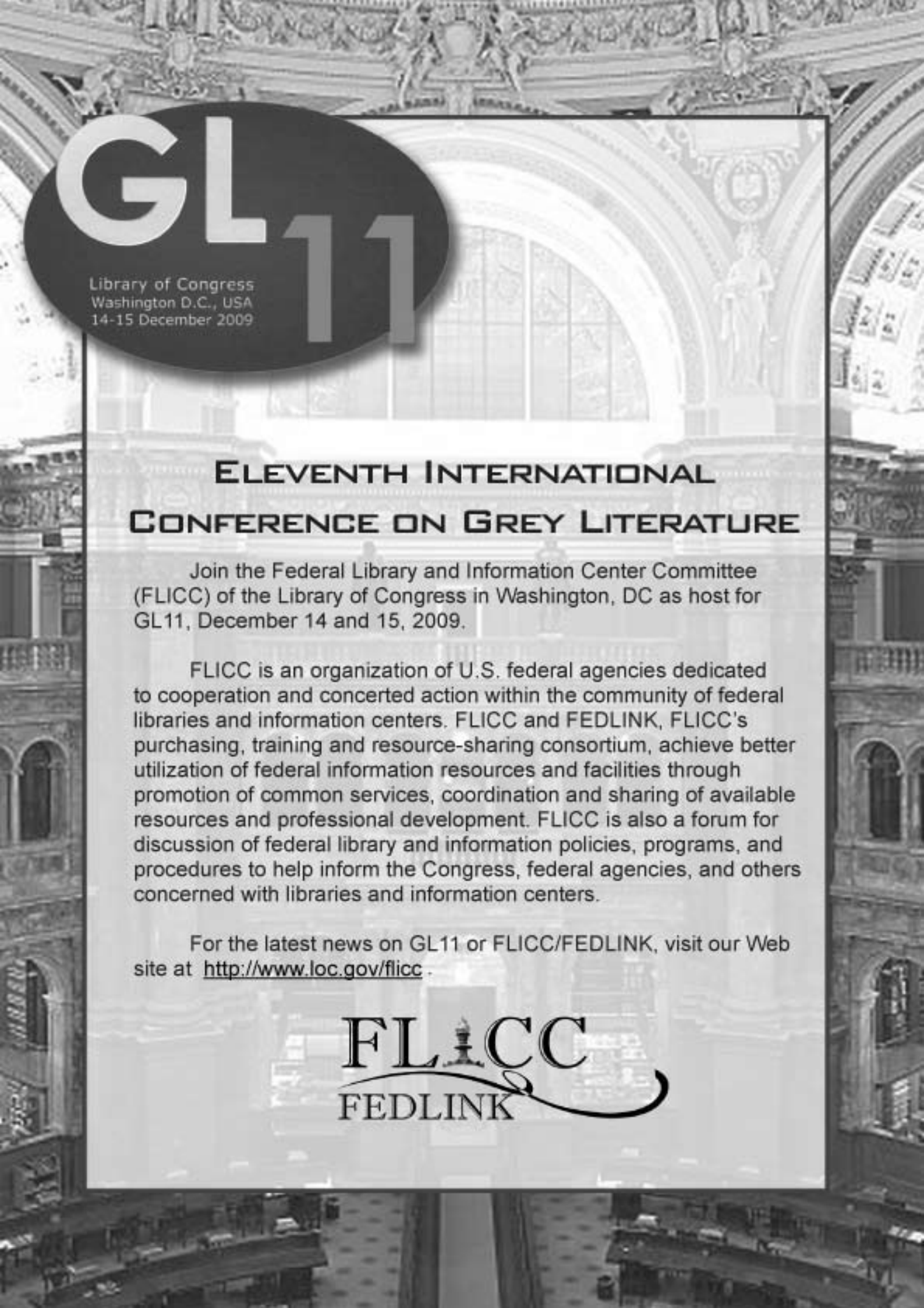
Open Access to Grey Literature on e-Infrastructures: the BELIEF Project Digital Library 94
Stefania Biagioni, Donatella Castelli, Franco Zoppi;
Institute of Information Science and Technologies, ISTI-CNR, Italy

A Hidden Treasure on Computer Science Pre-History in Pisa: the CSCE Collection 104
Stefania Biagioni and Silvia Giannini;
Institute of Information Science and Technologies, ISTI, Italy

Dissemination of Information on the Results of R&D through the JAEA Library 111
Kiyoshi Ikeda, Mayuki Gonda, Takashi Nozawa, Keizo Itabashi, JAEA, Japan Atomic Energy Agency
Naomi Ebisawa, TOSS, Total Support Systems Corporation, Japan

Grey Literature And Computational Linguistics: From Paper To Net 118
Claudia Marzi, Gabriella Pardelli, Manuela Sassi
Istituto di Linguistica Computazionale; Consiglio Nazionale delle Ricerche, CNR, Italy

OpenSIGLE - Crossroads for Libraries, Research and Educational Institutions in the field of Grey Literature 122
Dominic Farace and Jerry Frantzen, GreyNet, Grey Literature Network Service, Netherlands
Christiane Stock and Nathalie Henrot, INIST-CNRS, France;
Joachim Schöpfel, University of Lille, France;



Eleventh International Conference on Grey Literature

Join the Federal Library and Information Center Committee (FLICC) of the Library of Congress in Washington, DC as host for GL11, December 14 and 15, 2009.

FLICC is an organization of U.S. federal agencies dedicated to cooperation and concerted action within the community of federal libraries and information centers. FLICC and FEDLINK, FLICC's purchasing, training and resource-sharing consortium, achieve better utilization of federal information resources and facilities through promotion of common services, coordination and sharing of available resources and professional development. FLICC is also a forum for discussion of federal library and information policies, programs, and procedures to help inform the Congress, federal agencies, and others concerned with libraries and information centers.

For the latest news on GL11 or FLICC/FEDLINK, visit our Web site at http://www.loc.gov/flicc .

FLICC
FEDLINK

Keynote Address

Peter R. Young
Chief Asian Division, Library of Congress

Over much of the past decade, I have worked to understanding the impact of digital networking and Web 2.0 technologies on national and research libraries. This concentration was the focus of my work at the U.S. National Agricultural Library and guides my efforts at the Library of Congress. It is instructive to consider grey literature from the perspective of digital science data and eScience, and to consider the Eleventh International Grey Literature Conference themes[1] in relation to digital age challenges.

1. Grey Literature Challenges

GL consists of content "…produced on all levels of government, academics, business or industry in print and electronic formats, but is *not controlled by commercial publishers*."[2] GL consists of a "…body of materials that *cannot be found easily* through conventional channels such as publishers, but which is frequently original and usually recent." [3] GL in general consists of "…foreign or domestic open source material that usually is *available through specialized channels* and may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents."[4]

With the following characteristics, it is instructive to consider GL challenges:

- GL falls between open and classified/sensitive literature
- GL varies in quality
- GL intended to serve a limited audience
- GL issued in limited quantities
- Most GL is non-proprietary
- Hard-copy GL repositories are scarce & incomplete
- GL lacks outreach/marketing
- GL falls between "black & white"
- "Grey" implies incompleteness
- GL reflects gradations
- GL definitions are fluid

GL challenges form into the following questions:

1. What is needed to make GL more accessible and valuable?
2. Do we need to depend on information professionals to place this large body of valuable content under bibliographic control to enable search engines to provide better access?
3. Are GL content and metadata standards sufficient and universally applied so that GL could be accessible if sufficient resources were available?
4. What would it take for the corpus of current and retrospective GL to be accessible via the Web?
5. How would we decide what GL was going to be more valuable to future generations of scientists and researchers, especially if we do not have sufficient resources to assure access to all GL content?

2. Digital Grey Literature

Digital networking technology is affecting the nature of scientific and research communications. These same forces of change are affecting the nature of grey literature content. In addition, digital networking technology and the trend toward open access are changing the conduct of science and research through the rapid development of the Internet/Web. These same forces are having a profound impact on the distribution of and access to GL content.

> "What has been made public by being published is no longer a black and white issue. There is still plenty of room for judgments and distinctions to be made about the quality, type, and nature of this knowledge. *This growing openness around what is known assists in the very assessment and verification. I, for one, do not see grey skies ahead, but something brighter."*[5]
>
> "Since there has been an increase in publication and dissemination of materials from the producers of grey literature and other materials, establishing the means to work directly with users rather than relying on the traditional means of evaluating and collecting becomes necessary." [6]

It is difficult to see how the library and information community could successfully attract sufficient resources to address the growing body of GL content and to comprehensively work to provide consistent and uniform access to an ever-increasing tsunami of print and digital content. In fact, digital machines

are ideally suited for performing certain functions exceptionally well, but these functions are not identical to the things that humans are capable of replicating at anything near machine speed or accuracy.

The combination of demand-driven science, together with the development of new digital storage and Web exchange tools are creating a new dynamic model for conducting data-intensive research in new ways. These trends are creating a lot of excitement among an array of scientific domains from micro-biology to astrophysics, and from ecological to oceanographic studies, where the creation of new knowledge employs new digital tools and Web methods. These new dynamics require that the more traditional library and information science communities join in creating the emerging Data Web through which links and object relationships between and among content and users are established in a evolutionary and iterative way.

It is not difficult to see that the Web is escalating the GL challenge. A deluge of Web-based grey digital content is sufficient to drown any plan for placing these materials into a controlled and accessible environment through our traditional print-based bibliographic control methods. David Weinberger speaks on the "Future of Knowledge" in his *Everything is Miscellaneous: the Power of the New Digital Disorder (2007)*. His thoughts on the relationship between data, information, knowledge, and wisdom are both enlightening and challenging. From an epistemological standpoint David's ideas are interesting. But, from a librarian's standpoint, these concepts are revolutionary. He writes of a third order of knowledge (a new paradigm) as one where things that one can do easily with digital technology are those things that the real world makes really hard. In this third order, users (not owners) of information organize that information according to their interests.

Flickr offers an example of Web 2.0 tools that offer opportunities for individuals and communities to help organize Web content. The Library of Congress's project with Flickr to allow users to help identify assets and in so doing to assist curators in describing and tagging individual photographs with descriptive terms is but one example of engagement of users in helping to organize digital content. Del.ici.ous is another example, that offers tagging in the form of bookmarks to Web sits of interest. It offers personalization in tagging, but it also provides a social networking community where a user can see what everyone else using the public site has tagged. These tags are social links that grow and evolve through the addition of more and more users, illustrating the "Wisdom of Crowds"that WikiPedia reflects.

Another example of citizen science is drawn from astronomy. The Galaxy Zoo site contains a quarter of a million galaxies which have been imaged with a camera attached to a robotic telescope the Sloan Digital Sky Survey. More than 150,000 people have taken part in Galaxy Zoo to help classify images of galaxies.

> "Today, scientific publishers are production companies, specializing in services like editorial, copyediting, and, in some cases, sales and marketing… in ten to twenty years, scientific publishers will be technology companies…. they'll be technology-driven companies in a similar way to, say, Google or Apple. That is, their foundation will be technological innovation, and most key decision-makers will be people with deep technological expertise. Those publishers that don't become technology driven will die off."
>
> "Scientific publishers should be terrified that some of the world's best scientists, people at or near their research peak, people whose time is at a premium, are spending hundreds of hours each year creating original research content for their blogs, content that in many cases would be difficult or impossible to publish in a conventional journal. What we're seeing here is a spectacular expansion in the range of the blog medium. By comparison, the journals are standing still."[7]

The global library and information services community does not need to be standing still. We have the responsibility to address the growing need to both organize and to provide access to grey content, especially as digital data increases in the future. The astonishing growth of digital content requires that we elicit new methods and new models for assuring that the value of grey literature content is available to those scientists and researchers addressing the grand challenges of the future.

---

[1] Impact of GL on Net Citizens, Corporate governance of GL, Uses and applications of subject based GL, GL repositories revisited and Open access to Grey resources
[2] *ICGL Luxembourg definition, 1997 - Expanded in New York, 2004*.
[3] *Wikipedia*
[4] US Interagency Gray Literature Working Group 1995
[5] Blogger John Willinsky. Professor at the Dept of Language and Literacy Education at UBC, and the Public Knowledge Project.
[6] Heather Lehman and Janet Webster *Describing Grey Literature Again: a survey of collection policies* in Publishing Research Quarlerly, Spring 2005, pp. 65-72.
[7] Michael Nielsen blog.

Opening Paper

The Grey System for Monitoring Self-Funded Research

Aleksandr V. Starovoitov and Yuri M. Bogdanov,
CITIS, Centre of Information Technologies and Systems of Executive State Authorities, Russia

Aleksandr M. Bastrykin, and Leonid P. Pavlov,
VNTIC, The Scientific and Technical Information Centre of Russia

Abstract
To eliminate the defects in research monitoring in Russia a special Government decision was issued in 2006 with the idea of creating a system for self-funded research projects registration. The system was designed in the years of 2007 – 2008 and now put into operation as an integral part of the federal scientific and technical information grey literature system. Based on the output information from the system the annual summary report for the Joint Interministerial Commission specified in the Government Decision is prepared. A general outline of the system, the description of its input and output document forms and database structure are given in the paper.

Introduction

Today's global economy is often described as one in transition to a "knowledge economy", or "knowledge-based economy" which is being developed in the environment of a "knowledge society" as an extension of an "information society". The theoretical grounds for this kind of treatment as well as the very terms "information society", "knowledge society" and "knowledge economy" were put into scientific circulation as long ago as in the late sixties last century by the famous American economist and the founder of the modern management theory Peter Drucker in his book "The Age of Discontinuity" that was republished many times since then [1].

Now that the Drucker's forecast came true the decisive role of human capital in the form of knowledge and education is universally recognized as a powerful productive asset and the source of innovative solutions. In Russia of today it is well understood that the way from the backward raw materials economy to the modern knowledge economy lies through innovations generated by scientific research and applied in industry. Some burdens of the Soviet science strategies still haunt the Russian science even now and about 80% of financial support for science still comes from the state [2] while Russian business is reluctant to invest in scientific research. Also, it was typical for the Soviet science to carry out research at the widest front possible disregarding the practical results if any and this approach can not be followed today. So, new concepts are required in the state scientific policy.

Administrative measures

It is evident that the immediate aim of forming a knowledge society and an innovation economy suggests an adequate funding of scientific research and development. There has been a growth in scientific research state budgeting every recent fiscal year (except this year of 2009 when because of the world economic crisis the state budget of Russia was sequestrated including the item of science expenditure). So, the federal budget science expenditure in percentage to the total budget expenditure grew from 1,69% in 2000 to 2,27% in 2006 [3]. In absolute figures the growth of federal budget science expenditure is as follows: 72,4 billion rubles in 2006, 89 billion rubles in 2007 and 118.4 billion rubles in 2008 [4].

The figures of budget assignations to science being important it is no less important to optimize the distribution of financial means among the fields of science. To do so priorities in scientific research and critical subjects should be determined. Since there are about one hundred budget recipients in Russia - federal ministries and agencies that have their share in the total budget science expenditure to finance their scientific research and development projects - Joint Interministerial Commission was formed as a decision-making body for research policy and with the aim of working out or updating the list of science and technology priority development directions and the list of critical technologies in the Russian Federation at least once in four years.

Now there are totally 8 priority directions, for example bio-systems, energy and energy saving, information and telecommunication systems, nano-systems and materials, and 34 critical technologies like bioinformation, software design, hydrogen energy, distributed computing, nuclear energy and fuel, cell technologies, atmo- and hydrosphere monitoring, refuse utilization,

etc. The lists prepared by the Commission on the basis of different means of science monitoring and statistics are approved by President of the Russian Federation (the latest approval dated May 21, 2006).

Collecting and monitoring state-funded R&D

One of those means is the federal scientific and technical information system for grey literature that supports monitoring (both in financial and subject respect) of the state funded scientific research and development activities covering extensively all the territory of the Russian Federation [5]. The system collects and controls scientific and technical reports and dissertations concerned basically all scientific subjects ranging from mathematics, physics, electronics and engineering through to social sciences and the humanities.

The legal ground for the operation of the system is the Federal Law "On the obligatory copy of documents" that makes it mandatory for all the scientific research executors (scientific and higher school institutions, industrial organizations) to be the input documents suppliers to the system in the form of full-text research and development reports accompanied by registration and information cards with the abstracts and metadata of the documents. Also, data on contracts and funding arrive from federal ministries and agencies and dissertations come from dissertation councils and individual researchers.

Since 2004 the full-text documents have been scanned and converted into a digital representation to form a full-text digital repository (before 2004 the full-text reports and dissertations were microfilmed) and since 1982 the registration and information cards with the abstracts and bibliographic descriptions of the documents have been keyed-in into computer to form the corresponding retrospective database serving an electronic catalogue to the repository and a means of quick and easy online search and access to the information. Each document has a unique inventory number both in the database and in the repository. Now there are about 3 million documents in the database.

The users of the system are able to conduct online subject search and order copies of documents both via the Internet and in the reading room. Most individual users are researchers, experts, lecturers and tutors, under- and post-graduate students. Among collective users are different kinds of organizations – scientific research institutes, universities, industrial and business community, Higher Dissertation Commission, state executive and controlling authorities (ministries, agencies, audit and fiscal bodies, law-courts). Official users usually require statistical, financial and generalizing information on science.

Thus, the existing system for grey scientific and technical information sources is the most complete and reliable means of monitoring and controlling the situation in state-funded scientific research and development activities. The system's collection is an indispensable source for government agencies with an interest in the latest Russian contributions to science and technology.

Self-funded research monitoring Decision

No matter how much money is given to science from the state budget it can never be the only and sufficient financial source for research and development. So, the diversification of funding is inevitable and there is a growing trend that more and more scientific projects are being funded from research organizations' own financial resources.

Those organizations are commercial state-owned ones functioning in the forms of federal state unitary enterprises and open joint-stock companies with the state share-holding. Their self-funded research projects are not covered by the federal grey literature system and hence were out of centralized monitoring and were not taken into account by the Joint Interministerial Commission when preparing proposals for updating the lists of priority directions and critical technologies.

To eliminate the defects in research monitoring a special Government Decision was issued on November 4, 2006 (No. 645) with the idea of creating a system for self-funded research projects registration. The greatest difficulty with such systems of all-Russian scale is not its computer network and software realization (though not a simple thing in itself) but its, so to say, organizational technology - the ensuring that all the commercial organizations involved provide their reports to the system timely, complete and correct even if they are supposed to do so in accordance with the Government Decision. Unfortunately, the sad experience of the latest decades exposed a very low executive discipline of scientific and scholar institutions. So, Government Decision No. 645 not only stipulated the idea but also defined the mechanisms of its realization.

It is well known that financial reports are presented by all the organizations most unquestioningly and without demur. The Decision obliged the top managers of the federal state unitary enterprises and open joint-stock companies with the state share-holding to include in their annual financial report presented to their higher ministry or agency the information on their self-funded research and development projects and to submit this information to the Federal Agency on Science and Innovations responsible for the scientific activities assessment and monitoring in the Russian Federation. The Decision also obliged the Federal Agency on Science and Innovations to generalize the received information and issue the Annual Summary Report on self-funded research intended for the Joint Interministerial Commission so that it could take into consideration the data from the Report when working out the lists of priority directions in science and critical technologies.

In accordance with the Decision the information on self-funded research should be submitted in an approved unified form as an annex enclosed in the organization's annual financial report. The approved blank form is added to the Decision's text. The form's fields of data are important because their filling determines the information value of the document. Of course, the form is in the Russian language and we will consider its translation into English.

The heading of the form is as following: "Civil purpose self-funded scientific R&D projects of the federal state unitary enterprise (or the open joint-stock company with the state share-holding)" – the specific organization's name and report year are to be filled-in. Next come the following names of fields to be filled in:

- city of the organization
- postal address and phone number
- higher federal authority body (ministry, agency, etc.)
- branch of activity
- main kind of activity

After those attributes defining the organization the fields follow that characterize the content of research, its time limits and self-funding:

- subject(s) of scientific R&D (project-title)
- priority direction, critical technology
- term of carrying out (years)
- volume of financing for the fiscal year (in thousand rubles)
- basic results (in increasing way, for the fiscal year included).

The content fields are filled-in in a free way. The filled-in form is signed by the head of organization, stamped and directed to the Federal Agency on Science and Innovations. The form is to be submitted both in printed (on paper) and in digital (Microsoft Word) shape.

The System

To carry into effect the Government Decision a system based on computer network and database technologies was designed in the years of 2007 – 2008 and now put into operation as an integral part of the federal scientific and technical information grey literature system. The self-funded research monitoring system is evidently grey because its input form described above and output Annual Summary Report are typically grey documents.

The state contract for the system's design and introduction was won by the Centre of Information Technologies and Systems of Executive State Authorities (CITIS) that is now merged with the Scientific and Technical Information Centre of Russia (VNTIC) being the leading centre for grey literature in Russia. The system's architecture and software was developed by the specialists of CITIS and VNTIC.

The heart of the system is the database on self-funded research with the filled-in annex forms. It was not possible to completely avoid the paper forms since the annex was enclosed in the annual financial report submitted on paper. Nevertheless, an online mode of filling the annex form is provided for the authors of self-funded research projects who are able to address the CITIS site on the Internet (www.citis.ru), click "the annex form online filling-in" and have the form on the screen of their computer. There are many conveniences supporting the online filling-in such as the formal verification of numerical fields, the enclosed lists of priority directions and critical technologies and the list of correct names of the organizations that were previously registered in the system. The user just has to click the name instead of keying it in.

On completing the online filling-in of the annex form the user directs the document via the Internet straight to CITIS where it is entered into an intermediate (technological) database. At the same time the user prints out the completed form, this paper form (signed by head of the organization and stamped) is added to the annual financial report and submitted to the Federal Agency on Science and Innovations wherefrom it arrives in CITIS.

The documents in the technological database and the corresponding documents arrived on paper are verified, after verification are entered into the database on self-funded research and each assigned a unique registration number. This main database of the system is provided with a well-developed search and retrieval means and all the fields of the document are searchable. The database management system or rather information search and retrieval system is IRBIS designed by a group of Russian scientists and programmers under the supervision of Professor Nicolai Maximov [6] and now recognized as a de-facto standard DBMS in the sphere of scientific and technical information in Russia.

Now there is a two-year retrospective database (with the report documents of 2007 and 2008 years – totally about 1 thousand documents), next year (2010) the documents of 2009 will be entered and so on. Thus, the system ensures the registration of report documents on self-funded research, their permanent storage in the database format and both quantitative and qualitative analysis of self-funded research in Russia prepared in the form of the Annual Summary Report.

Since the field "Basic results" in the annex form often is filled-in by the report authors in rather a short way it is not informative at times. Sometimes the same is true for the field "Subjects of scientific R&D" when the title of a project is rather general and not specific. In these cases it is difficult to analyze the research results qualitatively. In order to enhance the system's potentialities a link between its database and the Federal registration and information cards database (the state-funded research database) was envisaged. The documents of the latter database – the information cards – are more detailed than the annex form and contain the abstracts of full-text reports.

To make the link work it is necessary to persuade the authors of self-funded research projects that they should also register their projects in the Federal grey literature system and fill-in the information card so that it could be entered into the Federal database. In case of the lack of information in the annex form for a detailed analysis it would be possible to address the corresponding information card. A circular letter with such a recommendation was directed to all the federal state unitary enterprises and open joint-stock companies with the state share-holding. Unfortunately, only three organizations followed the recommendation and registered their 22 projects in both the databases.

Annual Summary Report

The main output document of the system is the Annual Summary Report on self-funded research intended for the Joint Interministerial Commission. The Report consists of the textual analysis and the illustrative materials in the form of tables and pie charts. So far two Reports were prepared by means of the system – in 2008 (documents of 2007) and in 2009 (documents of 2008).

The major items of the Report contents are as follows:

- table of government bodies showing the amount of their dependent enterprises and companies submitted the self-funded R&D information
- distribution of organizations' financial means for self-funded R&D by the federal government bodies
- self-funded R&D financing distribution by priority directions and critical technologies
- the enterprises and companies spent more than 50 million rubles on self-funded R&D
- comparative tables of data for the reported and previous year
- generalized data and its analysis
- conclusions and recommendations.

In 2007 and 2008 the total average amount of financial means spent by enterprises and companies on self-funded research per year is more than 1 billion rubles. In 2008 the amount of financial means spent on self-funded research concerned the priority directions and critical technologies increased nearly three times as compared to the year of 2007. In 2008 totally 56 organizations (39 enterprises and 17 companies) funded 455 research and development projects (the enterprises – 297 projects and the companies – 158 projects) from their own

financial resources. Three enterprises spent more than 50 million rubles on self-funded research.

As an example let us consider a pie chart reflecting the self-funded research distribution by priority directions of scientific and technological development (in percentage of spent money):

- rational wildlife management – 41.14 %
- energy and energy saving – 24,92 %
- live systems – 19,20 %
- nanosystems and materials – 10,28 %
- transport, aircraft and space systems - 2.10 %
- security and antiterrorism - 1,46 %
- information and telecommunication systems - 0,90 %.

Concluding Remarks

Now that the system is put into operation it may be concluded that self-funded research is no longer out of monitoring and control. The information obtained by means of the system allows:

- to monitor the situation in the sphere of self-funded R&D all over Russia
- to update the lists of priority directions and critical technologies
- to improve the distribution of financial means for scientific R&D both federal budgeted and self-funded
- to reduce the unjustified duplication and overlapping of R&D projects.

References

1. Drucker, Peter F. The Age of Discontinuity: Guidelines to Our Changing Society. – Rutgers Univ., N.J.: Transaction Publishers, 1992. – 420 pp. - ISBN-10: 1560006188. ISBN-13: 978-1560006183.

2. Kuznetsov, Yuri. Federal budget funding of civil science in Russia. / Otechestvennye zapiski. – No.7, 2002 (In Russian).

3. Krichevsky, Nikita. Crisis, gentlemen! http://www.apn.ru/publications/article20966.htm (In Russian). Visited 22.09.2009.

4. Ministry of Finance of the Russian Federation official site http://www1.minfin.ru/ru/budget (In Russian). Visited 22.09.2009.

5. Pavlov, L.P. The State and Development of the Russian Grey Literature Collection and Dissemination Centre. – Interlending and Document Supply. - MCB Univ. Press. 1998, vol. 26, No. 4. Pp. 168-170.

6. Maximov, N.V. et al. Information Systems. – Moscow. – FORUM: INFRA-M, 2007. – 496 pp. (in Russian).

Raising Awareness of Grey Literature in an Academic Community Using the Cognitive Behavioral Theory

Yongtao Lin, Tom Baker Cancer Centre, and
Marcus Vaska, Health Sciences Library, University of Calgary, Canada

Abstract

Cognitive skill training, part of cognitive behavior management, is based on the cognitive behavioral theory. The principle that thinking controls behavior has been widely used by educators to develop methods to improve the performance of students; this concept can be further extended to clinicians who provide multifaceted health promotion and psychological counseling programs. The plan of our educational project is to raise awareness among students, faculty members, and researchers of material not produced via standard commercial publishing channels, emphasizing the role that grey literature plays in teaching and research in our academic community.

Pre-test surveys conducted prior to these ventures indicate that while researchers, students, and faculty members may have used grey literature resources at some point during their research pursuits, more than one-third (36.7%) believe they haven't done so, a significant number expressing uncertainty in not knowing how to find, effectively use, and evaluate grey literature. In terms of specific grey literature resources, the majority (85.7%) of users have had rich experiences with association and government websites, but lack familiarity with other useful resources such as subject-based directories, databases, or well-established grey literature repositories.

"When opportunity knocks, you should probably open the door!"(Schwann, Petermann, & Petz, 2008). This statement indicates the value and importance of theories in promoting new practices in health services. A theory can lay the foundation of a teaching or learning goal by describing the purpose, intervention, and assumed outcome of a proposed endeavor (Wayne State University). As health sciences librarians, the need to promote grey literature in terms of visibility and accessibility, thus raising awareness, forms the basis of our teaching goals and is thereby the focus of this paper.

Undoubtedly, as the barrier between white and grey literature becomes narrower with the advent of technology that seeks to uncover the unrecoverable, challenges will inevitably arise, especially when deciding what exactly about grey literature needs to be promoted. The cognitive-behavioral theory helps describe various factors regarding lack of awareness and misguided conceptions about searching for elusive material, while also guiding the selection of sources of grey literature and the methodologies we have adopted in our promotion project.

Three such educational initiatives that we have developed, based on the principle that change will occur only as you think differently, will be highlighted in this paper. These include the creation of a Grey Literature Speaker Series (http://glspeakers.wetpaint.com)[1], where subject specialists/liaison librarians at the University of Calgary discuss their experiences and challenges with uncovering and using grey literature in their subject areas; a joint presentation to faculty members and students at the University of Calgary Faculty Technology Days showcasing the impact and role that technology, such as mobile computing, has played and continues to play when it comes to the grey literature; and library information sessions to cancer researchers based on user experiences, expectations, and needs when it comes to seeking non-traditional material.

The One-Minute Paper, distributed after a session to assess participant learning outcomes, reveals that knowledge of different sources and current trends in grey literature has significantly increased. We therefore advocate the use of the cognitive behavioral theory as a channel for developing education programs and promoting grey literature in our academic community.

***Keywords**: cognitive behavioral theory, grey literature, raising awareness, teaching, academic community/setting, assessment*

THE COGNITIVE BEHAVIORAL THERORY (CBT)

"Knowledge is of no value unless you put it into practice"
(Anton Chekhov)

Having established its roots in Greek philosophy, Cognitive-Behavioral Theory (CBT), or cognitive-behavioral therapy (literature has shown that these terms are used interchangeably), is "a form of psychotherapy that emphasizes the important role of thinking in how we feel and what we do" (NACBT, 2007). CBT was introduced as a methodology in the health sciences sector in order to provide an opportunity for medical students and healthcare practitioners to apply evidence-based medicine to their research pursuits (Lloyd & Reyna, 2001).

The essence of cognitive behavioral theories clearly defines "what we know and think affects how we act" (Wayne State University, p.9), implying that one's behavior can be modified by one's knowledge. Another critical point of note posits that "knowledge is necessary but not sufficient to produce behavior change" (p.9). Therefore, our perceptions of using cognitive behavior theories to interpret how our users, namely health care providers and researchers, search and accept information, has determined our methodologies in raising awareness of grey literature (GL) in order to bring behavior changes to research in an academic environment.

Didactic reasoning, or the common sense approach, remains in the forefront of the library teaching sessions we deliver to our clients and our colleagues. When introducing the concept of GL, we ask the user to apply previously untapped cognition and decision-making skills to find, access, and retrieve relevant information. In the world of inquiry-based learning, we raise awareness and provide advice and guidance on how to recover the unrecoverable. While we infinitely emphasize the importance of considering GL resources in the information-seeking process, it is ultimately the researchers' thinking process that determines how, when, where, and why this literature will be used.

HEALTH BELIEF MODEL (HBM): COGNITIVE SKILL TRAINING IN HEALTH PROMOTION AND PRACTICE

There are several models that exist for discussing health behavior. These are exemplified at the individual and intrapersonal level and are often broadly categorized as cognitive-behavioral theories. In correlation, social and behavioral scientists have spent years creating numerous theories to understand and explain behavior (Tsang, 1999). Wayne State University proposes knowledge, attitude, beliefs, motivation, self-concept, developmental history, past experience, skills, and behavior as the key components of intrapersonal behavior (p.9). These aspects of behavior lead to the notion of cognitive skill training in health promotion and practice, namely the Health Belief Model (HBM).

One of the first models created to adapt theory from the behavioral sciences to health problems, HBM focuses on one's perceptions of a health problem threat, along with a recommended behavior for preventing or managing the problem. While its original definition was aimed at understanding one's idea of a threatening health problem, coupled with a countermeasure to help offset and/or manage the issue, the literature supports the notion that HBM can assist in scrutinizing one's inaction or noncompliance (Wayne State University). Developed and introduced in the 1950s by a team of psychologists working in the U.S. Public Health Service, HBM speculates that individuals fear disease because they fear the unknown. When first introducing the concept of GL to an audience, whether that audience is composed of healthcare practitioners, librarians, or the academic community in general, it is the fear of a lack of awareness about GL and its uses in research that can often lead to skepticism or quizzical looks.

Primarily focused on one's attitudes and beliefs (Tsang, 1999), HBM is comprised of six concepts that explain and account for an individual's decision to accept change and modify one's behavior accordingly: perceived susceptibility, perceived severity, perceived benefits, perceived barriers, cues to action, and self-efficacy (Wayne State University)(Table 1). While primarily adapted to fit the needs of those involved in the healthcare sector, we have modified and adapted the components of HBM towards our own GL educational pursuits. Before these initiatives are introduced, it is useful to briefly comment upon the original perceptions of HBM, using the current H1N1 epidemic as a case study.

(Table 1: Health Belief Model)

Concept	Definition
Perceived Susceptibility	Perception of chances of getting a condition (i.e. H1N1).
Perceived Severity	Perception of the severity of condition and its impact (modality and mobility of H1N1)
Perceived Benefits	Perception of effectiveness of the adverse action (benefit of existing measures such as H1N1 vaccine)
Perceived Barriers	Perception of tangible, psychological and other costs of adverse action (possible side effects of the vaccine or not getting vaccinated)
Cues to Action	Strategies to activate "readiness" (to convince individuals to receive the H1N1 vaccine)
Self-Efficacy	Confidence in one's ability to successfully perform action (individuals receive H1N1 vaccine)

Often referred to as perceived threat, perceived susceptibility and perceived severity discuss one's perception of contracting an illness or disease. In the realm of GL, an assessment study must be taken of focusing only on the traditional (i.e. scholarly, peer-reviewed) literature, as well as reserving judgment towards the negative impact of excluding GL in research. The perceived benefits comment on the various existing measures to combat or reduce the spread of the illness, often referred to as the effectiveness of the adverse action, while perceived barriers emphasize possible adverse effects of the action or inaction (Tsang, 1999). This translates into recognizing the usefulness of GL and its impact on research, while attempting to modify researcher opinion as to the barriers of using GL resources. In order to encourage an individual to take action on a particular perception, thereby activating 'readiness', cues to action are used to inform and entice. Finally, self-efficacy, seen as one's confidence in the ability to successfully perform an action can be interpreted as a personal belief in one's ability to use GL effectively and efficiently on a regular basis.

(Table 2: Application of HBM)

Concept	Application in GL Awareness
Perceived Susceptibility	Researchers believe that not including GL in research would miss relevant evidence.
Perceived Severity **Perceived Benefits**	Researchers believe that the consequences of excluding GL are significant enough to try to avoid.
Perceived Benefits	Researchers believe that inclusion of GL would present a full picture.
Perceived Barriers	Researchers identify their barriers to using GL and explore ways to reduce barriers.
Cues to Action	Researchers receive messages about the GL Speaker Series, training sessions, OA Poster, etc.
Self-Efficacy	Researchers are confident in using GL.

A model that delves into the perceived beliefs of the individual, HBM can indeed be applied across the board to promoting GL material. Its primary aim is to propose that "the likelihood of an individual making a particular action is determined by the perceived benefits of the action weighed against the perceived barriers" (Tsang, 1999, p.24). We begin GL awareness programs by investigating perceived susceptibility which defines the targeted user group, along with identifying the risks of focusing only on mainstream literature. Specifying consequences of lack of awareness with GL pertains to the HBM concept of perceived severity, where researchers believe that the drawbacks of overlooking grey material are significant enough to be avoided. This methodology serves as a prequel to the perceived benefits of HBM's role in cognitive behavioral theories, where defining actions to take, namely how, where, and when, help clarify

the positive effects of GL in both research and teaching. Based on studies in the healthcare sector which describe the potential fallout of failing to consider non-peer-reviewed studies in meta-analysis, it becomes evident that the inclusion of GL truly does present the researcher with a full picture, the complete spectrum within which to gather and retrieve information, serving as a complement to peer-reviewed journals.

There are numerous hurdles to overcome in the quest for the ideal document; not everything is available electronically, despite the wealth of information the Internet is able to provide. Skepticism on the part of the learner can create a perceived barrier towards searching, an issue that needs to be identified and reduced via reassurance and assistance interventions.

Deemed particularly useful in analyzing a person's inaction or incompliance, HBM remains one of the most widely recognized conceptual frameworks of health behavior. In our GL programs, which will be discussed in the next section of this paper, we will demonstrate how we have adapted HBM to explain why the desire to raise awareness of GL has been so eminent. People are fearful of the unknown; GL sessions motivate one to learn about the unknown (perceived threat), reducing fear by taking action ascertaining that benefits outweigh risks. The reason for using and relying on HBM is the fact that it not only explains what it takes to change behavior, but guides the search for why and how, persuading an audience to make decisions with the assistance of grey material.

RAISING AWARENESS OF GREY LITERATURE IN AN ACADEMIC WORLD

The goals of our study are twofold: firstly, to use HBM, one of the cognitive behavior theories, to describe factors regarding lack of GL awareness, coupled with the guiding methodologies we have adopted, and secondly, to determine and establish effective programs in an academic community to promote GL awareness.

In order to make consistent use of a new research tool, one must be equipped with knowledge of the tool and the importance of using the tool. By providing how-to information, sending email reminders about GL programs, and showcasing our GL poster, we are motivated to provide training and guidance in GL. All of these attempts, while certainly time-consuming, are proven factors that training and guidance in GL inevitably increases researcher confidence and subsequently skills in using GL.

A User-Education Program for Cancer Researchers

Held semi-annually, GL classes, part of a User Education program at the Tom Baker Cancer Knowledge Centre, help expand boundaries and create knowledge, awareness, and tolerance towards grey material.

In order to enhance meaning with the concepts introduced during the library user education program, each session we deliver revolves around a current topic and/or patient safety concern. In this way, we are able to not only provide our user with a definition, background and brief history of the world of grey, but also provide concrete examples of the impact of GL on a specific health perspective/topic.

The GL Speaker Series for University Librarians

Following a brief presentation given at the University of Calgary's Libraries and Cultural Resources (LCR) Academic Council meeting in January 2009, where we described our experiences at GL10, we held a lecture at the Faculty of Communication and Culture's Colloquium Speaker Series, introducing the concept of GL and addressing critical roles GL plays in research. During the course of this presentation, it became abundantly clear that "grey literature is not a term used by most faculty members" (Sulouff, 2005, p.1). Having read Sulouff et al's qualitative study, where subject librarians were interviewed in order to determine the level at which various disciplines make use of GL, we were interested in seeking out the methods and resources used by librarians across the University campus to retrieve GL. Upon sending out a call for volunteers via an e-mail listserv, and after filtering the responses received to ensure that there was adequate representation from each of the primary faculties on campus, a team of nine librarians formed the GL Speaker Series. The purpose of this team was to provide an opportunity for each liaison librarian to educate colleagues within LCR on how the term 'grey literature' can be applied and defined in the context of everyday research-related tasks in a particular discipline.

We also established a GL Speaker Series Wiki, http://glspeakers.wetpaint.com, as a forum to provide awareness of the nature of the untraditional material to a wider LCR audience. An introductory session was planned to introduce GL to our colleagues and invite participation for forthcoming GL Speaker Series over the next year. Working with our fellow librarians, we have come to realize once again that "while most librarians are familiar with the term grey literature…there was much that the subject librarians did not know about grey literature, especially in disciplines other than their own." (Sulouff et al., 2005, p.2). Nevertheless, as many University of Calgary librarians stated during the inaugural GL Speaker Series session, the

opportunity to learn how each of the team members made use of GL in their work strengthened the notion that GL was everywhere, across disciplines, not solely residing in the medical or science field.

The goal of the GL Speaker Series is to create awareness of GL in various disciplines within library sciences, and more importantly, to encourage communication and collaboration among librarians about using and promoting GL. The purpose of these sessions has been and will continue to be focused on a specific theme/topic approach to GL. The core subjects will consist of social sciences, arts, humanities, health sciences, education, business, and law. Based on discussions raised during the introductory session, we have observed how the use of GL evidently transcends across disciplines, changing one's research behavior by actively pursuing this type of material on a regular basis.

Showcasing GL at Faculty Technology Days

While the influx of institutional repositories and GL networks have attempted to gather as much research and teaching material as possible into their information vaults, "there is no central register for web sites distributing electronic publications" (Stempfhuber et al., 2008, p. 123). Internet search engines such as Google, particularly Google Scholar, are gaining widespread acceptance as a means of accessing and retrieving GL material. Nevertheless, despite the impact that technology has played when it comes to uncovering the 'hidden treasures' of information, "grey literature and open access publications are still seen as second-class publications…despite the fact that many of these publications went through the same type of peer review as traditional publications"(p.124). In an attempt to prove that GL extends far beyond one particular discipline, we set out to inform the University academic community as a whole, via a presentation during the University's annual Faculty Technology Days, about the impact that technology, such as mobile computing, plays in one's information-seeking behavior. Catering the session to faculty, staff, and graduate students, the hands-on interactive session discussed the role that mobile devices have played in specific types of GL, in particular social networking sites.

GL Poster During Open Access (OA) Week

The University of Calgary celebrated the international momentum of OA when LCR hosted several events and activities during OA Week (October 19-23, 2009). A poster entitled *Open Access & Shades of Grey* was displayed at the Student Centre during the OA Week keynote presentation on *Scholarship and Learning in an Open Access Knowledge Environment*. The poster showcased the role that OA publishing plays in the realm of GL. OA increases access, awareness, and visibility of previously hidden material, thereby acting as an essential complement to peer-reviewed findings. It was also a wonderful opportunity to broaden awareness and understanding of GL and OA to research and teaching as well as to raise questions for future discussions on how to facilitate better and more responsive access to digitally-created content.

METHODS AND FINDINGS

Pretest Surveys (Appendix 1)

Brief pre-test surveys were distributed via email to attendees approximately one week prior to all training sessions. Although participation in and completion of the survey was entirely voluntary, the responses received helped identify the different learning outcomes of each learning group, develop session content, and structure the classes accordingly. When asked about their previous experience with GL, 61% of respondents indicated that they either have not used or searched for GL in their field, while in fact almost all had been exposed to GL; they were merely unfamiliar with the definition and basic concepts of GL. Among all types of GL resources, researchers mostly consulted association and government websites when seeking out the latest trends or news releases. This was additionally followed by familiarity with search engines, and print materials as other alternate sources within which grey material could potentially be found.

The pre-test surveys also revealed that a majority of attendees (63%) used GL for their own research pursuits, while 30% focused on GL as a self-learning or teaching aid. The overlap in the percentage numbers indicates the multi-disciplinary nature of GL, where self-learning, teaching, and research can go hand-in-hand.

One-Minute Paper

Participants were given the opportunity at the end of GL sessions to anonymously provide feedback via the One-Minute Paper, on "two or three things learned from the session that may be most useful to future research and practice." This aspect of the study design was extremely beneficial, as it allowed for open and honest communication. We were able to evaluate the effectiveness of our program and ultimately determine whether behavior change with regards to using GL would be more likely to be carried out since the users perceived the consequences of

excluding GL in research. This supported our views that benefits can indeed be derived from including grey material particularly if there are fewer barriers to using and searching GL.

Responses were analyzed and three major common themes were identified (see Table 3 below):

(Table 3: Examples of One-Minute Paper Responses)

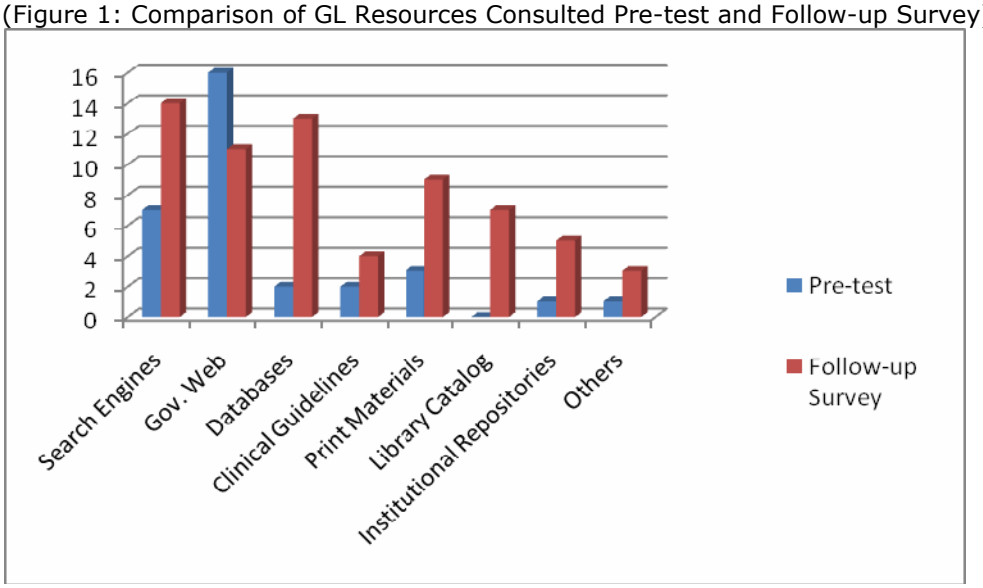
Common Themes	Examples
Concepts, Theories, and Introduction	• promoting GL to researchers is key in giving balanced/unbiased information • clarity about what GL is and isn't • breadth of search engines, databases, open access resources, etc. to search GL
Access, Resources, and Strategies	• lists of GL resources/websites is very practical and useful • strategies to search GL
Information Literacy	• this was my first exposure to the term GL. It's helpful to bring other perspectives to the realm of open or unofficial resources that are available. It will help in my own activities around making grey resources more readily apparent, findable, and accessible.

The majority of responses received were highly appreciative of being exposed to GL, and supported the inclusion of documents of a grey nature in research tasks. This proved that, with the guidance of HBM, it is possible to change one's behavior and perception.

Follow-up Survey (Appendix 2)

A follow-up survey was emailed to participants six months after the sessions with an aim of evaluating the impact of our GL awareness programs on our clients' research and teaching practices, as a reflection on whether there are behavior changes in searching and using GL. The comments received were positive. Almost all indicated that they have searched for GL since attending a GL session. Many have become more confident in recognizing the importance of GL. There is also a high level of confidence with regards to knowledge of GL resources. As with the pre-test survey mentioned earlier, there is a noticeable pattern between recognizing the need for GL coupled with the confidence and knowledge to be able to search it, hence the overlap between percentages reported. Further, while the purpose of consulting GL material is primarily to satisfy research requests (89%), two-thirds (67%) of respondents indicated that they also make use of grey material for their own self-learning. These statistics indicate that GL serves as key resource material beyond a one-sided research purpose. In addition, a growing trend has emerged with regards to types of GL resources consulted (Figure 1), as indicated by the follow-up survey. Search engines have established the top tier, with government resources and databases following closely behind. Due to extensive promotion of the importance of open access, DSpace, the University of Calgary's institutional repository was also reported to be consulted on a regular basis, along with the library catalogue.

(Figure 1: Comparison of GL Resources Consulted Pre-test and Follow-up Survey)



CONCLUDING REMARKS: THE FUTURE OF OUR GREY LITERATURE INITIATIVES

As exemplified by the methods undertaken to raise awareness of GL, the use of the Health Belief Model framework is helpful in the continued pursuit of promoting the importance of non-traditional material in an academic community. Based on feedback received from the pre-test and follow-up surveys, together with the One-Minute–Paper, the targeted audience of our programs presented here have undoubtedly undergone a transformation with regards to their beliefs, attitudes, and behaviors towards the nature of this elusive material. As promotion of GL in the academic sector continues, education programs must address relevant beliefs and continue to target different learners.

As many researchers can attest to, GL is "often among the first sources of information on a topic or activity [and]…is sometimes the best/only resource ever issued and may offer unique information and insights not available elsewhere" (Childress, 2003, p. 4). In order for GL to continue down its current path of success, questions inevitably arise as to how to ensure that various types of GL (from open access journals to institutional repositories) continue to be replenished and keep the academic community aware of the latest research projects.

Via the initiatives discussed in this paper, along with a few new ideas that are already forming in our minds, we believe that our continuous promotion of GL as a legitimate, required supplement to the literature review, will further narrow the distinction between grey and white, leading to new information-seeking standards prevalent across an entire academic institution. The study that we have undertaken, via an automated survey-collecting instrument, is in essence a feature or type of GL as well. Wide-sweeping generalizations of the survey data collected have allowed us to improve and enhance the delivery of our GL sessions by anticipating the needs of our users and focusing on a particular aspect of a lesson that may be more difficult to retain in memory.

There are a variety of ways in which an academic institution can choose to make GL material available to the clientele it serves. The educational interventions we have introduced are but one possibility. Nevertheless, efforts to continue promoting GL to our audience must be continued. Armed with a heightened awareness of the importance of including GL material in research, it is also the responsibility of the user to think and decide how to use this newfound GL information that determines how worthwhile a research endeavor will be.

References

Banks, M. (2004). Connections between open access publishing and access to gray literature. *Journal of the Medical Library Association, 92*(2), 164-166.

Childress, E. (2003). Going grey: Grey literature and metadata. *Journal of Internet Cataloguing, 6*(3), 3-6.

Leahy, R.L. (1996). *Cognitive therapy: Basic principles and applications*. Northvale, N.J.: Aronson.

Lloyd, F.J., & Reyna, V.F. (2001). A web exercise in evidence-based medicine using cognitive theory. *Journal of General Internal Medicine, 16*(2), 94-99.

National Association of Cognitive-Behavioral Therapists, NACBT (2007). *What is cognitive-behavioral theory*. Retrieved November 9, 2009 from http://nacbt.org/whatiscbt.htm

Schwann, M., Petermann, L., & Petz, G. (2008). The E2D2 Model: A dynamic approach to cancer prevention health interventions. Presented at the *Chronic Disease Prevention Alliance of Canada (CDPAC) Third National Conference, Ottawa, November 26-28, 2008.* Retrieved November 22, 2009 from http://www.cdpac.ca/media.php?mid=558

Siegel, G. (2004). Capturing academic grey literature: Starting at home. *Publishing Research Quarterly, 20*(1), 62-69.

Stempfhuber, M., Schaer, P., & Shen, W. (2008). Enhancing visibility: Integrating grey literature in the SOWIPORT information cycle. *The Grey Journal, 4*(3) 121-128.

Sulouff, P., Bell, S., Briden, J., Frontz, S., & Marshall, A. (2005). Learning about grey literature by interviewing subject librarians. *The Grey Journal, 1*(3), 131-136.

Tsang, R.L.M. (1999). *Teenagers and sun protection*. Unpublished Master of Science, University of Alberta, Edmonton. Retrieved November 22, 2009 from http://www.collectionscanada.gc.ca/obj/s4/f2/dsk1/tape9/PQDD_0027/MQ40120.pdf

Watkinson, A. (2006). Open access: A publisher's view. *LOGOS, 17*(1), 12-21.

Wayne State University. (n.d.). *Theories of Health Behavior, HPR 8540: Theory at a glance* (selected readings). Retrieved November 22, 2009 from http://www.kinesiology.wayne.edu/Theories%20of%20Health%20Behavior.pdf

[1] This Grey Literature Speaker Series Wiki was initiated as a private forum to facilitate grey literature discussions and planning details surrounding proposed presentations, with the future plan of making the resources freely available to the general public.

Appendix 1 - Pretest Surveys - Grey Literature Session Evaluation

1. Have you searched or used Grey Literature before? (Yes, No)
2. Under what circumstances have you used Grey Literature? (Research, Teaching, Self-learning, Other)
3. Where have you looked for grey literature? (Search Engines, Association and Government Websites, Databases, Clinical Guidelines, Print Materials, Library Catalogues, Institutional Repositories, Other (please specify)
4. What aspect of Grey Literature would you like us to inform you about at the session? (Introduction to grey literature, Current trends in grey literature, List of grey literature resources, Sample searches, Other (please specify)
5. What is your role? (Librarian, Faculty Member, Graduate Student, Undergraduate Student, Physician, Nurse, Administrator, Researcher, Research Associate, Other (please specify)

Appendix 2 – Follow-up Survey October 2009 - Grey Literature Sessions

1. Have you used/searched GL since you attended our session/participated in our GL discussions? (Yes, No)
2. Under what circumstances have you used Grey Literature recently? (Research, Teaching, Self-learning, Other)
3. Where have you looked for grey literature? (Search Engines, Association and Government Websites, Databases, Clinical Guidelines, Print materials, Library Catalogues, Institutional Repositories, Other (please specify)
4. Has your confidence changed in the following when you do research? (Awareness of including GL in research, Perceptions of GL resources, Comfort level of using GL resources, Abilities to search GL)
5. What is your role? (Librarian, Faculty Member, Graduate Student, Undergraduate Student, Physician, Nurse, Administrator, Researcher, Research Associate, Other (please specify).

The Integration of Grey Literature and Primary Research Data in Open Source Analysis: The Jordan Property Regime

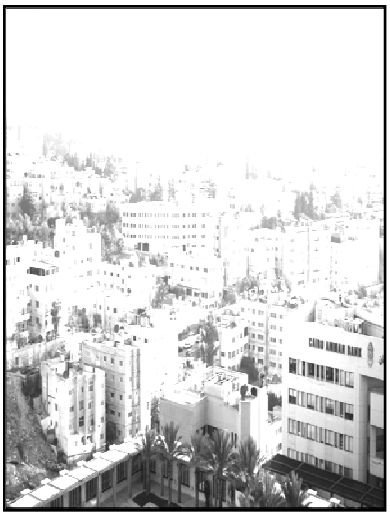
June Crowe and Bonnie C. Carroll,
Information International Associates, Inc., United States

Abstract
The Jordan Property Regime project was an exciting and challenging in-country research project that combined primary and secondary research, the latter containing substantial amounts of grey literature. Since there is not sufficient time to discuss all of our research in detail, this paper provides the general background of the project, our research methodologies, and highlights our research results and conclusions as they relate to women's rights to land ownership in Jordan. This paper also highlights the invaluable benefit that grey literature provided the researchers to fully discern the cultural, socio-political, and economic factors relevant to the study; this research certainly could not have been as readily substantiated using primary, traditional reference works alone.

Background

Bowman Expeditions
The Bowman Expedition is a global concept in which geographers, scholars, and researchers go to foreign countries to establish or enhance a mutual understanding and appreciation of the legal basis of land ownership, how it is used, mapped, registered, and taxed. The Bowman Expeditions are named in honor of the former American Geographical Society Director, Isaiah Bowman. The expeditions are based on the belief that geographical understanding is essential for maintaining peace, resolving conflicts, and for providing humanitarian assistance worldwide. The Bowman Expeditions provide insights on some of the human conflicts that spring from ruptures in the relationships between land and people. The information discovered in the research process can serve as a tool for foreign policy makers to capture the multi-dimensional and complex sets of land rights and interests that are keys to predicting and responding to global crises.[1] The México Indígena Project was launched as the first Bowman Expedition in 2005. Other Bowman Expeditions have been made to the Antilles Region, Colombia, and Kazakhstan. IIa's project on Jordan's Property Regime was a part of the Bowman Expeditions and was funded by the U.S. Foreign Military Studies Office.

Map 1. Map of Jordan



A view of Amman, the capital of Jordan.[2]

Background on Jordan

Geographically, the Hashemite Kingdom of Jordan (Jordan) is in the Middle East at the head of the Gulf of Aqaba as can be seen in Map 1. Amman is the capital and largest city with about 2.5 million people. The entire country has just over 6 million people, about 70% of whom are urban. There are 12 political governorates or administrative divisions. Jordan has a total area of about 89,342 sq. km and is slightly smaller than the state of Indiana.[3] The official language is Arabic, with English being widely spoken. Jordan has spectacular geographic diversity, ranging from desert plateaus in the east, highlands in the west, and the Jordan Valley, which is part of the Great Rift Valley that separates the East and West Banks of the Jordan River. About 80% of Jordan's land is owned by the state with approximately 95% of its cadastre being available electronically in a geographic information system.

Research Period and Participants

Our project was carried out over an intensive 12-month research period during September 2007 through September 2008 and was a cooperative enterprise between our American team, our Jordanian partners at Infograph, and the professors and students at Al-Balqa Applied University in Salt. We crafted a wide range of research methods to analyze the Jordan's current cadastral system and relate how the current system meets the demands of individuals and society as a whole. We implemented our research design through three in-country field research trips. Throughout the project, we worked with more than 25 local Jordanians who assisted in data collection and knowledge production. Among our partners were university professors, students, attorneys, surveyors, realtors, women activists, and geographic information system (GIS) experts. We are especially indebted to the Amman-based Infograph Company and Al-Balqa Applied University in Salt for guidance and facilitation of our access to various governmental sources and documentation.

Focus of the Research

A cadastre is usually a parcel-based land information system that contains land records (e.g., rights, restrictions, responsibilities). A geographic description of land parcels is linked to other records that may describe the ownership or control of the parcel, the value of the parcel, and any improvements to the parcel. Such records are used for recording ownership, taxation, and land management. As a first step, our research examined the accuracy and clarity of location, comprehensiveness, and transparency of the Jordanian cadastre.

We looked at the accuracy and reliability of the data and how it was mandated; whether the system was centralized or de-centralized; how the information was collected (via ground surveys tied to geodetic control, uncoordinated ground surveys and measurements, aerial photography, digitizing existing historical records); the types of rights recorded; as well as the history, culture, and traditional land tenure arrangements in Jordan.[4]

The focus of our research was on the overall property system as well as special topics, including women's issues in land ownership which will be discussed in this paper to illustrate the importance of socio-cultural factors as a source of grey information.

In order to fully understand perceptions regarding women's land ownership rights, it was essential to understand some of the complex cultural variables, including religion and historic traditions at work in Jordanian society. These factors include what is known as the Triangle - God, the King, and the State. Jordanians will not criticize these specifically, but will criticize women's roles based on underlying perceptions regarding the applicability of the law and practice. Grey literature discoveries during the primary research process provided the insights into some of these cultural variables.

Research Methodologies

Our data collection methods were based on accessing information from a wide array of open sources, grey literature, including government publications, technical reports, personal land plats, conference proceedings, maps, private sector company information, archived records and publications, theses, academic and trade journal articles, and websites. Prior to our in-country field work, we conducted a vast amount of open source research to discover as much as possible about Jordan's land system.

Surprisingly we found that there is a substantial amount of information about Jordan's land system on the Internet. In fact, Jordan's Department of Lands and Survey (DLS) has a

website that includes an online geographic information system of the cadastre, or ownership registry, available to the public. The collective efforts involved in the various research activities yielded substantive findings relating to the many facets – historical, socio-economic, cultural, and technological – of the land management system in Jordan.

The grey literature we discovered through our research design enriched our primary research, providing us with not only context but more details than we would have found otherwise. For example, through our contacts in Jordan we found company documents which gave us some background on Jordan's Department of Lands and Survey, the central agency responsible for all land transactions in the country. We might add that the documents we received were proprietary, which moves darker on the grey scale, but we did receive permission to use them.

As a result of our meetings with the various personnel at the DLS, we obtained brochures, presentations, a set of laws dealing specifically with land transactions, maps, and a copy of a master's thesis on our topic. Several of the DLS staff regularly made presentations about Jordan's cadastre at international conferences such ESRI and International Federation of Surveyors (FIG); we were able to locate these conferences and the papers. The DLS also publishes a monthly statistical report on land transactions for both commercial and private sectors. All of these grey resources played an essential role to our understanding of Jordan's property system.

Our primary research involved interviews with local experts, population surveys of representative samples, focus groups, and on-the-ground land surveys for comparisons with land measurements and descriptions on file at the DLS. Using multiple techniques and strategies, we were able to arrive at a well-informed understanding of Jordan's land property regime. The grey literature that provided an understanding of the region's culture was essential in understanding key aspects of land ownership. Cultural variables that most influenced our findings were based on gender, class, political and economic power, religious faith, and identity.

Table 1 below, "Secondary Open Source and Grey Literature Research Sources," is illustrative of the types of grey literature sources that were consulted and proved to be useful for our analysis. These sources were important to improving our understanding of our initial, primary research results because this information provided additional background, context, and insights into Jordanian politics, culture, and society concerning land ownership.

Table 1. Examples of Secondary Open Source and Grey Literature Research Sources.

Conference Proceedings	Government Publications	Non-governmental Organizations-Reports	Journals (Trade & Academic)	Other
9th Middle East and Africa Conference for ESRI Users 2006	Department of Lands and Survey Jordan – Land Registration System	United Nations	Photogrammetric Engineering & Remote Sensing – grids/datum	Options and Possibilities to Improve the Cadastre in Jordan (Thesis)
International Studies Association Mtg.	Official Newspaper of the Hashemite Kingdom of Jordan	Library of Congress	The Muslim World	Infograph –Land Management System of Jordan (Proprietary Company Report)
FIG Congress—Building Jordan's Digital Cadastral Data Base	Land Laws of Jordan	Institute of Women's Studies, Birzeit University	Journal of Middle East Women's Studies	GIS for Petra Region Authority (Proprietary Company Report)
	Government Publications, Presentations, and Brochures	Centre on Housing Rights and Evictions	The Ahfad Journal	Special Reports, i.e., Women's Rights in the Middle East and North Africa by Sanja Kelly
	Jordanian Library Collections	International Land Coalition	Journal of Development	

As a result of our primary research which involved interviews, meetings with government agencies, open source research, and focus group discussions, we discovered additional grey literature. The context and cultural insights extracted from these sources provided us with a

more in-depth understanding of the respondent's answers to the questionnaires. The grey literature information informed the primary research findings and influenced the overall conclusions of the research. The following high-level diagram illustrates how secondary grey literature research was combined with primary research that resulted in a more informed analysis of our data.

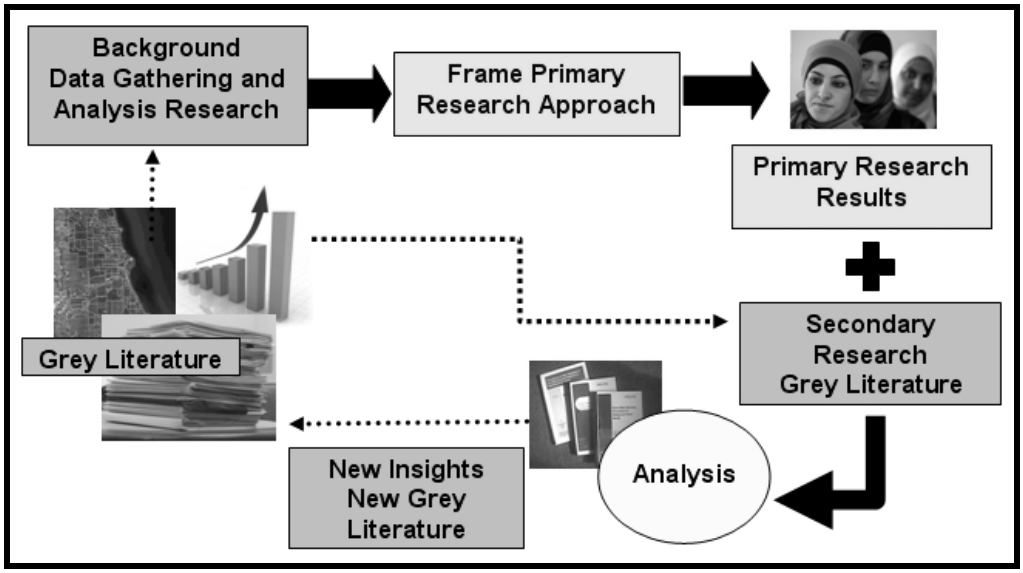


Illustration 1: Process Flow Diagram.

As can be seen, our background research helped to frame our approach to the primary research. Then, as we conducted our primary research, we uncovered yet more grey literature which assisted in our analysis. Highlights of secondary research provided key observations that led to the identification of complex issues impacting this study. Discussed in greater detail in the following section, these salient topics centered on women's highly unequal ownership of land, practices related to Islamic laws of succession, and current socio-economic trends related to land ownership issues.

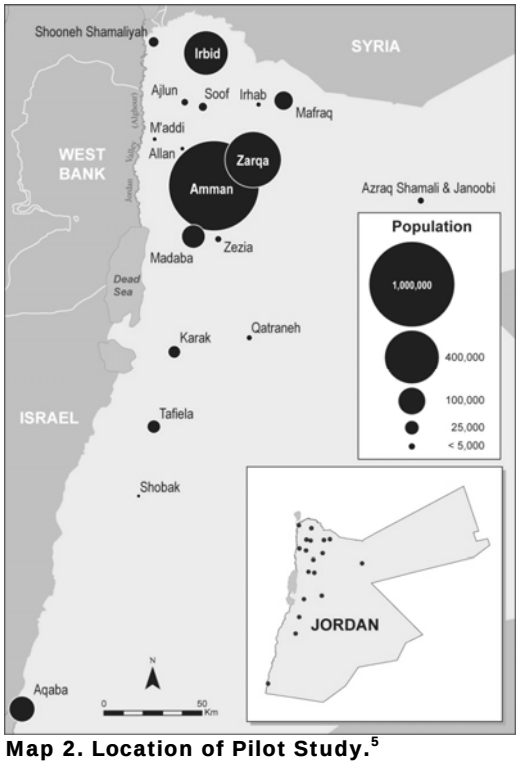
Primary Research Results

About the Three Surveys

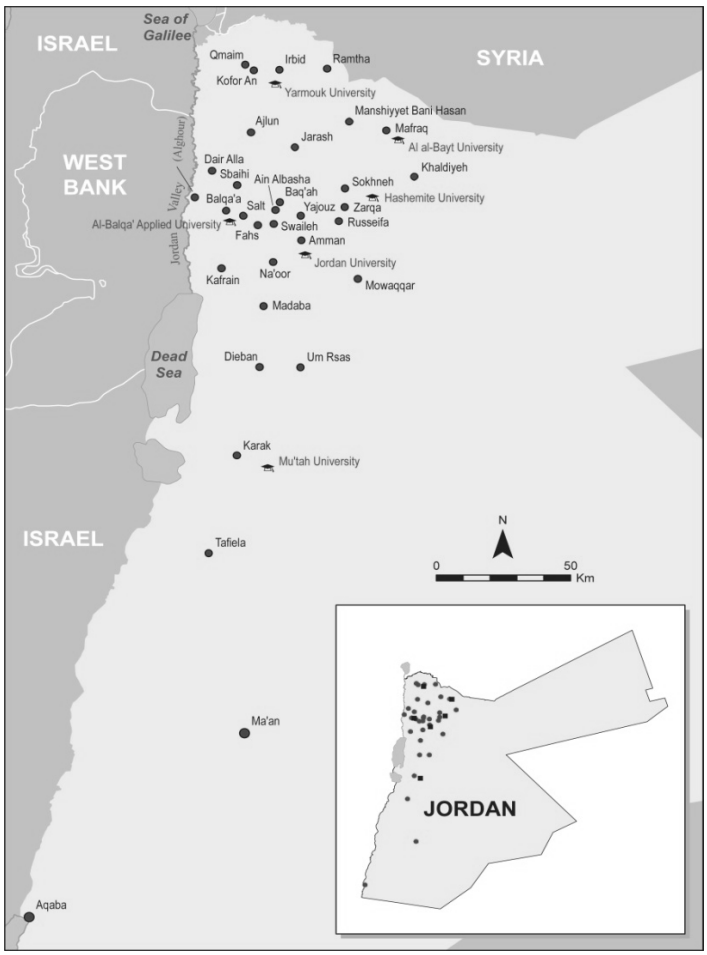
Our team devised three questionnaires in Arabic that were concerned with land issues for which we gained permission from the Jordanian government to administer.

- The first survey was a pilot survey conducted in 6 sample localities of 15 households each. The pilot served to validate the survey instrument, establish an interview protocol, and serve as a test base for the household selection process. The pilot also provided the means to validate the localities as being representative sites of Jordan's cadastre.
- The second survey was designed to gather data on student opinion regarding land issues. This survey was conducted at 6 universities with 150 student respondents at each university.
- The third survey was a comprehensive survey covering 22 widely dispersed localities with 1,660 respondents. The goal of the comprehensive survey was to administer a questionnaire to a representative sampling of the population and the geography of Jordan. We selected locations that provided a stratified approach with a variety of population densities and geographic diversity. This survey domain included urban and rural areas, mountainous regions, and desert areas across the country as well as refugee camps.

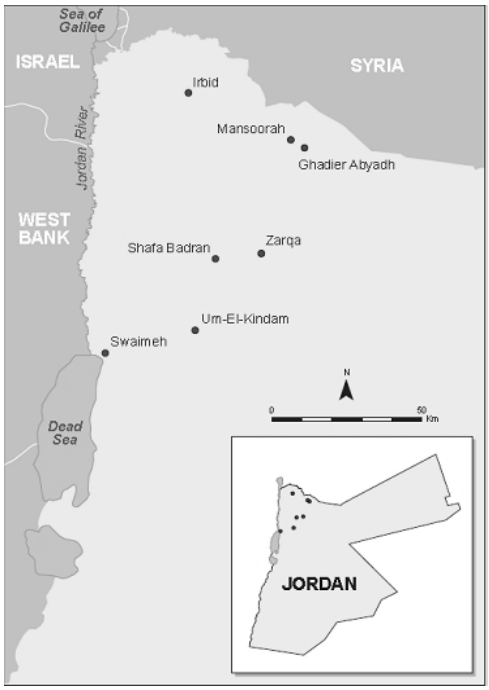
The areas surveyed covered a wide range of both Jordan's population and geographic areas ranging from the Dead Sea area to more densely populated areas in Zarqa in the northwestern region. In total, the three surveys had 2,600 respondents and were distributed among various groups of the Jordanian population and in various localities.



Map 2. Location of Pilot Study.[5]



Map 3. Locations of Student Respondents in Relation to 6 Major Universities. [6]



Map 4. Comprehensive Survey Locations Showing Population Density [7]

The comprehensive survey involved visiting 22 locations, including five neighborhoods in Amman City. These locations were divided into groups that were designed to provide a representative sample for the population of Jordan. The grouping was based on criteria involving size, population density, and type (e.g., rural, urban). A systematic, random sampling technique was used to determine the households to be surveyed. Of the 1,666 persons who participated in the survey 72.4% were male and 27.3% female, ranging in age from 18 to 85.

Table 2. Locations Grouped According to Geography, Population, and Income.

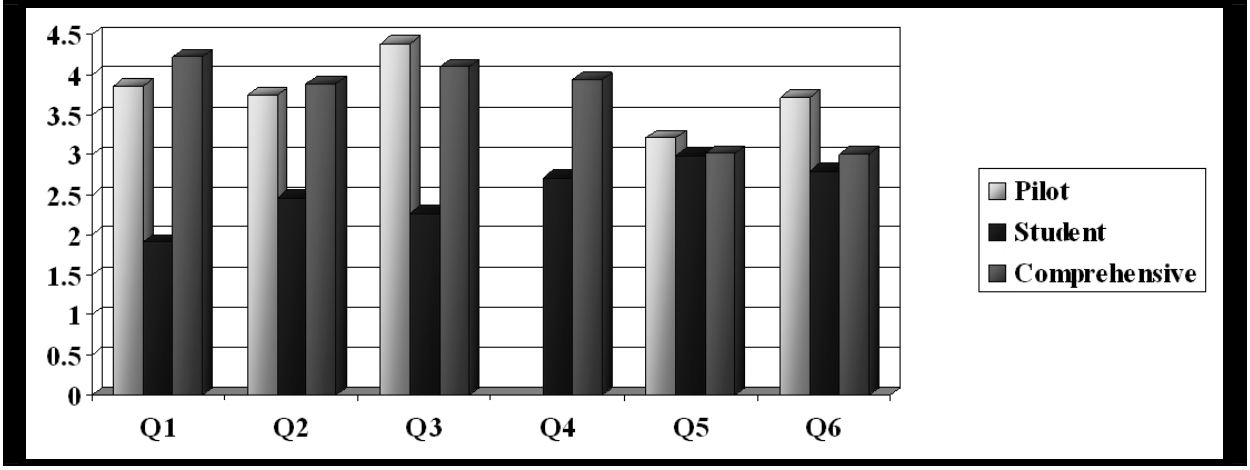
Setting	Group	City	Population in 2007	Income Level	Other Features	Number of Surveys
Urban	1	Population - 1,135,733*; Five residential areas (called "neighborhoods," but in fact each forms a sub-city within metropolitan)				
		Abdun	60,695	High	Included in Zahran	40
		Marka	102,008	Middle		40
		Ras Alain	104,782	Low		40
		Swaileh	70,830	Low		40
		Alweibdeh	51,933	Low to middle	Included in Alabdali	40
Urban	2	**Irbid**	275,054			120
		Zarqa	440,736			120
		Aqaba	94,251			120
Urban	3 (smaller and spread out across the country)	**Mafraq**	52, 622			90
		Ajlun	8,080		Substantial Christian population	90
		Madaba	77,457			90
		Karak	22,165			90
		Tafiela	25,021			90
Rural	4	**M'addi**	3,264		In central and northern	70
		Shooneh Shamaliyah	15,764			70
Mountain	5	**Allan**	3,126		One in north, one south; livelihood highly rural	70
		Shobak	2,357			70
***Palestinian refugee camps**	6	**Soof****	11,912		Near Jarash city	70
		Zezia***	7,262		Next to Madaba city	70
Desert	7	**Irhab** (near Mafraq to the southeast)	3,952		Bedouin, with strong tribal affiliation despite living in permanent dwellings	70
		Qatraneh (northeast of Al Karak)	5,072			70
Desert	8	**Azraq Shamali (North)** **Azraq Janoobi (South)**	5,278 1,960		Only Jordanian locality with substantial number of Druze as well as another ethnic Muslim group called Sheshan	70

The remainder of the paper discusses some of the key findings regarding women's land ownership rights. The discussion includes an analysis of six questions from each of the surveys concerning Jordanian Women's Land Ownership Rights; this analysis illustrates the importance of our grey literature resources.

Composite Survey Results

Table 3 shows the composite responses to the questions concerning women's rights to own land in the three surveys. The possible choices were based on a Likert scale from 1 – 5 where 1 denotes "strongly disagree" and 5 denotes "strongly agree" with the statement in the questionnaire.

Table 3. Composite Table of Statistical Mean of Responses Related to Women's Rights to Own Land.



Survey questions:

Q1. Jordanian law permits women equal opportunity to own land.
Q2. Jordanian society permits women equal opportunity to own land.
Q3. Jordanian women use their rights to own and register land.
Q4. Jordanian women practice religious rights to inherit land.
Q5. Jordanian women tend to forsake their rights to own land.
Q6. Fathers, in their wills, ensure their daughters inheritance to own land.

Pilot Survey Results

Except for question 5 (mean of 3.21), these results indicate a somewhat neutral position. This result implies that Jordanian law and society provide equal opportunity for men and women to own land. For example, respondent answers to question 3 at a mean of 4.39 are between "agree" and "strongly agree" with the statement indicating that women in Jordan are effectively using their rights to own and register land as their religion allows. There is a tendency here to agree with the statement because they do not want to criticize their religion. Yet, question 4 with a mean score of 3.21 suggests that respondents are more inclined to criticize women as individuals for their practice of forsaking their rights in favor of their brothers. Here a pattern emerges that shows respondents are more inclined to blame women as individuals (3.21) than the woman's father (3.72), "society" (3.74), State law (3.86), or least of all criticism if any of religion (4.39).

University Students Survey

The second survey was distributed to a sample of 900 students enrolled at Jordan's six major universities. The questionnaire contained 16 questions related to land issues with respect to their places of residence. This resulted in a fairly large poll and geographical cross section of students from 33 different localities in Jordan.

Unlike the pilot study answers, students tended to be more critical in that their responses leaned toward "strongly disagree." They disagreed most with the first statement "Jordanian law permits women equal opportunity to own land just as men do" with a mean response of 1.91. This can be interpreted as openly criticizing state law for not providing equal treatment for women. For the remaining statements relating to legal and traditional aspects of land ownership and inheritance for women, all students' mean answers fall in the range between 2.00 (disagree) and 3.00 (neutral).

The university student survey results were not unexpected, yet somewhat surprising given the conservative nature of Jordanian society. Student respondents to these questions may have been influenced by their generation and level of education. Students are often described as being critical in their views of society in many cultures.

Comprehensive Survey

This comprehensive survey contained 34 questions, which were extracted from the pilot survey instrument. The questionnaire was refined by eliminating nine questions from the pilot study questionnaire. These nine questions were either marginal to our research, or we were specifically instructed by the Jordanian Authority (Department of Statistics) to drop them from the survey.

The total number of questionnaires distributed and analyzed was for 1,666 households. We originally planned to distribute 1,640 questionnaires; however, in the process of distribution, some students interviewed more people than necessary, so these extra questionnaire responses were added to the totals in their respective localities.

For the first four statements, respondents in the comprehensive survey favorably agree with the statements; their answers are in the range 3.89 to 4.22. This indicates that they agree that Jordanian Law, society, and religion provide equal opportunity to women when it comes to owning land in the Kingdom. Yet, for the remaining two statements, 5 and 6, respondents provide a mean response of 3.02 and 3.01, respectively, indicating their answer is in effect neutral. In other words, when responsibility is placed on the women themselves as individuals or their fathers, the respondents are less positive and adopt a stance of neutrality.

A comparison of the three surveys yields the following trends:

a) Student responses to all six statements are less positive than the responses of pilot and comprehensive respondent answers, and their mean answer leans more toward disagreement with the statement;
b) For the two statements related to the rule of state law and society in arranging equality in land ownership for women, the pattern suggests that comprehensive household survey respondents agree the most with the statement, followed by pilot study respondents; students indicated the lowest level of agreement;
c) For the three statements related to religion (#3), women as individuals (#5), and fathers (#6), pilot study respondents tend to agree the most with the statements, followed by comprehensive household survey respondents; students agree the least. Student responses to these statements fall mostly between "neutral" and "disagree."

Focus Group

Finally, another interesting cultural factor was discovered during the focus group session. There were four participants, 3 Moslems and a Christian. Participants included a Senator, Government Employee, GIS Director at a university, and a Lawyer. These were women who specialized in women's rights in Jordan. The discussions lasted for three hours; they were held in Arabic and were recorded.

The discussion centered on distinguishing between theory and practice regarding women's inheritance rights to land. State laws and religion in Jordan provide women and men with equal opportunity to own, inherit, register, and benefit from land, and there is no discrimination whatsoever based on gender. In practice the reality is largely based on individual decisions made by male members of certain families.

There are differences between rural and urban societies in regards to women owning land. In rural localities, people are more traditional and usually do not allow females to inherit land; all shares are divided among the male members of the family after the death of the father. This practice is common in many Arab countries; it may be viewed as an inheritance tradition and not as discrimination against women.

The traditional justification for such practices is that once females are married and go to live with their husbands, they should not inherit the land of their father. The land is kept as the property of the family and thus has to be inherited by the male members, i.e., the brothers only. Surprisingly, at least to the Principal Investigators, Christians in Jordan also follow Moslem practice as far as land inheritance is concerned.

A short questionnaire was given to the women in the focus group. The group strongly agrees that the Jordanian law provides women with equal opportunity to own land as men do. However, Jordanian society is less enthusiastic about this option, and this is why women place their answers between 3 and 4 on the Likert scale.

In question 3, we have a similar result. The group does not think that women in Jordan are fully utilizing their right to register lands in their names as their religion allows them to do. Their answer to these two questions tends slightly toward agreeing with the statements.

Group answers of "3" for questions 4 and 5 are neutral or indicate that such a statement is 50% correct. In other words, the women in Jordan do not all choose to act on their rights to inherit land as Islamic practice permits, and there is a tendency to forsake lands in favor of brothers among at least 50% of women in Jordanian society.

For question 6, the group answer was placed between 2 and 3, suggesting some tendency to disagree with the statement. This suggests that there exists a probability that more parents than average fail to ensure in their wills that their daughters will inherit the land and register it in their own names after the father's death.

Women's Focus Group Results Strong Disagreement……..Strong Agreement				
1	2	3	4	**5**
(1) Jordanian law permits women equal opportunity to own land just as men do.				
1	2	**3**	**4**	5
(2) Jordanian society permits women equal opportunity to own land just as men do.				
1	2	**3**	**4**	5
(3) Jordanian women use their rights to own and register land in their own name as their religion allows.				
1	2	**3**	4	5
(4) Most Jordanian women practice their religious rights to inherit land and register it in their own name.				
1	2	**3**	4	5
(5) There is a tendency among Jordanian women to forsake their rights to own and register land and give it to their brothers.				
1	2	**3**	4	5
(6) Most fathers in Jordanian society ensure in their wills that their daughters inherit their land and register it in their own names (or benefit from the land) after the father's death.				
1	**2**	**3**	4	5

Conclusions

As noted previously, evaluating perceptions and practices regarding women's land ownership rights in Jordan offered insights into complex cultural variables and factors relative to the Triangle of God, the King, and the State. Our survey showed that students tend to be more radical in their thinking about women's rights in land ownership. As was seen from their responses, they openly criticized Jordan's state law but still remained neutral on Shari'a law.[8] [9] Shari'a is the textural stipulation of law based on the Qur'an.[10] Shari'a is the legal framework within which the public and private aspects of life are regulated for those living in a legal system based on Islamic principles of jurisprudence.

Our primary research involving focus groups, interviews, and surveys uncovered more grey literature which gave us the background of Jordan's system of law, the Islamic rules of inheritance, and explained some of the barriers to women's inheritance rights. From these we garnered key insights into the perceived impediments for Jordanian women to own land. In theory Jordanian women are equally treated by Jordanian law on a par with men. However, social practice and attitudes among women themselves may act as an obstacle for women to exercise their right and register real estate in their own names as men do. Awareness among women of their rights in this regard varies geographically. Women living in urban localities are more determined to exercise their equal right with men to own land than those who live in rural areas, where family members often pressure them to forsake their rights in favor of male family members.

Significantly, disempowerment of women in Jordan is often blamed on the actions of women themselves and society to a certain degree, but less on state law and religion. Based on such findings, women in Jordan are clearly "trapped" by social custom and practice, and it may take an extended period of time to achieve in concrete practice what Jordanian law offers them in theory.

[1] Douglas E. Batson, Registering the Human Terrain: A Valuation of Cadastre. National Defense Intelligence College. 2008. http://www.ltera.org/index.php?view=article&catid=127%3Apublications&id=252%3Aregistering-the-human-terrains-a-valuation-of-cadastre&option=com_content&Itemid=182&lang=en

[2] Photography courtesy of the author. December 2007.

[3] CIA World Fact Book. https://www.cia.gov/library/publications/the-world-factbook/geos/jo.html#

[4] Mouen Mousa Sayegh, "Options and Possibilities to Improve the Cadastre in Jordan." (Master's thesis, January 2003)

[5] Map created by Will Fontanez, Director of Cartography, University of Tennessee. September 2008.

[6] Map created by Will Fontanez, Director of Cartography, University of Tennessee. September 2008.

[7] Map created by Will Fontanez, Director of Cartography, University of Tennessee. September 2008.

[8] Yvonne Yazbeck Haddad, John L. Esposito Islam, gender, & social change. 1998. p. 105. http://books.google.com/books?id=RkbZWrU4UfUC&pg=PA105&lpg=PA105&dq=jordan+sharia+law&source=bl&ots=4k1Rn3l8-Q&sig=tfaOIahD0OzzuuXrBNt1QFKu13Q&hl=en&ei=OrYAS_SGFYfdnAfmpaWRCw&sa=X&oi=book_result&ct=result&resnum=3&ved=0CAwQ6AEwAjgK#v=onepage&q=jordan%20sharia%20law&f=false

[9] Anne Marie Goetz. Governing Women: Women's Political Effectiveness in Contexts of Democratization and Governance Reform. 2008. p. 54.

[10] In Search of Equality: A Survey of Law and Practice Related to Women's Inheritance Rights in the Middle East and North Africa (MENA) Region. Centre on Housing Rights and Evictions (COHRE). Geneva, Switzerland. October, 2006

The Secrecy of Grey Unveiled: Grey Literature and the Freedom of Information

Cees de Blaaij
Library of Zeeland, The Netherlands

"to pierce the veil of administrative secrecy and to open agency action to the light of public scrutiny"
US Supreme Court (US Department of the Air Force v. Rose, 425 U.S. 352 (1976))

Abstract
This paper examines the relationship between political culture and information control. What does 'freedom of information' mean? Governmental institutions produce a diffused body of grey literature. Some parts of these vast grey information resources are accessible but a significant number are not for a number of reasons. On one hand the public in democratic societies has a right to know as formulated in freedom of information laws. On the other hand problems of national security do prevent and restrict access to grey resources. Secrecy has to be maintained because it concerns the quintessential function of the state. There is a public need to know if open government is the preferred political culture. But in an age with an ongoing war on terror, governments are understandingly less eager to stimulate an open society. Openness might harm the national security.
This article evaluates (a) The historical background of government information resources in relation to freedom of information, (b) The working of the FOIA (Freedom of Information Act) in the United States and Europe in general and the public access to grey governmental resources, and (c) more specific the impact of the events of 9/11 and public access to grey government resources. Indicators show that government secrecy has continued to rise since September 11, 2001. However some progress has been made. In general the use of the different freedom of information laws in democratic countries to obtain information from the government continues to rise.

Methodology and materials
In this paper a case study is presented on the effects of 9/11 attacks and the availability of "sensitive" unclassified governmental information within the existing framework of national Freedom of Information Laws. The used methodology is a case study based on comparative research. Data for research was gathered and analyzed from reports and quantitative/qualitative data resources from different organizations: US government agencies i.e. the US General Accounting Office, European Union institutions and non-govermental organizations . In addition relevant international literature on the topic of Freedom of Information was researched.

Introduction
Steven Aftergood, director of the Project on Government Secrecy of the US Federation of American Scientists, once said: "Information is the oxygen of democracy"[1]. If citizens are supposed to participate effectively in a democracy and be able to scrutinize the acts of governments in respect to economics, social issues, and even in issues of war then the quality of information is vital. Unfortunately this is not always the case. According to a 2006 report on the future of digital government in the European Union[2] "there is a lack of proper quality of information and certified sources of information in an era of information overload. When systems become more and more connected, low information quality can spread and might even disrupt the functioning of the public administration."[3]
How did such a situation of *data smog*[4] come into existence? When digital government was introduced in the 1990s governments in the Western world began to realize that the Internet was as a great tool to disseminate information in an efficient and convenient way. Government agencies started to provide information and online services to citizens[5] and sought a direct connection with the user.[6] The drive was to post as much government information online as possible to support the idea of openness. A great variety of document collections became electronically available in respect to parliamentary debates, laws and agency activities. Reports and papers based on government-funded research on subjects like healthcare, geospatial issues, farming and education (among others) also went online. In total these collections represent a mer à boire of different types of grey literature available from public bodies.[7]
In general there was little consideration about the amount of useful information made available to the public, or simply put how much was too much.[8] The tragic events of 9/11 in however caused a reversal in governmental thinking in the US about what was useful information to the public and what might be harmful information into the wrong hands in the war against terror. As

a result the need for secrecy to protect national security counterweighted the public interest towards open government and public access to unclassified government information was restricted.
In this paper I will focus on the dualism between the need of governmental secrecy and the freedom of information. As a relevant case study I describe and analyze the effects of 9/11 events in relation to the right of access to public information in the US and the EU.

Freedom of information: a brief overview
Freedom of information refers specifically to the right to access information held by public bodies. Getting access to unpublished governmental information means making an official request for governmental disclosure of information following the designated procedure prescribed in the US Freedom of Information Act or the relevant EU regulation.[9]

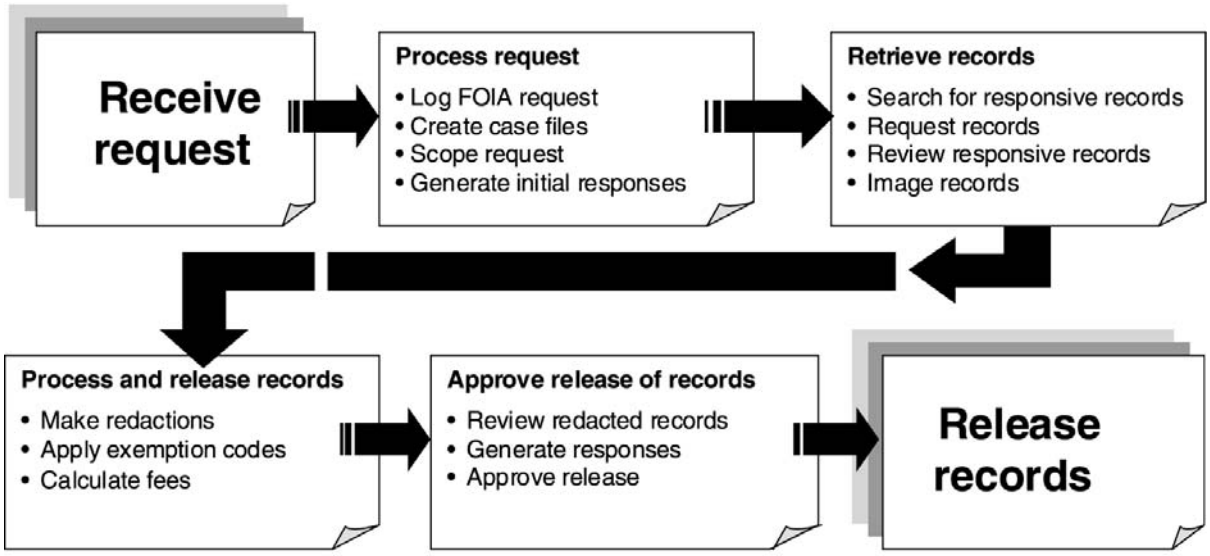


Figure 1. Overview of Generic FOIA Process. Source: GAO-01-378.

Freedom of information was never a distinct constitutional or universal right but evolved as a property of the right to freedom of expression into an independent – sui generis – human right in itself. Already in 1946 during its first session the UN General Assembly adopted a resolution that stated that freedom of information was a fundamental right[10] despite the fact that in the early stage after World War II, the international human rights instruments did not specify the right to information as an explicit right to access information held by public bodies. The implicit notion of a public right to know was nevertheless understood.
In 1966 the US implemented the first Freedom of Information Act[11] (FOIA), "to pierce the veil of administrative secrecy and to open agency action to the light of public scrutiny".[12]
Other governments were not eager to follow; twenty years later only eleven countries had similar legislation. The political climate of the Cold War caused governments to emphasize security interests outbalancing the concern for freedom of information.
After 1984 this changed profoundly, following the collapse of military regimes in Latin America. In Argentina an investigation commission published a report titled Nunca Mas (Never Again)[13] that documented the kidnapping, torturing and killing of more than 9,000 Argentines during the so-called "dirty war" in the 1970s. Evidence was based on systematic documented actions of the military regime and its bureaucracy. The official files gave concrete proof of atrocities. Such was the case also in Chili[14] and Paraguay[15]. In this way grey literature played an important role in establishing the truth that thousands of innocent people had been killed by a military regime.
In Latin America and in Eastern Europe (with the downfall of the Berlin Wall in 1989) democratic governments replaced authoritarian regimes. In the new political environment the public obtained access to once restricted sources. Archives of former security services were opened i.e. in Germany where the archives of the former Ministry of State Security in East-Berlin - known as Stasi - literally proved to be a *Fundgrube* for identifying former Stasi informants and the ways the Stasi had operated.[16] In Russia secret files were transferred to state archives with only limited access. In 1992 however the ex-KGB archivist Vasili Mitrokhin managed to deliver 20.000 pages of crucial intelligence information to the West.[17] At that moment KGB secrecy had received a severe blow and official documents were used to establish the truth about seven decades of Soviet government. Across the Atlantic in the US documents relating to the Cold War were subject to declassification.[18] [19]

The publication of documents from the intelligence agencies, police and military was a method for providing justice to victims of security agencies. It also made clear the dangers of a security sector as a state within a state when there is no accountability or control. The end of the Cold War stimulated a movement for transparent governments and the implementation of disclosure laws.[20]
The principles laid down in American FOIA were exemplary for other states. By the end of the 1990s the political ideal of open government was manifest and many believed that democracy and the relation between citizen and state would be redefined.
Supranational organizations began to promote the politics of open government with stronger statements. In 2004 the Special Rapporteur on Freedom of Opinion and Expression from the United Nations proclaimed the right to access information held by public authorities was a fundamental human right "based on the principle of maximum disclosure, establishing a presumption that all information is accessible subject only to a narrow system of exceptions".[21] In September 2006 this opinion was confirmed by a landmark decision from the Inter-American Court of Human Rights. It recognized the existence of a full right of access to information held by the government and other public bodies.[22] This court also made it clear that countries should train public officials on procedures for releasing information and that they must be guided by the principle of maximum disclosure.
In July 2006, 68 countries had the similar disclosure laws.[23] In 2009 90 countries do recognize a right to access but delays in responding to requests for information are notorious.

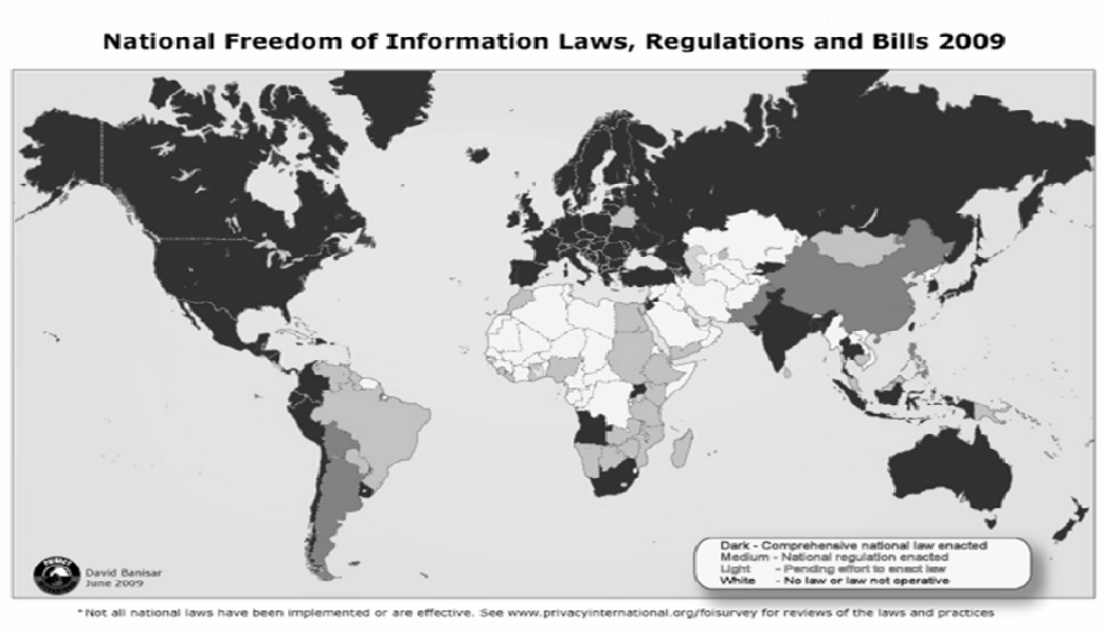


Figure 2. Source: D. Banisar, Privacy International

A survey from the Open Justice Initiative – an advocacy group - in 2006 compared how government agencies in 14 countries responded to information requests. Its analyses showed that 1,926 requests were filed. 56 percent of the request made in countries without freedom of information laws went unanswered and 38 percent of the requests made in countries with freedom of information laws went unanswered. So 47 percent of the requests did not get any response at all.[24]

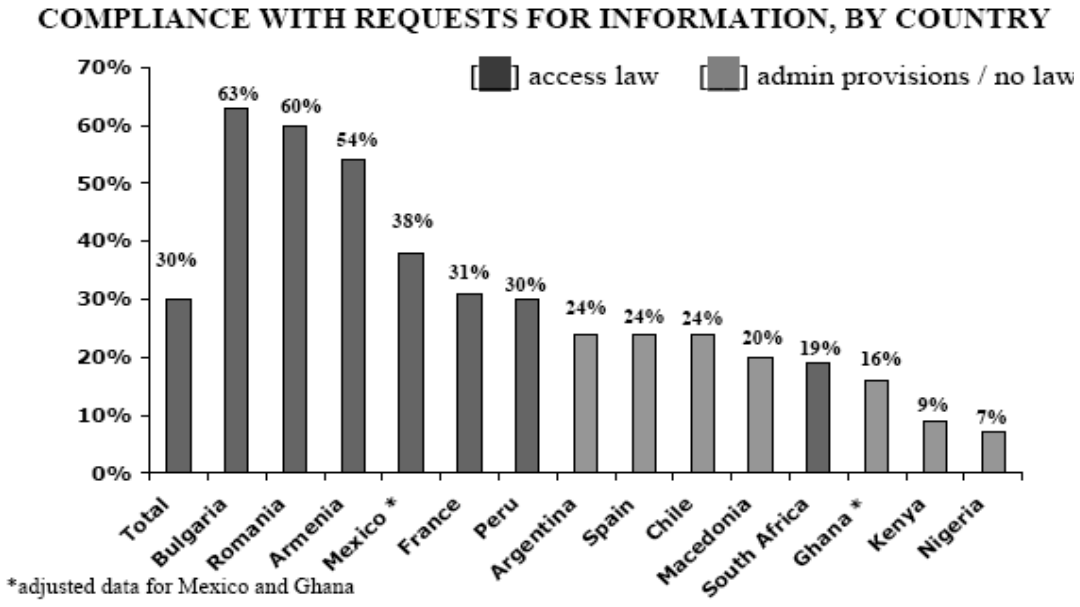


Figure 3. Source: Open Justice Initiative

The report from the Open Justice Initiative diagnoses the vitality of the FOI market. One of the remarkable outcomes is that in countries transitioning to democratic rule – like Armenia, Bulgaria, Romania, Mexico and Peru - a higher percentage of information requests got answered than did in two mature democracies France and Spain. This can be explained by the fact that the freedom of information is not only the right to request and receive information from public bodies but in a number of cases public bodies are obliged to publish information on a need to know basis even in the absence of a request. France and Spain make a significant amount of governmental information online available to the public.

Case study: 9/11 and the freedom of information

The attacks of 9/11 made government and the public aware that public information sources could be used for terrorist activities. The 9/11 Commission Report found that the terrorists had used the Internet for public information resources in mastering their plan such as flight times, research for flight schools etc. and that they communicated by e-mail. Much of this went undetected by using prearranged code words for target locations.[25]

The New York 9/11 attacks had a dramatic impact on the availability of American public information sources. Unclassified information became sensitive in the eyes of the Bush administration. Federal agencies realized that specific information in the wrong hands might be harmful and could be used for new terrorist attacks. Subsequently it was decided to remove a variety of electronic information sources and restricted access to unclassified information in libraries, archives, databases and websites.

Some examples:

- The Nuclear Regulatory Commission (NRC) closed down its website temporarily. Later on it came back online with revised content but data files were removed.
- A number of federal agencies like the Federal Aviation Administration (FAA) removed information about airline safety for security purposes.
- The Office of Pipeline Safety discontinued open access to its National Pipeline Mapping System.

There are many other examples of the removal of sensitive geospatial, environmental, (nuclear) energy and transportation information from sources.[26] The censoring of information was a kind of Pavlovian reflex and not based on a thoughtful process. To this date it is not clear to what extent US governmental information was removed from public access. Ironically, some of the removed information was already in the hands of nongovernmental sources like journalists groups, investigative reporters and science organizations.[27] In 2002 the Las Alamos National Laboratory terminated public access to thousands of unclassified reports from the Technical Report Library but the Federation of American Scientists put them back online.[28]

Estimation is that under the Bush administration the removal from public view of unclassified information went on until at least 2005. The Clinton administration's openness initiatives were reversed and the level of declassification lowered dramatically in the years after 2001.

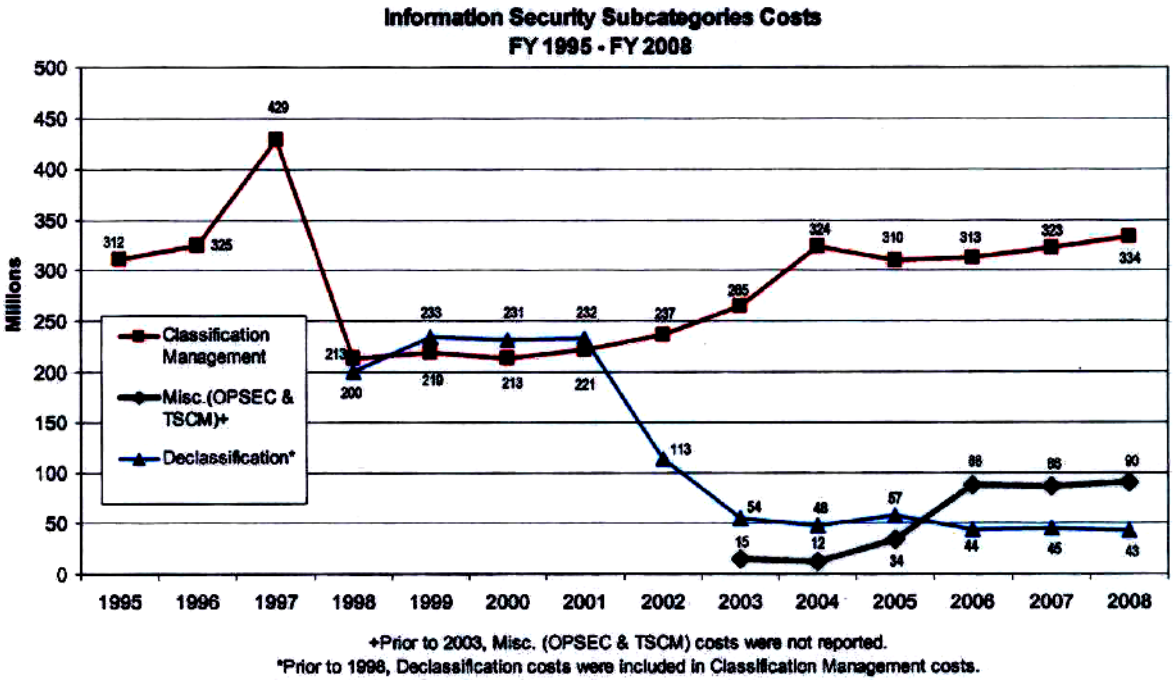


Figure 4. Source: GAO Cost Report 2008

The US Information Security Oversight Office's (ISOO)[29] Cost Report from 2008[30] shows that after 2001 declassification costs decreased with $ 157 million. After 9/11 secrecy was dramatically expanded measured by the number of newly classified documents. Roughly the government in 1995 classified 3.6 million documents. The total number of classification decisions increased from 9 million in 2001 to 14 million in 2003[31] and 16 million in 2004.[32] The director from the Information Security Oversight Office, J. William Leonard observed that even in these times of national security challenges there was excessive classification.[33] He has "seen information classified that [he's] also seen published in third-grade textbooks.[34] Thomas Kean, chairman of the 9/11 Commission observed that three-quarters should not have been classified in the first place.[35] Also the accompanying management classification costs reached unprecedented levels as Figure 5 shows.

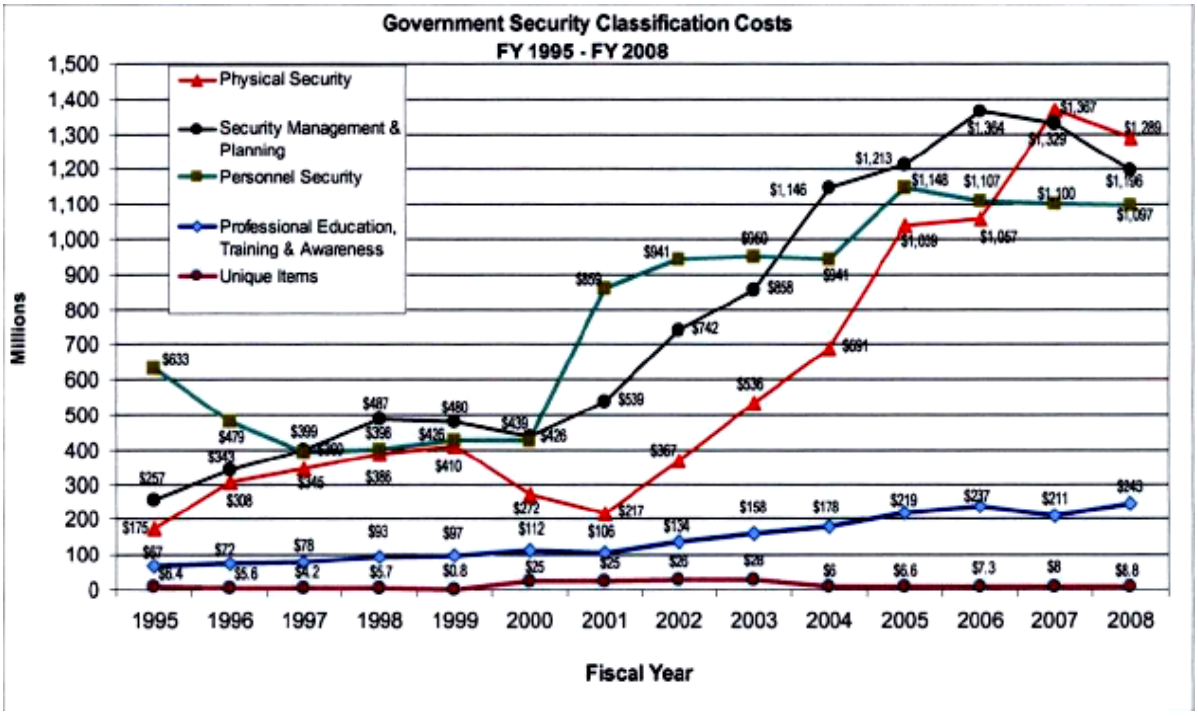


Figure 5. Source: GAO Cost Report 2008

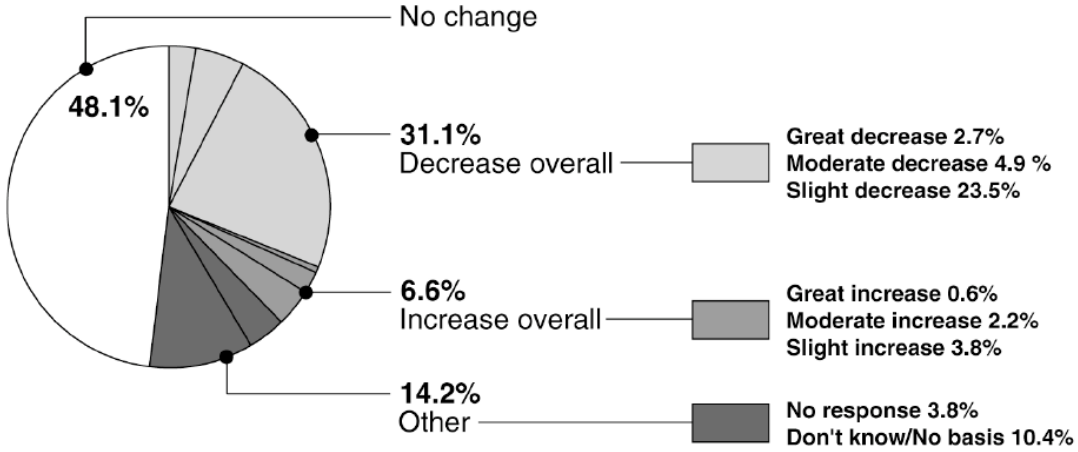
The impact of the Ashcroft and the Card memo's

The described information blackout started on October 12, 2001, when Attorney General John Ashcroft issued a memorandum to the heads of all departments and agencies concerning the FOIA policy.[36] The new Ashcroft FOIA Memorandum encouraged agencies to consider the values of exemptions in FOIA. The Justice department would defend any agency that had a "sound legal basis" for not disclosing information upon requests. In addition on March 19, 2002 White House Chief of Staff Andrew Card issued a similar memo to federal departments and agencies recommending withholding information regarding weapons of mass destruction. He referred to "sensitive but unclassified information". Because this type of undefined information didn't match the criteria for classification on national security grounds the memo was accompanied by two other memos from the Information Security Oversight Office and the Justice Department Office of Information and Privacy. The first memo explained the standards for classification. The second one explained the use of FOIA exemptions to be used to withhold information. This FOIA policy made it easier to deny FOIA requests.

The Ashcroft policy clearly contrasts with that of his predecessor Janet Reno who issued a FOIA policy of more openness. The Reno Memorandum from 1993 summarized that "it shall be the policy of the Department of Justice to defend the assertion of a FOIA exemption only in those cases where the agency reasonably foresees that disclosure would be harmful to an interest protected by that exemption."[37]

The US General Accounting Office – the investigative arm of the American Congress - afterwards conducted a survey concerning the Ashcroft memo. Of the FOIA officers who received the Ashcroft memorandum 31% indicated that they were less like likely to make discretionary disclosures.[38]

Effect of Ashcroft Policy on Likelihood of Agencies' Making Discretionary Disclosures



Source: GAO.

United States General Accounting Office

Quantitative data from annual FOIA reports shows that under the Bush administration agencies received fewer requests in 2001 and 2002 and more requests in the following years 2003, 2004 and 2005. The US Justice Department argued that posting information of public interest caused a reduction of incoming requests.[39] If this is true then the increased number of incoming FOIA requests in the years 2003, 2004 and 2005 could indicate that some public information was no longer available on the websites of different agencies. This supports the conclusion that critical information was removed from government websites after 11 September 2001. Also noteworthy is an observation from the General Accounting Office in 2002 that a number of agencies were paying insufficient attention to the online availability and the maintenance including accuracy and currency of the website content.[40]

Striking was the fact that the processing efficiency of Freedom of Information Act requests (FOIA) changed for the worse. Backlogs dramatically increased during the Bush administration[41] in comparison with the Clinton administration.

On December 14, 2005, President Bush issued Executive Order 13,392 on "Improving Agency Disclosure of Information."[42] Among other mandates, the Order calls for Agencies to review their FOIA programs and "identify ways to eliminate or reduce [their] FOIA backlog."

The delay problem however was not solved. This has been confirmed by a survey from the National Security Archive[43] from July 2007.[44]

In January 2007, the NSA filed FOIA requests with the 87 leading federal agencies and components for copies of their "ten oldest open or pending" FOIA requests. 57 agencies responded to this request, 53 reported backlogs. Twelve of the agencies that responded had

requests pending ten years or more. Of those 12, five had requests pending 15 years or more. The oldest pending FOIA requests in the federal government date back to the 1980s—one from 1987, two from 1988, and three from 1989.[45]
The impact of the Ashcroft and Card memos was also subject for investigation in 2003 by the National Security Archive.[46] They initiated a Freedom of Information Act Audit and approached a total of 35 agencies. 17 of 33 Federal departments or agencies surveyed (52 %) indicated awareness and dissemination of the Ashcroft Memo, but reported little change in regulations, guidance or training materials reflecting the new policy. Only 5 of 33 Federal departments or agencies surveyed (15 %) indicated significant changes. These agencies had a Defense background and were security related. The research of NSA confirms the outcome of the survey of the GAO. In general only small changes in policy or procedure, such as re-review of records to identify sensitive or classifiable materials were implemented however the overall guidance was to increase secrecy in security related matters as we have seen for example in the Abu Ghraib prisoner abuse scandal. The amount of decrease in declassification costs from 2001-2008 supports this conclusion.

The European Union and public access
The political situation in the European Union is very different from the US as the EU is not a federal state but an organization of national member states using methods of "multilevel governance". The position of the national state has gradually changed over the years and new ways were found to fine-tune constitutional systems in the long-term process of political integration. The shift of power to a supranational organization has caused some Eurosceptism including complaints about secretiveness. The lack of transparency was one reason why the Treaty of the European Union was rejected by a Danish referendum in 1992. Scandinavian countries like Denmark and Sweden always favored more openness but most other national governments supported the system of closed decision making of the European Council. To challenge the secrecy system the Danish ministers announced that they would use their presidency in the first six months of 1993 to declassify and make documents from the Council of ministers' meeting public.[47] The earlier rejection of the Treaty by the Danish caused some turbulence. The council made some concessions and the adoption by the Council of the Decision on Public Access to Documents on 20 December 1993 showed some progress. The influence from the Scandinavian countries gradually became stronger as France and Germany recognized the need for a new agreement towards a right of access to information. On May 30, 2001 a new regulation became applicable.[48] About the same time the Council adopted its decision regarding security regulations.[49]
This framework of two legal instruments provides citizens a general right of access to the documents of the EU institutions, although this right of access is not unlimited.
Exceptions to disclosure are established in Article 4 of Regulation (EC) No.1049/2001, which places certain limits on the general right of access. EU institutions have to (1) apply the exceptions in a restrictive manner, (2) to consider all documents requested with a view to determine whether the documents concerned can be released in full or in part, and (3) to state reasons for refusal in such a way that it does not harm the interests to be protected.
The developments in the last decades show clearly that the balance has shifted from institutional secrecy towards more institutional transparency. The Council's security regulations make it clear that documents should not be over-classified otherwise this could otherwise result in a loss of confidence in the validity of the classification system.[50] The number of Council documents classified "EU Confidential" and above is therefore relatively limited. In practice, most of these documents concern the Common Security and Defence Policy.[51]
Problems in relation to right of access are particularly more present on a national level. Instructions for officials of the national authorities are not clear. Confidential or EU secret documents are sometimes circulating among officials of a ministerial department because officials don't know what EU classification means.[52]
Many new disclosure laws in the new democracies in Eastern Europe included protections for security organizations. This situation refers to the culture of secrecy in the years of communism and it will take more than one generation to end such a tradition. Therefore training public officials how to handle laws on how to access public information needs attention.
The events of 9/11 and the subsequent war on terror did not lead to visibly removing public information from websites in the countries of EU members, as was the case in the US. However the implementation of anti-terrorist laws and the changes to domestic intelligence structures and practices did have an impact[53] on the right of privacy. Questions about the powers domestic intelligence agencies should have in relation to civil liberties in democratic countries remain contested.
New laws were introduced like the Patriot Act in the US and the Anti-Terrorism Bill in the UK. In the post-era of 9/11 each country in the EU has approached the security problem in its

own tradition and political culture, as defined by its historical context. Where Germany and the UK were rather slow to recognize the threat of terrorism, the French experience in the 1980s and 1990s helped them to take effective measures. However the French system also received a great deal of criticism because of the use of unlawful preemptive arrests and the disregard of civil rights. France did not reform it intelligence agencies. By its centralized approach it can offer strict surveillance programs. In Germany on the other hand the intelligence structure was highly decentralized. The different political entities have each their intelligence agencies. This makes the structure and communication between stakeholders very complex. There is some improvement on a federal level also in the context of EU security policy.

Conclusions

In the present situation the issues of oversight and accountability are a necessity in the context of an increasing growth of security organizations and domestic intelligence in the war against terror to protect the public interest. Public access to governmental grey information resources and the freedom of information are vital for a democracy. Developments are hopeful in regard to the number of countries providing public access to governmental information resources. That is just only a part of the story. For a balanced approach it is essential that the need for secrecy is to be the counterweighted by involvement of public scrutiny. The 9/11 information blackout has showed that culture of secrecy is undermining the values of democracy. Also troubling is the fact that the public sometimes seems to be detached from the political reality. Fortunately there are a number of advocacy and watchdog groups attracting and stimulating awareness for the freedom of information and access to a wide variety of grey resources like energy data sources and document collections from legislative institutions like Congress, the EU and other national parliaments. In some cases advocacy groups produce their own copied document collections to replace the original ones in case of removal.

Sometimes it is necessary for watchdog groups to provide for grey document repositories themselves i.e. to support journalists in investigative activities, like the website of Investigative Reporters and Editors does.[54] The IRE website contains more than 23,000 investigative stories to support research. This type of information is difficult to retrieve elsewhere. A more radical approach in promoting freedom of information is a website like Wikileaks, which tries to protect "internal dissidents, whistleblowers, journalists and bloggers" who might be endangered by publishing sensitive information. Wikileaks[55] put this information online anonymously and offers "assistance to people of all nations who wish to reveal unethical behavior in their governments and corporations".[56] In total it harbors 1.2 million grey documents. All these initiatives make clear that there is a global need for transparency to pierce the veil.

---

[1] Steven Aftergood, The age of missing information. The Bush administration's campaign against openness, in: Slate Magazine, March 17, 2005. See: http://slate.msn.com/id/2114963/ (retrieved Oct 21, 2009).
[2] See: http://www.egovrt d2020.org/ (retrieved Oct. 21, 2009).
[3] Cristiano Codagnone and Maria A. Wimmer (Eds.), Roadmapping eGovernment Research. Visions and Measures towards Innovative Governments in 2020, Delft University of Technology, The Netherlands. Results from the EC-funded Project eGovRTD2020, IST-2004-027139. See: http://www.egovrtd2020.org/ (retrieved Oct. 22, 2009).a
[4] David Shenk (1997), Data Smog : Surviving the Info Glut. Technology Review, 100(4), p. 18-26.
[5] Dawes, S. S., Evolution and continuing challenges of e-governance. Public Administration Review, 68(6), p. 86–102.
[6] Gellman, R. (1996). Disintermediation and the internet. Government Information Quarterly, 13(1), pp. 1 – 8.
[7] Hsinchun Chen, Lawrence Brandt, Valerie Gregg (et. al.), Digital Government: E-Government Research, Case Studies, and Implementation, New York : Springer, 2007.
[8] David Shenk, Data smog, New York : Harper Collins, 1997.
[9] The website of US Department of Justice provides for a handbook for making a FOIA request. See http://www.usdoj.gov/oip/04_1.html. The European Union provides access to Council document subject to the conditions laid down in Regulation (EC) No 1049/2001 regarding public access to European Parliament, Council and Commission documents (Official Journal L 145/43 of 31.05.2001) and the specific provisions regarding public access to Council documents in Annex II to the Council's Rules of Procedure, as amended by Council Decision 2006/683/EC of 15 September 2006 (Official Journal L 285/47 of 16.10.2006).
[10] See http://www.un.org/documents/ga/res/1/ares1.htm (retrieved Oct. 21, 2009)
[11] Public Law 89-554, 80 Stat. 383; Amended 1996, 2002, 2007. See text at http://epic.org/open_gov/foia/us_foia_act.html
[12] Department of the Air Force v. Rose, 425 U.S. 352, 361 (1976), See: http://supreme.justia.com/us/425/352/case.html (retrieved Nov. 7, 2009)
[13] Report of Conadep (National Commission on the Disappearance of Persons) – 1984, see: http://www.nuncamas.org/english/library/nevagain/nevagain_000.htm (retrieved Nov. 7, 2009)
[14] Pamela Constable, Arturo Valezuela, Chili's return to democracy, in: Foreign Affairs, vol. 68, nr. 5, 1989, pp. 169-186.
[15] Peter Lambert, "Paraguay" in Countries at the Crossroads: A Survey of Democratic Governance, Freedom House, 2007, pp.467-488.

[16] Not long after the fall of the Berlin Wall the archives of the former Secret Service of the German Democratic Republic were opened. See: http://tinyurl.com/y9w859s (retrieved Nov. 7, 2009)
[17] The Cold War International Project, Collection: The Mitrokhin Archive, see: http://tinyurl.com/yhcjjk4 (retrieved Oct. 21, 2009)
[18] "U.S. to Declassify Secrets at Age 25" from New York Times, December 21, 2006. See: http://www.nytimes.com/2006/12/21/washington/21declassify.html (retrieved Nov. 7, 2009)
[19] Nazi War Crimes & Japanese Imperial Government Records Interagency Working Group. Final Report to the United States Congress, April 2007, see: http://www.archives.gov/iwg/reports/final-report-2007.html (retrieved Nov. 7, 2009)
[20] Kate Doyle, The End of Secrecy. U.S. National Security and the New Openess Movement, in: National Insecurity. U.S. Intelligence after the Cold War, ed. by Craig Eisendrath, Philadelphia : Temple University Press, 2000, p. 92-93.
[21] Adopted on 6 December 2004. See: http:/www.unhcrh.ch/hurricane/hurricane (retrieved Oct. 23, 2009)
[22] Claude Reyes et al v. Chile: Summary of Judgment, Inter-American Court of Human Rights. Judgement of September 19, 2006. Summary of the Ruling available at http://www.justiceinitiative.org/db/resource2/fs/?file_id=17527 (retrieved Oct. 22, 2009)
[23] "Freedom of Information Around the World 2006: A Global Survey of Access to Government Records Laws," by David Banisar. See: http://www.freedominfo.org/documents/global_survey2006.pdf (retrieved Oct. 21, 2009)
[24] Transparency and Silence: a Survey of Access to Information Laws and Practices in 14 Countries, see: http://www.soros.org/initiatives/justice/focus/foi/articles_publications/publications/transparency_20060928 (retrieved Oct. 24, 2009)
[25] National Commission on Terrorist Attacks upon the United States. (2004). The 9/11 Commission report: Final report of the National Commission on Terrorist Attacks upon the United States. New York: Norton, p. 157, 164, 245.
[26] A list of closed or revised information resources is available at http://www.ombwatch.org/node/182 (retrieved Oct. 21, 2009)
[27] The website of the group Investigative Reporters and Editors dedicated to improving the quality of investigative reporting provides access to a number of databases. See: http://www.ire.org/inthenews_archive/ (retrieved Nov. 7, 2009)
[28] See: http://www.fas.org/sgp/othergov/doe/lanl/index.html (retrieved Oct. 25, 2009)
[29] The Information Security Oversight Office (ISOO) is responsible to the US president for policy and oversight of the Government-wide security classification system and the National Industrial Security Program.
[30] See: http://www.archives.gov/isoo/reports/2008-annual-report.pdf (retrieved Oct. 25, 2009)
[31] Ibidem
[32] Formal Statement. J. William Leonard, Director, Information Security Oversight Office. National Archives and Records Administration before the Committee on Government Reform. Subcommittee on National Security, Emerging Threats, and International Relations U.S. House of Representatives, March 2, 2005. See: http://www.fas.org/sgp/congress/2005/030205leonard.html (retrieved Oct., 26 2009)
[33] Formal Statement, J. William Leonard, Director, Information Security Oversight Office. National Archives and Records Administration before the Committee of Government Reform. Subcommittee on National Security, Emerging Threats, and International Relations. US House of Representatives, August 24, 2004. See: http://www.fas.org/sgp/congress/2004/082404leonard.html (retrieved Oct. 25, 2007)
[34] Scott Shane, Since 2001, Sharp Increase in the Number of Documents Classified by the Government, N.Y. TIMES, July 3, 2005
[35] 108 Cong. Rec. S9714 (2004) (statement of Sen. Wyden).
[36] http://www.doi.gov/foia/foia.pdf (retrieved Oct. 27, 2009)
[37] Attorney General Reno's FOIA Memorandum http://www.justice.gov/oip/foia_updates/Vol_XIV_3/page3.htm (retrieved Nov. 10, 2009)
[38] U.S. General Accounting Office, Report to the Ranking Minority Member, Committee on the Judiciary, U.S. Senate, Freedom of Information Act: Agency Views on Changes Resulting from New Administration Policy, GAO No. 03-981 (2003), see http://www.gao.gov/new.items/d03981.pdf (retrieved Nov. 3, 2009)
[39] Summary for Annual FOIA Reports for Year 2003, FOAI POST (2004), see: http://www.usdoj.gov/oip/foiaost/2004foiapost22.htm (retrieved Nov. 3, 2009)
[40] GAO-02-493 Electronic Freedom of Information Act Amendments, Augustus 2002, Washington DC, p. 54
[41] The Public Citizen, a consumer advocacy organization, examined statistically the number of days needed to process requests from the years 2000 to 2004, see: http://tinyurl.com/yfblyya (retrieved Oct. 29, 2009)
[42] Executive Order 13392 of December 14, 2005, Improving Agency Disclosure of Information, see: http://edocket.access.gpo.gov/2005/pdf/05-24255.pdf (retrieved Oct. 28, 2009)
[43] The National Security Archive is an independent non-governmental research institute and library located at The George Washington University in Washington, D.C. The Archive collects and publishes declassified documents acquired through the Freedom of Information Act (FOIA). See: http://www.gwu.edu/~nsarchiv/ (retrieved Nov. 8, 2009)
[44] 40 years of FOIA, 20 Years of Delay: Oldest Pending FOIA requests. Date Back to the 1980s, The National Security Archive (George Washington University), The Knight Open Government Survey, Washington DC, 2 July 2007, see: http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB224/index.htm (retrieved Oct. 21, 2009)
[45] Ibidem
[46] The Ashcroft Memo: "Drastic" Change Or "More Thunder Than Lightning"? The National Security Archive, Freedom of Information Act Audit. Phase One Presented March 14, 2003 / 2003 National Freedom of Information Day Conference. See: http://tinyurl.com/yjdw9zw (retrieved Oct. 21, 2009)
[47] John Palmer, Danish Plan to Go Public and Lift Plans of Secrecy", The Guardian, December 30, 1992, 6.
[48] Regulation (EC) No 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission documents.
[49] Decision 2001/264/EC
[50] Council Decision of 19 March 2001 adopting the Council's Security Regulations (2001/264/EC), see: http://www.eulex-kosovo.eu/training/security/docs/COUNCIL_DECISION_sec.pdf (retrieved Oct. 24, 2009)
[51] Annual Report of the Council on the Implementation of Regulation No. 1049/2001 of the European Parliament and of

the Council of 30 May 2001 Regarding Public Access to European Parliament, Council and Commission Documents. Available years: 2002 – 2008, see: http://www.consilium.europa.eu/showPage.aspx?id=305&lang=en (retrieved Oct. 25, 2009)

[52] For example several confidential documents have been discovered on the website of the Icelandic government providing insight how European governments handled the financial crisis which led to the economic collapse of Iceland. See: http://www.google.com/#hl=en&safe=off&q=site:island.is/media/frettir/&start=0&sa=N&fp=1&cad=b For a sample see: www.island.is/media/frettir/31.pdf (retrieved Nov. 10, 2009)

[53] Aidan Kirby, Domestic Intelligence Agencies After September 11, 2001: How Five Nations Have Grappled with he Evolving Threat, in: Considering the creation of a domestic intelligence agency in the United States: lessons from experiences of Australia, Canada, France, Germany, and the United Kingdom/ Brian A. Jackson, editor; Rand Corporation, Santa Monica, CA, 2009, pp. 143-169.

[54] IRE Resource Center, see: http://www.ire.org/resourcecenter/ (retrieved Nov. 10, 2009)

[55] See: http://www.wikileaks.org (retrieved Nov. 10, 2009)

[56] Ibidem

Documenting an Environmental Disaster: The River Valley Collection at Marion Public Library

Ian Fairclough, George Mason University, United States[1]

Abstract

Marion Public Library (located in Marion, Ohio, 45 miles north of Columbus) serves as repository for a collection of documents produced in response to local concern about the incidence of leukemia and other ailments among the population. Former athletes at the River Valley High School, several miles east of town, were particularly at risk. Inquiries revealed that the school was built on a former WW II military site, and that the school's ballpark was above a dump site for noxious chemicals.

From being considered a relatively minor issue, local outrage grew until something was done. The school was relocated and the old buildings closed. Eventually the U.S. Army Corps of Engineers, and its contracting agency Montgomery Watson, cleaned up the site.

One consequence was the production of voluminous documentation, requiring the attention of a Documents Librarian and a Catalog Librarian. The resulting repository is of interest to scholars, and the library has received a surprising number of requests via interlibrary loan. Such requests, made possible via appropriate bibliographic description and access, demonstrate that a collection of primarily local interest can serve a wider community, provided that proper attention is given to its maintenance.

Of particular interest to scholars of grey literature is the issue that arose during the course of the documentation project of the partial transition from a paper-based paradigm to electronic resources. Initially, documentation was generated, in most cases using electronic technology, with a view to the paper version as end product. The realization that documentation in an electronic medium can serve as the objective happened during the course of the project, but was not fully realized. Indeed, the persons who generated the documentation are not known to have expressed the transition from paper to electronic version as a concept. As a consequence, further work entailing retrieval of the electronic originals where still available, and document scanning where not, remains a possibility. Such work will likely fall to librarians; it is unlikely that the original authors will assume responsibility.

The Social and Historical Context of the River Valley Collection

Surely no material is more obscure, more "grey," than what is classified as secret! Perhaps had secrecy not been a primary concern of the U.S. Government during and after World War II, Marion and its surrounding communities might have avoided the tragedy that ensued when the River Valley High and Middle Schools were constructed atop the dump site of the Formerly Used Defense Site known as Marion Engineer Depot (MED). Hindsight is easy, however. Another way of phrasing the question is, "Who could have known?" Even in 1988 with the publication of Charles Mosher's compilation[2] of anecdotes and newspaper cuttings from the local paper, the Marion Star, given the benefit of four decades of hindsight little was uncovered. The existence of the dump site was either unknown, ignored or disregarded during the process of acquiring the tainted land for educational purposes. And no mention of leukemia is found in Mosher's anthology.

Two formerly used defense sites are located in the Marion area: Scioto Ordnance Plant (SOP) and Marion Engineer Depot. Both of them are covered in the extensive documentation known as the River Valley Collection (RVC) at Marion Public Library (MPL), although only MED was implicated in the leukemia tragedies. The investigation and cleanup activities of the 1990s onwards took place - and continue - in both locations. SOP was by far the larger - and the more dangerous to work in during the war, for it produced explosives destined for use against enemy powers. SOP was constructed not long after Pearl Harbor and employed thousands of workers, however production stopped after just one year of operation. After VJ day the SOP site was deactivated and decontaminated, then parcels of land were sold off for agricultural and other uses, including a prison, an airport, and a housing subdivision. A historical marker at the airport documents the relocation of area residents as their land was expropriated via eminent domain.

Some buildings constructed during the war were converted to other uses. Other buildings from pre-war days had never been demolished and reverted to their previous usage. Notable is

Likens Church, a prominent local landmark since the 19th century. Among the hearsay documented in Mosher's book, Likens Church was rumored to have had, at one time, uranium stored in the basement[3]. Uranium storage was however unsubstantiated, and investigations including radiological tests provided no evidence. But the Atomic Energy Commission did have a minor role in SOP's future, with Monsanto labs occupying a small part of the SOP property for a while.

Marion Engineer Depot meanwhile continued in active service until 1960. It adjoined a long-distance highway (formerly U.S. 30 South, now state route 309) and a major east-west railroad. After going out of service, most of it was transformed into the Marion Industrial Center. The River Valley School District acquired a small tract of the land, upon which the now disused school buildings were constructed. During its occupancy by MED, some of this land had been used as a prisoner-of-war (POW) camp, Camp Marion, the residence of Germans captured in North Africa after the allied overthrow of the Axis powers there. The POWs were occupied in activities lawful under the terms of the Geneva convention[4]. Again, secrecy was an issue: Fraternization of locals with POWs was, in a word, verboten.

No exact picture emerges, from either official documentation or the competent journalistic reporting of area newspapers, of just how or when it became known among some sufferers of leukemia that they shared a common bond as graduates of the River Valley schools. Perhaps it was the production of Mosher's book - to this day the only published account of SOP and MED - that initiated and stimulated discussion. And since the research was inconclusive, neither establishing nor refuting an association between the schools, the former dump site and the patients, locals continue to argue amongst themselves - vehemently at times[5]. Whether or not the controversy was a contributing factor, the decision was to "err on the side of caution." It was decided to take action, close the schools, and clean up - again (for archival reports contain memoranda indicating that decontamination did take place in the years immediately following the war).

The Marion Public Library was a logical location for the repository, given the close connection between the repository and the community. Strictly speaking it is a "city-school district" library, serving the city of Marion and its associated school district. Unofficially (that is, without dedicated funding from the county), it also functions as a library for the whole of Marion County. Other options that might have been chosen are the Marion County Historical Society and the Ohio Historical Society (OHS). Some related documentation is available via OHS. But the choice of the library was a good one, because it has greater ease of access in terms of both location and hours of service. It also has the advantage of substantial technical support, as well as ample shelf space in the Reference area.

Decisions in Document Processing

Beginning in 2003 Fairclough began to process donated materials that had accumulated during the past several years. Dawn McCleery, at that time Adult Services Librarian with responsibility for the River Valley materials, served as contact person for negotiation and receipt, and together with Fairclough made a set of general decisions for processing to be applied to the whole collection. These decisions were:

- To use full cataloging according to Anglo-American Catalog rules, 2nd ed., revised (AARC2R)
- To assign Library of Congress subject headings
- To provide a full Dewey Decimal Classification (DDC) number
- To prefix the DDC number with notation specially for the RVC
- To provide a series added entry, River Valley collection, in each local record
- To provide local subject added entries for the high and middle schools respectively
- To make the bibliographic records available internationally using OCLC WorldCat.

The decisions are far-reaching. Their importance is best illustrated using the alternatives as points of comparison. AACR2R cataloging mandates a uniformity of treatment for materials, in particular in the choice of the source of information from which the record is prepared. Most materials had a title page, or suitable substitute, with a title that the document author had prepared. In contrast, non-AACR2R cataloging might have been derived from an unknown and unspecified source, or the cataloger might have invented the data. Such invention is permissible under circumstance preventing use of the prescribed source; when done, AACR2R requires the cataloger to enclose the supplied information in brackets. Furthermore, full cataloging is contrasted with the preparation of brief records, which might have little

information other than a provided title. Such records, regrettably, abound in the bibliographic universe. Marion Public Library, through assigning the task to a professional cataloger, avoided the pitfalls that inadequate documentation can create, such as the inability to properly and uniquely match a bibliographic record in the catalog with an item in hand.

Use of Dewey Decimal Classification has many benefits. It allows the records to be treated on a par with other materials in the collection: For nonfiction and reference materials, MPL is a DDC-classed library. It facilitates shelving of like materials - known popularly if not totally accurately as the primary use in public libraries. A better understanding is available in terms of scholarly organization. Researchers can use the DDC to retrieve information to a high degree of precision - again, provided that the job is done properly, in terms of the full number available built according to the DDC schedules and tables. Creation of this number is a professional job, not to be undertaken lightly. In contrast, many libraries content themselves with a modified DDC number, usually by truncating after an arbitrary number of digits. Pressure from circulation staff can even be a deciding factor in such a policy!

Table 1 is a conspectus of DDC numbers used with the RVC, as assigned locally. It illustrates concisely what the collection is about.

Table 1: Dewey Decimal Numbers in the River Valley Collection

363.17	Hazardous materials
363.1763	Monitoring, surveillance, reporting
363.179	Specific types of hazardous materials
363.1791	Toxic chemicals
363.1799	Radioactive materials
363.7	Environmental problems
363.73	Pollution
363.735	Social action (class here remedial actions)
363.7356	Government action
363.7358	Community action
363.736	Control
363.7363	Monitoring, surveillance, reporting
363.7364	Inspection and testing
363.737	Measures to prevent, protect against, limit effects of pollution
363.738	Pollutants
363.739	Pollution of specific environments
363.7392	Air pollution
363.7394	Water pollution
363.7396	Soil pollution
363.73966	Control (elimination and reduction of hazards, of sources and causes of difficulty)
363.739664	Inspection and testing
616.99419	Leukemia
658.4083	Social responsibility of executive management - protection of environment (class interdisciplinary works in 363.7)
917.71514	Marion County (used for maps, etc.)

One additional element in the call number, the prefix, allows for separate shelving from the rest of the collection. It contains two parts, the letter R to indicate it is part of the Reference collection, followed on the next line by RIV VALL to uniquely designate the RVC. Depending on the online system in use locally, prefix data can be included in the filing order (as at present) or omitted and processed separately (the result being to display other materials the library holds in the same subject area). Separate processing is advantageous but regrettably not available in MPL's current integrated library system. Without such notation materials in the collection would be interspersed with other works on the same subject but not referring to the River Valley situation - although other materials on that subject having to do with Marion County might be accommodated by use of the full DDC number.

By way of illustration of the structure of a DDC number: This title, *Soil, Air and Particulate Monitoring Data Package : River Valley School Site Investigation, Former Marion Engineer Depot, Marion, Ohio*, has a number built thus:

- Base number for Hazardous materials[6]: 363.17
- Further notation for " Monitoring, surveillance, reporting" added to base number 363.17 as instructed there: 63
- Notation to indicate subsequent addition of a geographic subdivision: 09
- Geographic number for Marion County: 771514
- Built number: 363.176309771514

This number was retained in the OCLC WorldCat master record for the benefit of all users of that database. Locally, notation for Marion County was removed as redundant and the prefix was added to form: R RIV VALL 363.1763

By browsing the local catalog, using the search key R RIV VALL, one can retrieve a full list of all titles in the RVC, arranged in order of the subjects represented in the DDC scheme. If the decision had gone differently, that is, not to use DDC numbering but simply to lump all materials together, no such ordering would be available.

We further decided that another unifying feature, identifying locally all documents in the collection, would best be served by the presence in each bibliographic record of a searchable data field. In order to be distinct from all other fields in the records, and to avoid confusion with data actually present in the items, we used a series added entry for this purpose. A user of the MPL catalog[7] can thus search for the series "River Valley collection" and retrieve a list of all titles in the collection. Properly - and this depends again on the display capabilities of a local system - this display will be in alphabetical order by main entry (that is usually for these materials the document title). This added entry provides a different perspective to the aforementioned browse by DDC number. It is provided locally, since the "River Valley collection" is a locally created entity rather than a published series. A local researcher wishing to know if a particular document is in the collection but unsure of the exact wording of the title can retrieve the records and peruse them. Had the decision not been made to include this field, no such capability would have been built into the records.

Local subject added entries were also created and added in most cases to the records. Since MPL is not a contributor to the Name Authority Cooperative Project (NACO)[8] and no other agency had already provided a heading in a name authority record for either the River Valley High School or the River Valley Middle School, these names were used as the bases for the headings in the records. In assigning these subject added entries, we followed a fairly liberal application. Strictly speaking the documents for the Scioto Ordnance Plant - and even some from Marion Engineer Depot - had no direct connection with the future schools. But in order to facilitate access to the documents locally, the subject headings were provided in local records. So someone searching in the MPL catalog for the schools and not already aware of the existence of the RVC will, it is hoped, discover the collection. Again, had no such decision been made, the capability would have been omitted.

The last decision listed, to make the bibliographic records available using OCLC WorldCat, was not the last one made, but is the most significant in terms of bringing awareness of the materials to a wider audience. Many institutions collecting such materials might decide that they are purely of local interest, and catalog them only using their local system. Had MPL done so, it is a foregone conclusion that the library would have received no requests for these materials outside our local community. But scholars working on environmental concerns elsewhere in the United States have discovered MPL's resources, and librarians have requested selected materials.[9]

A corollary to using WorldCat is that the records were cataloged using the *MARC21 Format for Bibliographic Data*, as implemented in WorldCat and locally. This was not so much a decision in itself as a consequence of having the records in WorldCat and in the MPL local catalog. Not only is use of MARC required for the records to be included in the MPL catalog, it also has more granularity than some of the other metadata options, such as Dublin Core, currently being applied to institutional repositories. If encoding in another metadata scheme should prove necessary, other agencies with such requirements can use a data crosswalk scheme with an appropriate processing tool such as MarcEdit to get the bibliographic record into the locally preferred format.

In part the use of MARC also arose from the fact that the bulk of the collection at the time processing began was in print format - and remains so today although some of the materials are now electronic. Within the profession of librarianship and in scholarship in general, the phrase "institutional repository" is in transition: Increasingly, though that phrase does not necessarily signify that it is in electronic format, it has that connotation. A similar semantic shift took place with the word "database" a generation ago. But the RVC is truly a repository although only part of its contents is electronic.

Telling the Story Through the Documents

The RVC consists of materials occupying approximately ten shelves of the MPL collection. An exact figure of the number of items is unavailable: In fact, documentation is still in process, and the collection continues to grow. Some bibliographic records, such as those of the agendas and transcripts of the hearings of the Restoration Advisory Board (RAB), cover several items.

Thanks to the judicious choice of wording for the titles of documents that Montgomery Watson (MWH)[10] produced, we can review the collection by title and get a pretty good overview of the activity of the whole project. Here are selected titles, with the dates of production. Initial investigations took place between 1994 and 1997:

- Ordnance and explosive waste, archives search report for the former Scioto Ordnance Plant, Marion, Ohio (1994)
- Final limited site investigation report for radiological contamination at the former Scioto Ordnance Plant, Marion, Ohio (1995)
- Documents prepared for preliminary investigation of former Marion Engineer Depot site, River Valley schools, Marion County, Ohio (1997)
- River Valley High School, Marion, OH, sampling … in response to public concerns about leukemia and cancer incidence. (1997)
- Environmental investigation at the River Valley local school property : final report (1997)
- Work plan for River Valley local schools (1997)

It was at about this time (the late 1990s) that matters began to escalate. Whereas it appeared that local concern would soon be resolved, the reverse happened. A four-year study led to a conclusion that no significant association could be made between the incidence of leukemia and the River Valley schools. But far from quieting people's opinions, they became further inflamed. After a lull, documentation began again in earnest with titles such as:

- Marion Engineer Depot and Scioto Ordnance Plant Restoration Advisory Board agenda. Transcripts provided by Hammond Reporting Services. (2002)
- Soil, air and particulate monitoring data package : River Valley school site investigation, former Marion Engineer Depot, Marion, Ohio (2001)
- Draft revisions to the River Valley school property, operable unit 1 - former disposal area, draft remedial investigative report, Marion, Ohio. (2001)
- Analytical report for 204087 for DLZ Laboratories, Inc. Accompanying CD-ROM has text of base report (file: lims.pdf) plus other files. (2002)
- Public hearing held at the ... Marion, Ohio on ... at ... / Marion Engineer Depot, Scioto Ordnance Plant and Army Reserve Local Training Area Restoration Advisory Board. Transcripts provided by Florine I. Varner & Associates, Inc. (2003-)

It is interesting that the RAB documentation is based on transcripts. The original sound recordings, produced only to facilitate keeping an accurate if not necessarily verbatim written record, are of interest for a variety of reasons, some quite tangential (a study of regional accents and pronunciation, for example, can draw from it). Whether or not the sound recordings continue to exist is not immediately known. The laboratory report is an early example in the collection of the provision of information in CD-ROM format: Much of what was documented in supplementary files is data that would require many reams of paper to print, and which would be unusable in that format. The term "draft" is a recurrent one within the collection, as is "draft final." In fact it is the norm rather than the exception for the most recently deposited version of a document to be designated as "draft." It seems that in many cases the authors lacked the nerve to assert that a document is "final." An exception which "proves the rule" is the radiological report of 1995, which illustrates the notion prevalent at that time that the project was a small matter and would soon be over. In the same vein and as another side of the same coin: When documenting laboratory reports and the like, a "chain of custody" is recorded as security against tampering with specimens.

Document Formats

In the first instance, the agencies responsible for the clean-up conducted archival studies, made available in 1994 and 1998, to discover what is now available about the establishment of SOP and MED. These studies, drawing upon resources much of which were formerly confidential but now have been declassified, contain copies of material retrieved from archival sources. Whether or not the archival search was exhaustive, and successfully so, the benefit to users is substantial. One does not have to visit the National Archives in order to access at least some of the information: It can be found in the Marion Public Library. Much of it is

photocopied from the original, and the copy that winds up in the binders on the shelves at MPL might be a third generation copy if not even more remote from the original - which was itself perhaps a carbon copy. These materials, then, constituted the initial foundation upon which the research into the local situation during the 1990s and 2000s was based. From half a century ago, materials whose present-day use in this context was not likely to have been foreseen emerged as the beginning of the story that the River Valley Collection tells.

It is not the objective of this essay to delve in detail into the document texts themselves. That project awaits the attention of a historian, and it is hoped that such a person will assume interest. This essay marks the status of the documents as grey literature and their migration from archival material to a local repository.

It is the various circumstances under which the River Valley Collection documents were produced that give it such a range of characteristics. The archival documents, created long before the electronic era, likely were dictated: An army officer - who probably never touched a typewriter - would have his secretary take a memo. Such a process constituted transcription of an oral message formulated according to stylistic conventions signifying rank, sub- and superordination, and so forth. By the 1990s this hierarchical situation was likely long gone, as all military and commercial ranks of personnel adopted the keyboard as the primary means of written communication. The secretary's supporting role was replaced by that of information technologists whose task in maintaining word processing software, keyboards, monitors, printers, and the central processing unit functioning correctly and in a timely fashion was critical to the proper functioning of the organization. Given this infrastructure, and the availability of devices such as templates to expedite memo writing, the preparation of the written document underwent a significant paradigm shift. Yet it was still the preparation of the print product that was paramount in an author's thought. It took considerable time before transmission of a document could be effected otherwise than in the paper medium. The word printout is significant here: It indicates the beginning of the thought process that the electronic document is primary. To my knowledge, none of the documentation in the River Valley Collection was considered to be a "printout" until quite recently.

Montgomery Watson were exemplary in their production of documentation for the U.S. Army Corps of Engineers. It is not known at this time whether MWH have retained the original files: Whether they were retained on the original hard drives, moved or copied to inhouse archival storage, or other steps taken. One hopes that they were preserved, and moreover that the files can eventually be made available for public use. The company might regard their obligation to the public as fulfilled by provision of the paper copies, and decline to participate further. Or they might view the additional steps required to make documents available as an act of *pro bono* community service that is in the company's interests. All these ideas are avenues for possible pursuit.

It was during the middle of the first decade of the 2000s that MWH's documentation in print format began to be submitted with an accompanying CD-ROM. Whether the CD-ROM was created with the text document primarily in mind or secondarily is a question. An example is this title, from 2006: *Closure Report, SOP-F : Artillery Fuse Load Line F4, Underground Storage Tank Removal Program, Former Scioto Ordnance Plant, Marion, Ohio*. A bibliographic note reads "CD-ROM contains full text of document plus appendices of analyses by Kemron Environmental Services, etc." Thus the question is whether the agents supplying this material were doing so primarily in order to transmit the appendices, which contain data not suitable for reproduction in print format. If so, then the actual text of the document might have been provided as an after-thought, as if to say, "Since we're providing the analyses in a CD-ROM we'll also provide the text."

Three years later (2009) this document appeared:

Former Scioto Ordnance Plant, Marion, Ohio [electronic resource] : Engineering Evaluation/Cost Analysis : Final Report for the Burning Field (SOP-M) and the Incendiary Fuel Disposal Area (SOP-Z).

From the general material designation "electronic resource" presented after the title proper, one knows immediately that the record describes an item not in paper format. Later in the record it is stated that the item is a CD-ROM. When printed out, the text comes to approximately 85 pages (irregular pagination prevents an exact statement). Tables, attachments, appendices, and photographs bring the total to approx. 200 pages.

The ultimate irony is that such documents are indeed printed out! McCleery herself makes a printout and places it in a binder. That such a practice is expected is a feature of the document, for it contains a cover sheet, intended for insertion in the outside pocket of the binder. That the practice is done at MPL stems from two concerns: Usefulness to the community, and preservation of the material. MPL users of the RVC want to come into the library and peruse the documents. Having to use a computer to do so would be, for them, an inconvenience. Plus, it would be very easy for the CD-ROM to be (illegally) removed from the collection.

The next stage of development for such documentation is to use online electronic media for storage and transmission. It was noted that scanning might take place. Judging from the quality of the documents from the 1940s it is unlikely that a scanning project would successfully transcribe these texts at first pass and without human intervention. It is likely to be a labor-intensive task. In mentioning quality, we intend not the historical worth of the texts, but their typographical and photoreproduced condition. Whether scanning of more recent documents is necessary remains to be determined. If the cooperation of MWH and other agencies can be engaged, those developing and maintaining the collection can avoid scanning. Perhaps another agency will collaborate with MPL to serve as a host site for the electronic materials, providing access via the Internet. The Ohio Historical Society already has an interest in the subject, and is a possible ally.

Conclusion

The "official" history of World War II, as written from the perspective of the United States army, is recorded in an extensive series[11] of books published in the post-war years, many of which have been reprinted in the early 21st century. One might think that the title, *The Ordnance Department: Procurement and Supply*[12], a volume of over 500 pages, would contain information about the Scioto Ordnance Plant. It does - a one-word mention (p. 218) documenting its closing in 1943. This paucity of information is not so much a presumption of the comparative insignificance of SOP as a reflection on the vastness of the scale of operations of World War II. This scale of operations is such that a site so large and productive as SOP is but a drop in the ocean. It is provocative that half a century later so much documentation has been generated on a matter that is considered so tiny in the grand scale of things.

What this documentation calls for is further attention. The documentation itself records the attention that was given during the site investigations, lab testings, plans of action, and reports of completion. Now it falls on the current generation of librarians and scholars respectively to preserve the documentation and to optimize its exploitation in order to contribute to the body of human understanding.

---

[1] Dawn McCleery (Head of Reference, Marion Public Library) advised on this presentation and related matters. Fairclough gratefully acknowledges McCleery's generous use of her time and support.

[2] Mosher, Charles D. The Scioto Ordnance Plant and the Marion Engineer Depot of Marion, Ohio : a profile after forty years. -- [Marion, Ohio] **:** C.D. and D.R. Mosher in cooperation with the Marion County (Ohio) Historical Society, c1987.

[3] Mosher, op. cit., p. 76.

[4] Mosher, op. cit., p. 133.

[5] A septuagenarian county resident once engaged me in such a conversation. I had to plead that, as a librarian and not native to the area, I simply do not know anything beyond what is recorded!

[6] In DDC number building the "base number" is *not* the three numbers to the left of the decimal point, but rather, all numbers that together have a particular significance and to which the classifier can add others to bring out additional facets.

[7] http://www.marion.lib.oh.us

[8] http://www.loc.gov/catdir/pcc/naco/naco.html

[9] Since RVC materials do not circulate outside MPL, when loaning them it is stipulated that the borrowing library must also restrict their use to on the premises.

10 During the course of the project, Montgomery Watson, an engineering company specializing in wetlands infrastructure, changed its name to Montgomery Watson Harza as the result of a merger, and then began also using the form MWH Americas, Inc. Inconsistency in company representatives' application of these names in the documentation has resulted in chronological overlap in their application in bibliographic records. Their website http://www.mw.com/ simply uses MWH.

[11] Stetson Conn, general editor. *United States Army in World War II. The Technical Services*.

[12] Washington, D.C. : Center of Military History, United States Army, 1960, reprinted 2003.

MOSAIC: Shades of Grey

Keith G. Jeffery, Science and Technology Facilities Council, United Kingdom
and Anne Asserson, University of Bergen, Norway

Abstract
Grey literature covers a wide range of artifacts. As indicated in a previous paper (Jeffery and Asserson 2004), the authors consider a grey literature repository in a research-based organisation to record the intellectual property of that organisation. However, this is only usable effectively if the repository contains the grey objects and the metadata is formalised (Jeffery 1999, Jeffery and Asserson 2004) or – better - stored and accessed in a CERIF-CRIS (Common European Research Information Format – Current Research Information System) (Jeffery and Asserson 2005). In this way the grey resource is available in the context of the work of the research organisation and/or its stakeholders managing research strategy, research evaluation, funding and cost-accounting, innovation and knowledge transfer and public information (Jeffery and Asserson 2005). This was further refined as 'Greyscape' (Jeffery and Asserson 2007) and the technologies for interoperation surveyed (Jeffery and Asserson 2008). A suggestion for using advanced hyperactive objects for research output workflow linked to a grey repository (Jeffery and Asserson 2006c) was set aside to await later reconsideration.

The key messages are:
1. *conventional grey literature repository metadata (usually based on DC (Dublin Core)) is insufficient;*
2. *great advantages are achieved when a grey repository is linked to a CERIF-CRIS: contextual metadata, workflow, interoperation and organisational integration;*

The adoption of such an architecture ensures:
- *a lower effort threshold on input (workflow, formal metadata) and hence increased repository fill;*
- *improved retrieval (formal metadata, semantic links);*
- *improved support for workflow and the research process (formal metadata, semantic links);*
- *improved links to other within-organisation systems (formal metadata, semantic links) including repositories of research datasets and software, library catalogs and systems for finance, HR, project management, directories, web-pages;*
- *improved interoperation with the systems of other organisations (formal metadata, semantic links);*

Mosaic was the original graphical user interface web browser. It provided a new way of accessing information – although it required considerable human effort/time to browse and click on links.
A CERIF-CRIS provides a new way to access and utilise grey information but encourages the user to let the computer system do the tedious work leaving the end-user free to do their research. This is achieved by the use of the formal syntax and defined semantics of CERIF entities, attributes and linkage together with as much or as little intelligence in the system as the user requires. The information space (the grey information mosaic) is navigated reliably and reproducibly by the computer, not the user. Presented through a web browser (the viewer mosaic – perhaps based on Mosaic) the ease of use remains but with a much more powerful information management capability.

Background

Mosaics

A mosaic is a form of decoration utilising small pieces of natural material (e.g. stone) or artificial material (e.g. glass) to form artistic patterns. Mosaics are usually found as floor, wall or ceiling decoration. The best mosaics have the pieces ordered by colour or greyshade into geometric forms. Thus like pieces are composed together to form a representation of an abstract geometric form (commonly squares, 'greek key pattern' swastikas…) or a form representing observed (e.g. the famous dolphin mosaic at Delos) or imagined (e.g. representations of God in Orthodox churches) nature. The key points about mosaics are (a) they are designed; (b) they have a formal structure; (c) the formal structure consists of small pieces composed into structures and those structures composed into larger structures; (d) they represent something in the human mind; (e) they communicate the idea in the human mind from originator to others; (f) they take considerable effort to construct.

Grey Literature
A collection of grey literature (or more generally grey objects) may be likened to a mosaic since the pieces represent individual objects and are grouped into structures statically or dynamically depending on the purpose. A grey literature collection exhibits the characteristics of a mosaic outlined above and can be extended to the concept of a dynamic mosaic – more like a kaleidoscope – as different patterns of collection of grey objects emerge depending on the query or classification aspect..

To understand a mosaic requires more contextual knowledge about the civilisation (including philosophy and system of belief) within which it was constructed. Much of this contextual knowledge is not recorded. With the posit that the grey literature collection is like a mosaic, then grey literature can only be truly appreciated with the contextual knowledge. The hypothesis is that this contextual knowledge (and indeed better metadata describing the grey literature itself) is provided by a CERIF-CRIS.

The Hypothesis
The hypothesis of this paper is that a grey literature collection is much better collected, structured, catalogued, utilised and maintained within the context of a research environment (commonly known as e-Research or e-Science) (Jeffery 2004a), (Jeffery 2004b) which relies on CERIF-CRIS to provide improved metadata for each GL object as well as contextual research information to other recorded research information thus improving the integration and publicising of grey within the research scene. The key is improved data collection, improved interoperation and improved query relevance and recall, all based on the formal syntax and declared semantics of a CERIF-CRIS.

Proposed Architecture

Objects, Data and Metadata
Metadata (Jeffery 2000) is data about data. Usually metadata describes objects. However, information that is metadata to one application is data to another. Take the library catalogue card within a computer system; it is used as metadata to find a particular book or publication; but as data if the librarian wants to take stock of the collection or answer queries concerning the number of books on e.g. grey literature.

Metadata is commonly machine-readable; that is metadata in one system can be read by another system and a simple (text-based) search can be done across systems. This is achieved by various protocols; in the WWW (World-Wide-Web) environment commonly by using Dublin Core (DC) for description, (OAI-PMH) (Open Archives Initiative Protocol for Metadata Harvesting) for file exchange and (OAISTER) for cross-system searching. A similar set occurs in the more conventional library catalogue world based on (MARC) using (Z39.50) for homogeneous retrieval over heterogeneous sources and many other sets exist in different research or commercial areas.

However, if the metadata is richer with formal syntax (structure) and declared semantics (meaning) then it is not only machine-readable - with its limitations in relevance and recall related to the limited description and simple query support - but also machine-understandable. This means that the computer system can do most of the work – providing acceptable recall and relevance due to the improved description and greater query capability supported - instead of the end-user having manually to refine the search to improve recall and relevance by browsing a screen of 'hits'.

A CERIF-CRIS provides a much richer metadata environment and a much more powerful query environment for grey literature utilising the computer system intelligently. Furthermore, a CERIF-CRIS relates the grey literature together with research context data to give the end-user a much richer contextual experience as well as assisting in ensuring improved relevance and recall.

Requirement
There are ever-increasing requirements for access to and processing of grey objects.

The first requirement is to capture and input the grey object with associated metadata; the metadata being required for later retrieval and processing. The largest problem is the threshold barrier faced by the end-user (usually author) in inputting the metadata; faced with screen after

screen demanding input the user gives up, or inputs hastily and incorrectly. The solution is to use pre-existing information in the system (a) to pre-fill form fields for the end-user and (b) to validate the input to improve accuracy. Both are much more effective if the pre-existing data is stored in a CERIF-CRIS with its inherent formalism and structure (syntax) and declared semantics. Furthermore, if the input process is part of an organisational workflow and thus progressive, the input is demanded in 'bite sized chunks' which facilitates input quality and timeliness. (Jeffery and Asserson 2005), (Jeffery and Asserson 2006b).

The second requirement is retrieval. Retrieval may be likened to finding a needle in a haystack or – more appropriately in today's internet world – several needles in several different haystacks. This has implications if rapid, accurate (recall and relevance) retrieval and subsequent processing is to be provided:

1. the grey objects must be described by metadata with formal syntax and declared semantics that are sufficiently rich to describe (to enhance location) and proscribe (to ensure access restrictions, rights and costs are observed) the object. The semantics must be supported by adequate domain metadata such as dictionaries, thesauri and domain ontologies;
2. there should be homogeneous access as seen by the end-user over the heterogeneous sources from which the objects required are to be retrieved. This implies (to reduce n*(n-1) interfaces to n) agreement on a canonical schema for metadata. This schema is used:
    a) to frame the end-user query ready for transmission to the target sources;
    b) to convert from the end-user query to local query against the local schema
    c) to convert local results to the canonical form for transmission back to the user;
    d) combination of local results with the canonical results of queries to other sources in the final composed result presented to the end-user.

Of course, if all grey literature sources used the canonical schema as their storage schema then the conversions are unnecessary.

The third requirement is subsequent processing. Retrieval and reporting of the results may suffice, but increasingly end-users wish to know the number of relevant grey literature objects (count function), the average number of grey objects retrieved classified by some other attribute e.g. author or organisation (group by and average functions) and so on including graphical representations to represent the data or even video to show the change with time. For such outputs to be useful the base data must be (a) accurate (b) structured formally (i.e. no semi-structured or unstructured data). Clearly a CERIF-CRIS meets this requirement whereas a metadata-poor repository does not.

The fourth requirement is relating information retrieved from the grey literature environment with other information – both in the research context (projects, persons, organisations etc) but also more widely within a research organization (finance, human resources, project planning) or even outside the research organisation (government statistics on policy, social structures, budgets, industrial output, education…). The ability of a CERIF-CRIS to link outward to such other systems due to the linking relation mechanism is a pre-requisite for this to be done maintaining integrity and avoiding costly ad-hoc coding (Jeffery and Asserson 2006a).

The Architectural Solution

As indicated above, an ideal architecture means that

1. all grey sources adhere to the same canonical schema;
2. that schema has formal syntax and declared semantics;
3. data under the schema can be used as data for some purposes and as metadata for others;
4. the schema allows linking relations between entities with date/.time stamp and role such that the structure is articulated flexibly, new entities can be added and related and links to external systems can be made using the same framework;
5. the above provide an optimal base framework for the processing required including input within a progressive workflow, retrieval and reporting, subsequent processing including statistical and graphical reports and interlinking to other systems both within and outside of the research organisation.

The hypothesis of this paper – based on previous work on the schema (Jeffery 1999), (Asserson and Jeffery 2005), and the architecture (Jeffery 2005), (Jeffery and Asserson 2008) is that CERIF forms the ideal canonical schema and is suitable for the architectural purpose.

One solution is to store the grey objects in the CERIF-CRIS; this approach has been used in the design of the (PURE) system from Atira in Denmark. Another solution is to store the grey objects in a repository (such as an institutional repository powered by one of (ePrints), (DSpace), (Fedora) or (ePubs)) linked to the CERIF CRIS wherein the metadata is stored in the canonical form. The metadata may be replicated from the CERIF-CRIS into another form such

as DC in the repository to allow for OAISTER searching over OAI-PMH. A third solution is to convert the repository software and metadata schema to CERIF; this is being done for ePrints under the R4R (Ready for REF (Research Excellence Framework)) project in UK. However that project is becoming convinced that either the first or second solution is better, not least because it fits better with the management information requirements of research institutions.

Conclusion

The mosaic of grey literature is not yet revealed easily. Its complex patterns representing structures, and the beauty of the complete form are not recognised. This is because of the heterogeneity of the sources, the lack of a canonical schema either for storage/query/results management or for interoperation over heterogeneous systems. Worse, existing sources use metadata schemas that do not have sufficiently formal syntax and lack declared semantics – both of which can be rectified by the use of CERIF.

The take-home message is clear: use CERIF as the canonical schema for grey literature. Ideally all systems would be conforming to the CERIF schema but – to accommodate legacy systems – for a while with heterogeneous systems having a CERIF wrapper. This would mean that:

1. query and retrieval provide better relevance and recall;
3. data input quality is improved;
4. systems can interoperate, to provide the end-user with a homogeneous view over heterogeneous distributed systems;
5. statistical and graphical processing can be reliable;
6. interoperation with other systems within and outwith the research organisation is facilitated.

References

- (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)
- (Asserson and Jeffery 2005) Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' The Grey Journal volume 1 number 1: Spring 2005 TextRelease/Greynet ISSN 1574-1796 pp5-8
- (CERIF) www.eurocris.org/cerif
- (DC) http://www.oclc.org:5046/research/dublin_core/
- (DSpace) DSpace repository product homepage http://www.dspace.org/
- (ePubs) ePubs repository product homepage http://epubs.cclrc.ac.uk/
- (eprints) ePrints repository product homepage http://www.eprints.org/
- (Fedora) Fedora repository product homepage http://www.fedora.info/
- (FRIDA) http://frida.uio.no
- (Jeffery 1999) Jeffery, K G: 'An Architecture for Grey Literature in a R and 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 2004a) 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 2004b) 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 2005) K G Jeffery CRISs, Architectures and CERIF CCLRC-RAL Technical Report RAL-TR-2005-003 (2005)
- (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
- (Jeffery and Asserson 2005) K G Jeffery, A G S Asserson 'Grey in the R&D Process'; Proceedings Grey Literature 7 Conference, Nancy, December 2005; TextRelease; ISBN 90-77484-06-X ISSN 1386-2316
- (Jeffery and Asserson 2006a) Keith G Jeffery, Anne Asserson: 'CRIS Central Relating Information System' in Anne Gams Steine Asserson, Eduard J Simons (Eds) 'Enabling Interaction and Quality: Beyond he Hanseatic

League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp109-120 Leuven University Press ISBN 978 90 5867 536 1

- (Jeffery and Asserson 2006b) Keith G Jeffery, Anne Asserson: 'Supporting the Research Process with a CRIS' in Anne Gams Steine Asserson, Eduard J Simons (Eds) 'Enabling Interaction and Quality: Beyond the Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp 121-130 Leuven University Press ISBN 978 90 5867 536 1
- (Jeffery and Asserson 2006c) Keith G Jeffery, Anne Asserson: 'Hyperactive Grey Objects' Proceedings Grey Literature 8 Conference (GL8), New Orleans, December 2006; TextRelease; ISBN 90-77484-08-6. ISSN 1386-2316 ; No. 8-06-X
- (Jeffery and Asserson 2007) Keith G Jeffery, Anne Asserson: 'Greyscape' Opening Paper in Proceedings Grey Literature 9 Conference Antwerp (GL9) 10-11 December 2007 pp9-14; Textrelease, Amsterdam; ISSN 1386-2316
- (Jeffery and Asserson 2008) Keith G Jeffery, Anne Asserson: 'INTEREST' Proceedings Grey Literature Conference Amsterdam 8-9 December 2008 in Tenth International Conference on Grey Literature : Designing the Grey Grid for Information Society, 8-9 December 2008, Science Park Amsterdam, The Netherlands ed. by Dominic J. Farace and Jerry Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, February 2009. GL-Conference series, ISSN 1386-2316; No. 10. - ISBN 978-90-77484-11-1.
- (MARC) http:// http://www.loc.gov/marc/
- (PURE) http:// http://www.atira.dk/en/pure/
- (RDF) http://www.w3.org/RDF/
- (XML) http://www.w3.org/XML/
- (Z39.50) http:// http://www.loc.gov/z3950/agency/

Are Legal Texts Grey Literature? Toward a definition of Grey Literature that invites the Preservation of Authentic and Complete Originals

Michael Lines,
University of Victoria, Canada

Law books in the popular imagination

Legal texts, though they exist in a wide variety of forms, are most typically thought of as Law Books. Law books, hardbound volumes in expensive bindings of browns and blacks, are heavy, difficult, and technical. They are a prop to popular conceptions of the law itself, and resemble more closely than most other earthly books the Platonic form of the 'weighty tome.'

In fact, some law libraries do a regular, if not exactly brisk, trade in renting their law books to TV and film productions. And the more dour the entertainment, the more likely it is to include law books in the background. Perry Mason was too active a man to spend much time in his law office, so we did not see the floor-to-ceiling oak shelves of reserved and wise volumes which undoubtedly offered him nightly counsel. In the '80s, LA Law was too concerned with exciting power suits and hairdos to have much need of our serious friends. The '90s brought us Law and Order, the dourest thing Americans have thus far been able to stomach on a regular basis. Much like marmite in the UK, or beef jerky in Canada, this distinctive fare seems to have become a regular and much-loved part of the daily American diet. As a result, US citizens, and those of us in the provinces of the empire as well, are treated to regular glimpses of the formidable law book in law office scenes. I would guess that Law and Order also offers the very occasional library scene, though I'm not enough of a devotee to attest to this "so help me God" on a stack of Bibles.

To the extent that there actually is a popular conception of law books, it is undoubtedly one of authority, tradition and power. Of course, it is naturally allied to the reputation of lawyers and of the justice systems generally, which are not unalloyed gold. Between the Innocence Project, which has had many successes exonerating those wrongly convicted and awaiting execution, and vagaries of international 'law' which were brought to public notice regarding the Guantanamo detainees, there exists much evidence of the failings of our systems of justice. In these dramas, law books are sources of traditions that can communicate both rights and wrongs.

The mixed character of law books can be summarized in two additional qualities, which it is safe to say escape entirely the purview of the casual onlooker: Law books are both official and public. In their official character they reveal their intimate connection with bureaucracy. Law books are perhaps the most official and authoritative printed result of any bureaucratic process, and in this they are a pinnacle in our doomed efforts to exert systematic control over the wildness of everyday life. Technical and convoluted, law books also remind us of the grim possibility of state-sanctioned force overriding our human rights: law books can send you to jail, and it is this powerful capacity, their official nature, that most distinctly sets them apart from grey literature, which, to say the least, has credibility issues.

Law books are also public, or at least they are in less paternalistic societies. In this aspect law books, and the legal texts they contain, represent trust in the good sense of our fellow citizens, and in the traditions of conversation and consensus building that create individual freedoms and which have led to the discovery, recognition, and protection of human rights in the first place. The public character of law, the legal texts which articulate them and law books where they are physically expressed is one of the most promising characteristics of any literature.

Characteristics of Primary Legal Texts

Legal texts are either statements of the law or statements about the law. Statements of the law are known as primary legal sources, and statements about the law are known as secondary legal sources. Primary legal sources carry much more authority in court than do secondary. Inasmuch as a court of law can be viewed as a laboratory in which the facts of everyday experience are examined with a view to discovering, refining, and elaborating our understanding of the laws under which we naturally operate, the primary legal texts represent the established body of knowledge, and they are for the most part accepted without incident as 'fact', that is, as a reality. The exceptions, of course, are where new facts bring into

question the completeness of the established understanding. In this regard, primary legal texts are authoritative, but not unchallengeable.

From an insider's point of view, and over many years of close association, law books and the legal texts they embody, like spouses, shed much of their original mystique, and reveal more complex characters. It is rewarding to believe that these continually uncovered qualities are even more fascinating than the diverting features that encouraged the initial love affair. This process of discovery reminds us that, in spite of their strongly conceptual nature, legal texts and law books are not abstractions; they take part in and reflect the evolving complexity of the real world.

Primary legal texts find their physical expression in a variety of formats, but they are almost exclusively of a serial nature. In accord with their reputation as imposing and their origins in bureaucracy, they are substantial hardbound volumes bearing the signs of serious intent: elaborate insignia, overlong and syntactically obscure titles, and the prominent reproduction of the names of officials. Generally, and with the notable exception of the US, they are produced by government printing houses, not by commercial publishers, a fact that nominally though not definitively brings them within the domain of Grey Literature. These works also contain few indexing tools, no cataloguing in publication information, and little bibliographic data.

Good Canadian examples of primary legal texts are the *Acts of the Parliament of Canada*, and the decisions of the Supreme Court of Canada. The *Acts*, which are the laws created by Canada's House of Commons and Senate, offer no tables of contents, no cataloguing information, and the barest of publisher information. Thankfully, there are a couple of officially produced indexing tools for this series which help researchers navigate the cross-currents of amendments, repeals, coming into force information, and numerous other complicating factors. Figure 1, the verso of the title page of volume 1 of the 1993 set, captures the full extent of the publisher data provided.

Figure 1

The *Canada Supreme Court Reports* is similar, with a different publication attribution, and minus the index. As shown in Figure 2, the publisher data in volume 1 of the 2001 set is concise, if not cryptic:

Figure 2

Where access to the information is not adequately supplied in an official source, the work is left to commercial publishers, should they choose to provide it. Thus, in paper at least, and relying on official sources alone, the decisions of the Supreme Court of Canada must be read diligently from start to finish in order to adequately gain access to the content. Naturally, the success of that project relies entirely on the capacity of the reader. This lack of tools for access to the text is characteristic more of Grey Literature than it is of commercial publications

In another respect these law books do resemble commercial publications: their method of distribution; like commercial publishers, Canadian government printing operations make their materials available on standing order, and thus they arrive bound and shelf-ready at libraries

across the country. And they certainly *appear* to be 'books' as opposed to Grey Literature. Plus, their status as official statements of the law separates them from Grey Literature.

However, not every legal text has this official status. A good example is the proceedings of the Canadian Houses of Parliament, known here as The Hansards. These 'verbatim' accounts of the public debates in our nation's capital are the 'official' record, and as such they do have a special importance in courts. However, they are not statements of the law in the same sense as the statutes, to which they are a precursor. The Hansards arrive weekly in subscribing libraries as stapled supplements. We then bind them in-house in a very traditional style. If we chose, we could bind them in pink with Hello Kitty knockoffs on the flyleaves. They do not include indexes, tables of contents, or cataloguing in publication information.

By some accounts, the Hansards are Government Documents and not legal literature more closely defined. If we accept this definition, the result is that an important and officially sanctioned resource often used in courts to resolve legal disputes is a species of Grey Literature. Thus to the extent that the Hansards step over the line between Government Documents and Legal Literature, we can claim ground from legal texts for Grey literature.

In the same way that the Hansards are official legal texts that provide background to legislation, the courts also produce a variety of documents that have an official character, and which are primary legal textual sources, but which are of a more specific or narrow character. Here we find the factums of the parties (formal statements of the arguments of the parties to an action), injunctions issued by the court on behalf of parties, and a welter of other documents, such as statements of claim, expert reports, and other forms of evidence.

What all these documents have in common is their resemblance to business records. They are documents produced for the accomplishment of a practical activity, and are very typical of bureaucracies. These documents share with primary legal documents a special status before the courts which is important to examine.

Standards for the Verification of Authenticity

As evidence in courts, documents are generally understood to be an exception to the hearsay rule. The hearsay rule asserts that statements made by third parties cannot be referred to by witnesses as evidence, because the reporting of them is liable to be inaccurate. In the case of documents, an exception is made because it is understood that a document tends to record matters in a manner and especially at an earlier time when the selfish motivations and other factors influencing parties are naturally minimized. Despite the ease with which a false receipt or handwritten IOU can be produced, it is recognized by courts that the forethought required of parties to produce such documents months or years in advance of an appearance in court on a particular action is, for the most part, beyond the capacities of the average sorry sod who finds himself enmeshed in the desperately slow and expensive machine of justice. What is true for the most informal of documents is that much truer for business documents produced by an office, or even better, a bureaucracy, in the regular course of its operations. Thus emails showing the transactions between a client and his agent, communications describing progress on a file, and the contracts the parties produce are all accepted with regularity into courts of law.

It is for the individual judge to assign an evidentiary value to these documents, and part of that assessment will be a consideration of the likelihood of the texts being accurate, i.e. not tampered with or otherwise compromised. In the particular case of court documents, there is an understandable if not entirely excusable presumption of accuracy when they arrive before a court. A simple ink stamp from the court registry, even if it be from an entirely unconnected court system, goes a long way, at least in Canadian courts. The fact that there is no systematic and reliable method of independently verifying the accuracy of the contents of court documents says a lot about the general level of assurances a court requires against fraud and error.

The reliability of the text of legal and quasi-legal sources is not, as seems to be assumed, secure. Without an inside view of government publishing processes, not much can be said, Still, the Hansards provide a suggestive sidelight. The Hansards are 'verbatim' accounts of the speeches delivered by the Members of Parliament. However, any researcher who has collected and transcribed interviews can tell at a glance that these are heavily edited versions. They have no ums, no ahs, few false starts, and minimal repetition. Behind the scenes there must exist multiple stages of textual preparation. In some respects this process may resemble those of commercial publishers, and in other respects it is likely to rely on records management systems, since these documents are produced on a continual basis. As the New Bibliographers demonstrated in the early 1900s, with each new transition of the text

differences, both intentional and not, are introduced. As Derrida put it (and I paraphrase): "iteration occurs: something new appears." [1]

Summary: Are Legal Texts Grey Literature?

No:

- Legal texts look and are official
- Their special status as statements of the law distinguishes them
- They are produced and distributed in a commercial-like manner

Yes, in most other respects:

- Not commercially published
- Opaque textual controls
- Lack of bibliographic and content tools
- Bureaucratic in character: "records produced in the normal course of business"

What Legal Texts Can Tell Us about Grey Literature

What the preceding demonstrates is that primary legal sources and many quasi-legal documents fall on a scale between commercial publications and bureaucratic records. The same can be said for many varieties of Grey Literature. Primary legal sources likely partake of a mix of traditional records management and commercial publishing methods of textual control. That is also likely true for many kinds of Grey Literature.

Archival Studies is the discipline that explores how to maintain collections of records in a series, be they current or historical. In archival terms, the definition of a record is "A document made or received in the course of the conduct of affairs and preserved."[2] The characteristic of a record that most clearly distinguishes it from a commercial publication is the 'archival bond' that exists between records in a series, and which implies or flows from the purpose of the creating body, as expressed in its activities[3]. That is, there is an intrinsic connection between records and the purpose of their creating body. This connection is everywhere apparently in Grey Literature. These two qualities have been summarized in the following graphic:

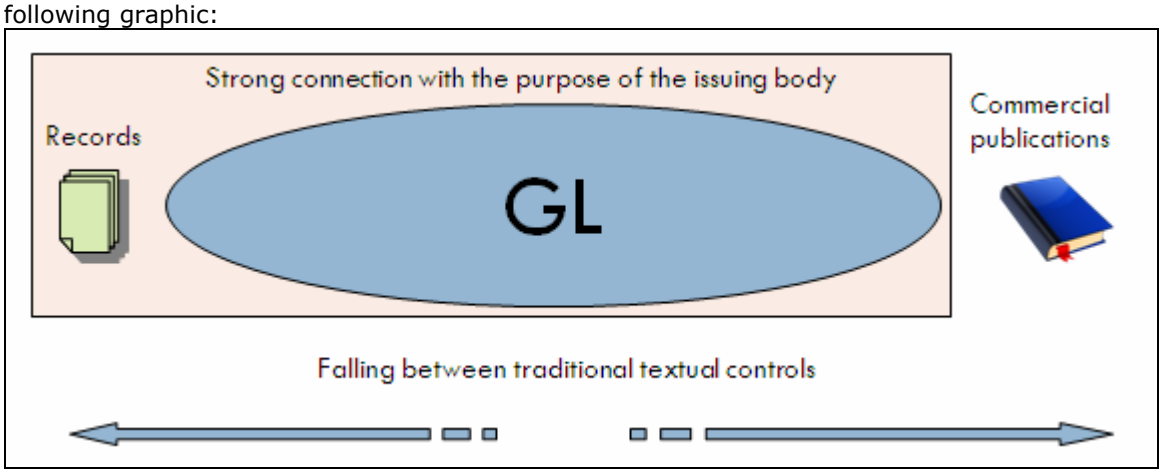


Figure 3

Reliability of the Text: Concepts Relevant to Legal Texts from Archival Studies that Can Bring Better Control and Access to Grey Literature

The strong intrinsic connection between works of Grey Literature and the purpose of their issuing bodies should be considered a candidate for inclusion in the definition of Grey Literature. Doing so would encourage the Grey Literature community to take advantage of numerous archival research initiatives that provide direction to the establishing and preserving the accuracy and context of the GL text. As a start, designers of Grey Literature repositories could take advantage of three foundational archival concepts: originality, authenticity, and completeness:

- Originality: The quality of being the first complete and effective record.
- Authenticity: The quality of being authentic, or entitled to acceptance. As being authoritative or duly authorized, as being what it professes in origin or authorship, as being genuine.

- Completeness The characteristic of a record that refers to the presence within it of all the elements required by the creator and the juridical system for it to be capable of generating consequences.[4]

Without entering into a detailed description of these terms and their implications, I invite you to explore the InterPARES Project, a multi-year project that is determining the basic qualities of electronic documents and the ways in which they can be preserved.

Relevance of These Concerns to the Online Environment

To demonstrate some of the issues that can be addressed by taking the security of the text more seriously, I return to the area of legal texts as they exist in the online environment. I will concentrate on a single example, though there are numerous issues that deserve attention.

In recent years with the acceleration of the Open Access movement, there has been a growing awareness, especially in US circles, that access to the law as expressed in primary legal texts has lagged. Though Open Access to the law initiatives were among the very earliest following on the advent of the internet, at this point the law remains in many respects difficult to access, and sometimes it is entirely sequestered by commercial interests. Carl Malamud established Bulk.resource.org for the purpose of remedying this situation. With the help of commercial law publisher West and others, he obtained the text of hundreds of thousands of US court decisions, and made them available for free on the internet as easily-downloaded files, either individually or in groups.

The texts he provides come marked up in XML; they are available in a sophisticated form that allows processing on a large scale with specificity. Thus they provide easy access to programmers who wish to present them in any number of forms. Bulk.resource.org is an invitation to innovate. Now anyone has access to the materials and can offer them to the world with new features for exploring them. This is a positive development, but it also gives us pause. The opportunity for mischief and error is very great.

For instance, in the Fall of 2009, Google made use of this resource to offer a new service to users of the Google Scholar portal. Called "Legal opinions and journals', this service comes with a disclaimer:

> Legal opinions in Google Scholar are provided for informational purposes only and should not be relied on as a substitute for legal advice from a licensed lawyer. Google does not warrant that the information is complete or accurate.[5]

And rightly so, since in the past Google's efforts to provide complete access to specific bodies of work have not always been entirely successful[6]. That fact has not escaped Mr. Malamud and his colleagues. Recently they initiated Law.Gov, which is "an effort to create a report documenting exactly what it would take to create a distributed registry and repository of all primary legal materials in the United States."[7] As part of the report, they anticipate addressing the issue of authenticity, and also of preserving authenticity.

---

[1] This issue has been recognized by the American Association of Law Libraries, which prepared a report on it in 2007 and continues to work on the question. It is particularly important regarding online versions of primary legal texts. See Access to Electronic Legal Information Committee and Washington Affairs Office, *State-by-State Report on Authentication of Online Legal Resources* (Chicago: AALL, 2007).

[2] Planning Committee on Descriptive Standards, *Rules for Archival Description* (Bureau of Canadian Archivists: Ottawa, 1990, revised 2008, p. D-7) http://www.cdncouncilarchives.ca/RAD/RADComplete_July2008.pdf.

[3] "The relationship that links each record, incrementally, to the previous and subsequent ones and to all those which participate in the same activity. It is originary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record).

[4] Luciana Duranti et al. The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project (UBC: Vancouver, 2001) http://www.interpares.org/book/interpares_book_q_gloss.pdf

[5] Google Scholar Help: http://scholar.google.ca/intl/en/scholar/help.html.

[6] See HeinOnline or Google Scholar? Why You Should Start Your Research in HeinOnline First (http://heinonline.blogspot.com/2009/11/heinonline-or-google-scholar-why-you.html) and HeinOnline or Google Scholar? Why You Should Start Your Research in HeinOnline First - Part II (http://heinonline.blogspot.com/2009/12/heinonline-or-google-scholar-why-you.html)

[7] Law.Gov. http://public.resource.org/law.gov/

ERIC℠ Education Resources Information Center

The Source for Grey Literature in Education

More than 100,000 open access full-text documents including:

- Research reports sponsored by the U.S. Department of Education
- Conference proceedings and papers
- State education agency reports
- Research and technical reports from scholarly societies
- Policy papers, working papers, position papers
- Other education-related materials

ERIC provides:

- More than 1.2 million bibliographic records of documents, journal articles, and books
- Education grey materials indexed from the 1960s to the present
- A robust search interface to aid the discovery of materials
- Access to the online *Thesaurus of ERIC Descriptors*

Search the extensive ERIC Collection at ***www.eric.ed.gov***

ERIC is a program of the U.S. Department of Education, Institute of Education Sciences, National Library of Education.
This is a public service announcement from the U.S. Department of Education.

ies INSTITUTE OF EDUCATION SCIENCES

Multi-Institutional Approach to Technical Report Literature: Development of the Technical Report Archive & Image Library (TRAIL)

Maliaca Oxnam
University of Arizona Libraries, United States

Abstract

While availability and access to more recent technical report literature has greatly improved through electronic delivery, older technical report literature still remains elusive to users. The challenges posed by these collections are two-fold: 1) the diverse nature of distribution and collection building practices across institutions; and, 2) the variability in the selection and acquisition methods used, resulting in incomplete series, confusion in depository and non-depository status, lack of awareness and definition of a "complete collection," and a host of other inconsistencies.

Through a national dialogue among academic librarians led by University of Arizona Libraries (UAL) beginning in 2005, common discoveries were made about the various barriers libraries were experiencing in managing and making accessible legacy technical report collections, and the urgent need to address these issues. The positive response to these dialogues and preliminary assessment results led to an extraordinary opportunity to resolve a difficult challenge in academic and special libraries. These problems created a strong case for establishing the Technical Report Archive & Image Library (TRAIL) collectively maintained by the library community.

The Technical Report Archive & Image Library (TRAIL) aims to identify, digitize, archive, and provide persistent and unrestricted access to federal technical reports issued prior to 1975. TRAIL is a Greater Western Library Alliance initiative led by the University of Arizona in collaboration with the Center for Research Libraries. TRAIL is funded by the member institutions of the Greater Western Library Alliance with an estimated annual budget of $128,000. To date, TRAIL has scanned more than one million pages of technical reports issued by the US federal government agencies and has involved participation and contribution from more than twenty universities. This paper will describe TRAIL, including the need for its development, current structure, and future directions.

Why a focus on federally funded technical reports?

In 2004, the University of Arizona Libraries (UAL) focused efforts on eliminating backlogs of purchased materials needing to be processed and cataloged. One of the results of this focused effort was the recognition that our library had collections of materials, sitting on the shelves available to users, that were deemed valuable, but with little or no title level cataloging, or other finding aids available. These collections were virtually "hidden" from our users. While available, most of these collections required that users employ some form of mediation or assistance from the library, which was the exact opposite direction from the Library's strategic direction at the time for creating "self-sufficient users" who could access the library "anytime, anyplace". One of the identified collections was a large collection of print technical research report series from various US federal agencies. The collection was housed in our Science-Engineering Library, separate from our other government document collections. These collections were purchased prior to UAL's establishment as a selective depository library and/or were purchased directly from the agencies and hence were not part of our official government documents depository collection. In addition to our print collection, there were also many thousands of reports that we owned issued either in micro-opaque or micro-card format and microfiche format. To further the complexity, these collections were full of gaps in the numbering sequences of the report series. These gaps were caused by reports that weren't publicly issued, reports that were not available at the time of the purchase of our collections and never filled in as the reports may later have been released, and also likely to natural theft and loss over time.

The UAL spent considerable time trying to locate shipping and/or receiving lists to get an item level inventory for all the technical reports we'd received in our collection. It was hoped that we might be able to create some form of inventory of items that could be added to our local OPAC. However, it didn't take us long to realize that we weren't going to be able to find records of what items had been shipped and/or received, and hiring students to manually inventory our collection was going to be time-consuming and costly. Providing better access to these collections wasn't a project that could realistically be undertaken by any one institution. And UAL temporarily ceased its project until funding and or technologies to help with the project could be secured.

In the summer of 2005, a discussion on technical report collection access was brought forward to interested members of the Engineering Libraries Division of the American Society of Engineering Education. A group email list was started for the more than 80 individuals who expressed interest in creating better access to these early technical report collections that so many of us held in our individual library collections. Ideas for preferred access points for libraries and users were discussed. Unanimously, librarians expressed the desire to have online access to these older reports with the recognition that few of the federal agencies were budgeted to digitize or would be able to provide online access to all of the older materials held in our collections. Through discussion, it was reinforced that many libraries, including government depository libraries and federal agency libraries held incomplete collections. To fulfill user requests for items, we really needed the collective whole of all the collections in order to ensure that any given user request could be met. Discussions also led to informal surveys regarding which report collections were thought to be most useful and most desired by the group as a whole.

Development of the Technical Report Archive & Image Library

Later the next year, after discussing the idea of a collaborative project with the Center for Research Libraries (CRL)[1] and a regional library consortium called the Greater Western Library Alliance (GWLA)[2], it was agreed that a joint project would be formed to explore and pilot the development of the Technical Report Archive & Image Library (TRAIL)[3]. The project would be supported and managed under the 32 member institutions of the Greater Western Library Alliance. The charge for TRAIL was to:

Develop a collaborative project with the Center for Research Libraries to identify, digitize, archive, and provide persistent and unrestricted access to federal technical reports issued prior to 1975.

A ten-member taskforce was appointed and an initial meeting was held in the fall of 2006. Ideas for workflows and costs of digitizing and hosting digital copies of technical reports were discussed. Throughout the months that followed discussions on the many pros and cons regarding collection acquisitions, workflows, digital conversion processes, digital archiving, and print archiving ensued. From these discussions, the guiding principles for TRAIL were developed:

1. TRAIL would have a primary digital collection (and archive) with a corresponding print archive.
2. Opportunities for participation and involvement should be available to all institutions.
3. Efforts would be made to digitize and archive all items in a selected report series, before working on additional series.
4. The taskforce would only focus on reports issued in print format.
5. The taskforce would only focus on documents in the public domain or for which they received explicit copyright permissions from the copyright holder.

The taskforce started to work on digitizing its first collections. By the Spring of 2007, a working pilot collection was released. Upon review, GWLA agreed to fund the project for a year to continue the development of TRAIL. GWLA members each agreed to pay $4,000 dollars providing TRAIL with $128,000 to cover digitization, overhead and staff costs for the year. During this time, a number of partnership opportunities were presented that greatly aided the project in working towards its objective. TRAIL partnered with the University of Michigan (UM) and Google through the University of Michigan's Google Books project to aid with the scanning and digital hosting and archiving of many reports. Items that met the Google scanning requirements were cataloged, barcoded and scanned through this process. A year later when the University of Michigan and the Center for Institutional Cooperation (CIC)[4] released the Hathi Trust Digital Archive[5], these items were deposited by the University of Michigan, as part of their collection. Items which cannot be scanned through the Google Books because of fragility, inclusion of maps or other special handling needs are outsourced to a digitization vendor. These items are then hosted and made available through the TRAIL public interface.

With the UM partnership came many changes to our processing workflows, one of which was creating a MARC record for each individual report processed. The benefit of this workflow change was that the project could then work with OCLC to provide collection sets for all of the digital versions of these items. Libraries could then purchase sets of records to be added to their OPACS increasing access to these report collections. (To date, many of the individual item records are searchable and available in OCLC. However, they have not yet been pulled into electronic collection sets by OCLC. This is due to the process of needing to convert the OCLC records for the print items into new electronic records with the receipt of the URL from UM and Google. It is hoped that this will be resolved January 2009.)

The work of TRAIL is overseen by the TRAIL taskforce. As a new report series is identified to be acquired and digitized a lead organizer is appointed. (It is not necessary for the lead to be a member of the taskforce.) This person is responsible for identifying and

acquiring two complete sets of reports for the given series. (Items are contributed by libraries and organizations nationally.) One set will be used for digitization and the other set is secured by the print archive. As items are acquired they are sent to the University of Arizona Library (also referred to as the "Central processing site") where the items are checked for special format needs and copyright compliance. They are then catalogued (using either copy cataloging or original cataloging) and packed for shipping to either Google or another digitization vendor. UAL then oversees the digitization status and quality checks for all scanned items. As a report series are completed, UAL works with OCLC to close out the cataloging authorizations and create the collection sets.

Future Directions

After two and a half years (not including the pilot), TRAIL has digitized more than 15,000 technical reports, which is more than 1.2 million pages of federally funded research. TRAIL has received requests and comments from all over the world from researchers ranging from Chile to Greece whom otherwise have had difficulty locating, ordering and purchasing reports.

At present, the digitized collection is being moved from the pilot site to a more permanent location, which is split between the University of North Texas and the Hathi Trust (via partnership with the University of Michigan). A new public interface and search service is being created by the University of Washington. The new interface being developed will allow for a federated search of servers in both locations. This will greatly improve access to the more than 15,000 reports that are currently available.

The success of TRAIL continues to bring new opportunities. Most recently the TRAIL taskforce is exploring moving TRAIL under the umbrella of the Global Resources Network[6] (GRN) overseen by the Center for Research Libraries. Moving to the GRN will allow greater ability to explore partnerships at a global as well as national level to partner on making technical research reports more widely available. It will also allow institutions to provide funding support and participate in the program without having to be a member of either the Greater Western Library Alliance or the Center for Research Libraries. Under the arm of the GRN, TARIL will be better positioned to advocate on the importance of ensuring access to these important documents in the history of our research and technology development in the US. Individuals or institutions wanting to learn how to get involved with the Technical Report Archive & Image Library can find additional information at http://trail.gwla.org.

References

[1] CRL, Center for Resaerch Library http://www.crl.edu
[2] GWLA, The Greater Western Library Alliance http://www.gwla.org
[3] TRAIL, Technical Report Archive & Image Library http://trail.gwla.org
[4] CIC, Center for Institutional Cooperation http://www.cic.net
[5] Hathi Trust Digital Archive, http://www.hathitrust.org
[6] GRN, Global Resources Network http://www.crl.edu/grn

Digitizing Grey Literature from the *Antarctic Bibliography* Collection

Tina Gheen and Sue Olmsted
National Science Foundation, United States

Abstract

In 1962 the National Science Foundation (NSF) created a clearinghouse for Antarctic information intended to foster the global, free exchange of data and publications between scientists and researchers. With funding from NSF, the Library of Congress began assembling the Antarctic Bibliography in 1963, and full-text of the items listed in the bibliography was later captured on microfiche for preservation. The Antarctic Bibliography primarily consists of journal articles, monographs, technical reports and conference proceedings collected by the Library of Congress; however, other polar libraries contributed a generous amount of grey literature material to the bibliography as well. Only a few copies of the microfiche collection were created and distributed, so the full-text version is a rare resource preserved in an increasingly less accessible format. The goal of the Polar Digitization project at the National Science Foundation Library is to make the full-text grey literature materials from the Antarctic Bibliography microfiche collection, including unpublished works, materials from conferences that are not readily available and rare government reports, available to the public electronically on the web through digitization and cataloging. This paper addresses the challenges and importance of making the full text of grey matter from the Antarctic Bibliography accessible in digital form.

Introduction

The U.S. National Science Foundation (NSF) is a government grant agency which funds research and education in most fields of science and engineering. NSF's Office of Polar Programs (OPP) manages and funds the United States Antarctic Program (USAP)[1]. In 1962, USAP established a clearinghouse for Antarctic Information at the NSF. The Antarctic Bibliography was a product of this initiative. From 1963 to 1998, the Library of Congress prepared the *Antarctic Bibliography*, a multi-volume set of citations and abstracts encompassing the world's scientific literature regarding Antarctica and the Southern Ocean [2]. The bibliography coverage spans from 1951 to 1998, includes Antarctic biology, geology, engineering, medicine, meteorology, oceanography, atmospheric physics, geophysics, and political and social science. The bibliography contains two sets of information: 1) bibliographic records with very brief abstracts of the world's Antarctic research literature from 1962 to 1998 in all of the scientific disciplines relevant to Antarctica (approximately 65,000 titles) and 2) bibliographic records without abstracts of the world's Antarctic research literature from 1951 to 1961 in all of the scientific disciplines relevant to Antarctica (approximately 5,000 titles)[2]. All material included in the printed bibliography was indexed by a team of subject specialists in the Science and Technology Division of the Library of Congress.

In addition to the printed volumes, microform copies of the full text of most of the cited items in the bibliography were produced by the Library of Congress as part of the funded project. These microfiche contain scientific articles, conference proceedings, technical reports, official memoranda, policy documents and other relevant publications. NSF has microfiche sets of the full texts of Antarctic items for on-site use only at the NSF building in Virginia, its three research stations in Antarctica, and the University of Canterbury in New Zealand (for NSF's Antarctic facility in Christchurch)[3]. The American Geophysical Institute (AGI) also has a set of the fiche at their headquarters in Alexandria, Virginia.

The National Science Foundation Library is an information and reference resource supporting the USAP and the NSF OPP. As custodians of one of the rare archival sets of microfiche of full-text materials collected from the *Antarctic Bibliography*, we have become increasingly aware of the need to make the grey literature information contained on this fairly inaccessible microfiche collection more readily available to the scientific community. The bibliography holds a wealth of historical information related to scientific data, climate studies, environmental changes, policy, and historical events. The technical reports alone cover topics such as snow road construction, seismic wave dispersion, studies of ice cores, geodetic surveys, chemical profiles of ice shelves, marine biological studies, and cartography. Many of these technical reports and government-funded studies are not accessible and are no longer available as monographs. In addition, the fiche version of the bibliography captured proceedings from societal conferences held all over the world.

The *Antarctic Bibliography* is currently maintained and updated online by the American Geological Institute (AGI)[4]. AGI distributes the bibliography free on the web through the Cold

Regions Bibliography Project (http://www.coldregions.org/) and provides links to full-text when possible, but much of the full text grey literature must be digitized on demand for a fee. The National Science Foundation Library began a pilot project to digitize the grey literature contained on the *Antarctic Bibliography* microfiche. However, digitizing and delivering the content currently preserved on microfiche presents a host of challenges including material selection, language barriers, original image quality, inclusion of metadata, and copyright issues.

Methodology

Much of the material contained in the *Antarctic Bibliography* was acquired from journals and books. Grey literature can be found throughout the collection, particularly in the materials submitted by polar libraries. Because much of the information captured during the microfilming process is now copyrighted and available from journal publishers, the selection of materials to digitize becomes critical. Therefore, we concentrated on the evaluation of technical reports, rare conference proceedings, and other information which falls into the category of grey literature. For this paper, we define grey literature as information that is not widely published or distributed, and which is usually difficult to discover or obtain, particularly unpublished works, materials from conferences that are not readily available and rare government reports.

Because NSF does not have the necessary equipment or staff to produce quality digital reproductions of the fiche, samples of the fiche were sent to the OCLC Preservation Service Center for automated digitization. The deliverables for the project included uncompressed 300 dpi grayscale archival TIF files and derivative bitonal bound PDF files with hidden text.

For our pilot study, we selected approximately five hundred fiche (~3800 individual cells) of grey literature materials from the full-text version of the *Antarctic Bibliography*. The samples were selected based upon a manual, internal evaluation to determine whether the fiche contained grey literature materials and whether the materials existed in digitized form on the web.

Anatomy of the Antarctic Bibliography Fiche

Manual identification and selection was a major, time consuming component of the initial phase of our pilot. The full text materials of the *Antarctic Bibliography* are not often easy to distinguish separately. An individual fiche may contain only one article, book, or proceeding, or it may contain a combination of several articles, books, and reports. Careful manual inspection for a black cell or unique identifier is usually required to distinguish one item from another. Similarly, a packet of several fiche may contain one long article, book, proceeding, or report, or a packet of fiche might span a number of individual publications.

Since our project specifically targeted grey literature, we did not provide a single block of hundreds of consecutive fiche because of the likelihood the block would contain mass produced full text books, commercial online journal products, and other materials which involved copyright issues with publishers. The consecutive packets often contained some desired technical reports and proceedings, but also included one or more journal articles, which were already digitally available from a publisher. Because of this, it was often impossible to select fiche that contained only grey literature publications. It was likely materials which had already been converted to digital format would be included in the span.

Evaluation and Selection

In order to avoid duplication of efforts and best use the financial resources available to us, we needed to carefully evaluate and select sample items to send to the preservation specialists at OCLC. Using the index of the *Antarctic Bibliography*, staff selected desirable samples of technical reports, proceedings, and other materials. These samples were identified by a unique accession number from the Library of Congress, which was added to the *Antarctic Bibliography*. Staff then manually recorded fiche number, item type, author, article title or name of conference, number of pages, and publication year. This information was used in searches on various online databases such as OCLC WorldCat and Public STINET to determine if the material was readily available. If the selection was available from several lending institutions or agencies, it was not a candidate for our pilot.

After evaluating the accessibility of the samples, staff manually assessed and evaluated the fiche. The number of cells per fiche, the number of images per cell, title, and number of owners on WorldCat were recorded. Staff then compiled a master list which contained:

- Library of Congress accession number (ID #)
- Corresponding volume of the print version of the Antarctic Bibliography where the citation was found

- Material type:
  - Book
  - Book Chapter
  - Booklet
  - Information Bullietin
  - Proceeding
  - Report
  - Technical Note
  - Technical Report
  - Etc.
- Number of images contained in the sample
- Publication year
- Additional Comments
  - Language
  - Availablilty

Digitization

We analyzed this data to choose a variety of appropriate grey literature materials for our pilot in one small sample set, which was sent to the OCLC Preservation Service Center in Bethlehem, Pennsylvania. From these samples, OCLC created a single TIFF per fiche frame. Because the process is automated, target or divider pages could not be excluded and were also digitized. This smaller initial sample set allowed us to evaluate quality before continuing with more digitization.

Analyzing and selecting specific publications led us to discover issues related to the initial filming of the microfiche. Nonuniformity in microfilming and standards created challenges for automated digitization processing. Fiche were filmed by various technicians using different standards, equipment, and procedures over the years. In particular, the background used for some of the fiche appeared to be a highly textured fabric. This background caused problems for the automated OCLC equipment since automatic detection of the page edge was impossible. To solve the problem, the fiche were scanned using an automatic crop box. Unfortunately, some of the pages were so badly skewed during the initial fiche filming that part of the page fell outside the cropped area (Image 1).

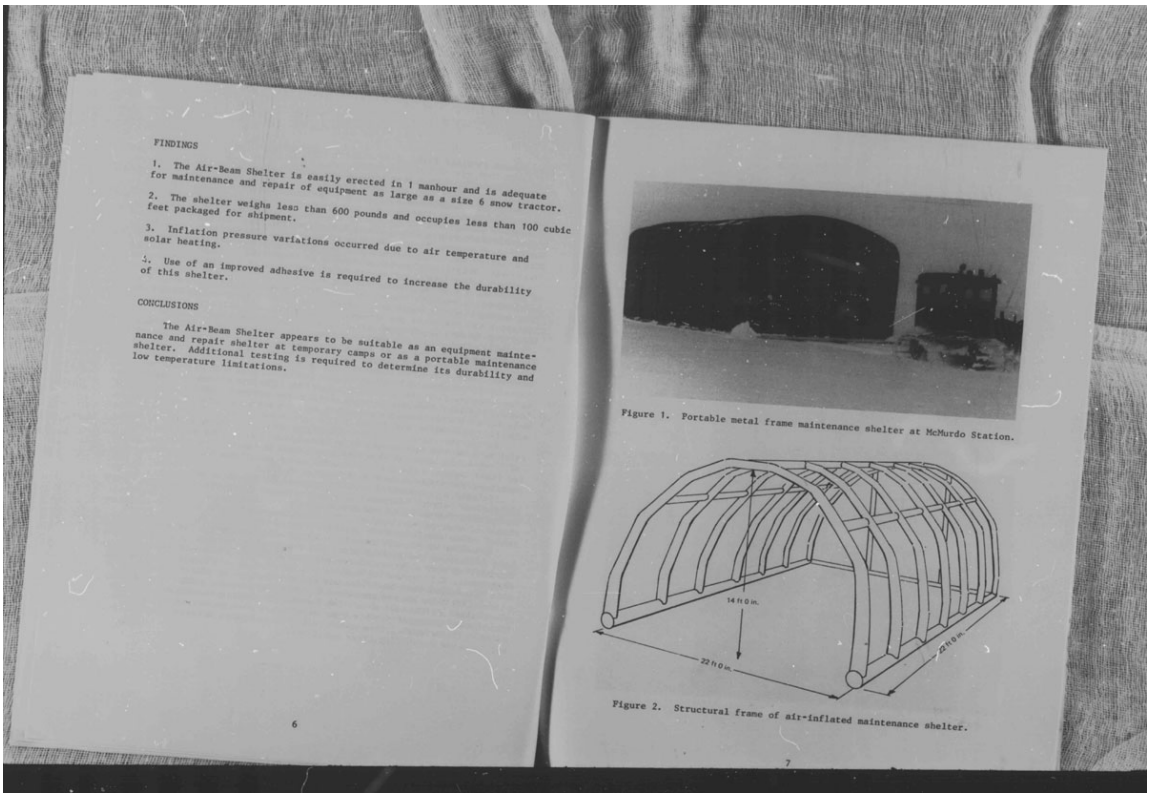
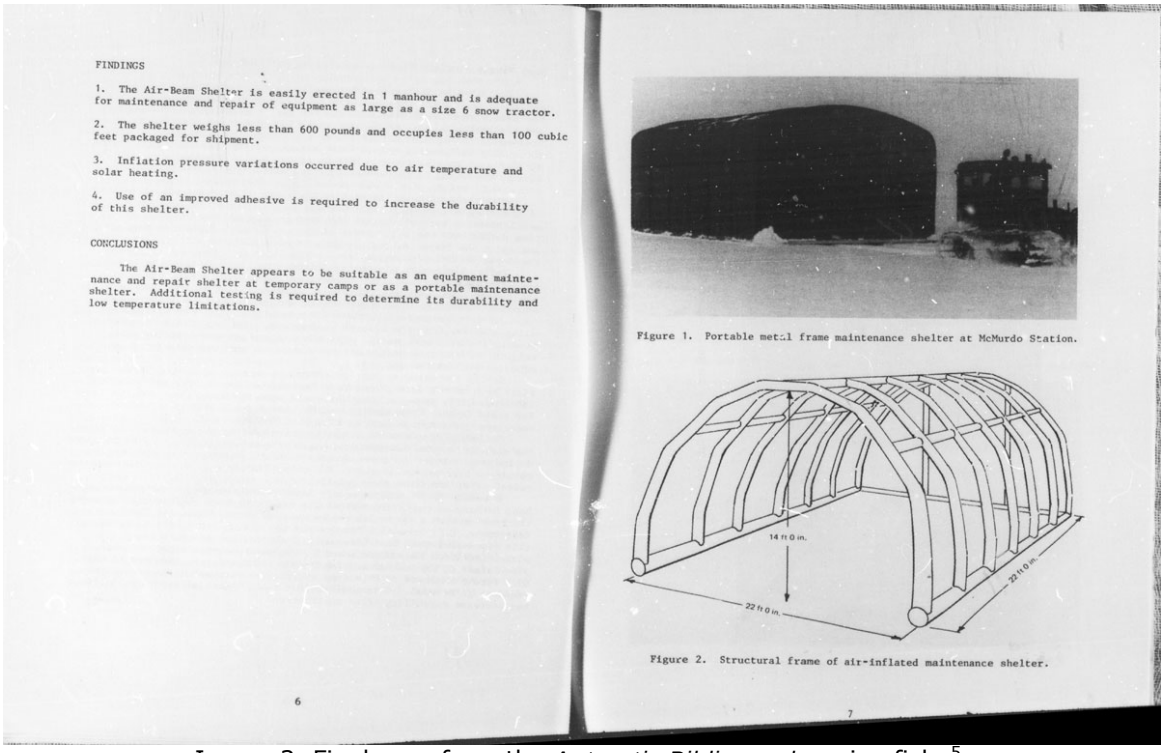


Image 1. A scan from the *Antarctic Bibliography* microfiche showing the textured background and skewed filming[5].

Rigid human quality control at the OCLC Preservation Service Center was required to remedy the problems with skewing and background noise. The resulting images were far better (Image 2).



FINDINGS

1. The Air-Beam Shelter is easily erected in 1 manhour and is adequate for maintenance and repair of equipment as large as a size 6 snow tractor.

2. The shelter weighs less than 600 pounds and occupies less than 100 cubic feet packaged for shipment.

3. Inflation pressure variations occurred due to air temperature and solar heating.

4. Use of an improved adhesive is required to increase the durability of this shelter.

CONCLUSIONS

The Air-Beam Shelter appears to be suitable as an equipment maintenance and repair shelter at temporary camps or as a portable maintenance shelter. Additional testing is required to determine its durability and low temperature limitations.

6

Figure 1. Portable metal frame maintenance shelter at McMurdo Station.

Figure 2. Structural frame of air-inflated maintenance shelter.

7

Image 2. Final scan from the *Antarctic Bibliography* microfiche[5].

Additionally, there was no standard for the number of pages included per frame on the fiche. Some fiche contained two images or pages per frame. Other fiche only contained one image per frame. Individual frame widths varied by several millimeters. Additionally, some fiche contained thirty-six frames per fiche, while others contained sixty to one hundred frames per fiche.

Ideally, we wanted a digitized image for each individual page, rather than a digitized image for each frame. However, after working with OCLC to find a solution, we realized we could not achieve one page per image with their automated process because of the issues and inconsistencies noted above. The OCLC automated process could produce a single TIFF per fiche frame up to 17" X 11". Because our goal was to create a quality digitized image that could be scanned using OCR technology for accessibility, we decided to forsake one image per page for now. We again recognized additional library staff time would be required to separate the pages of the digitized documents included in a single frame.

Results

OCLC delivered a hard drive containing a little over 2500 TIFF files containing one frame per TIFF. This translates to around 4500 pages and approximately 77 articles. The quality of our fiche images was poor, so OCLC advised against performing OCR and converting to PDF.

Conclusions

We learned much from our pilot. Sufficient time and a defined protocol for the selection of materials for digitization are critical. Although we spent a great deal of time evaluating the fiche and considering the best methodology, more time dedicated to this part of the process would have been beneficial. It is important to carefully assess the quality of the material to be digitized. Automated digitization systems have limitations and materials should be evaluated to determine if an automated process is a viable option. In our case, manual scanning would perhaps be more cost effective in the long run. Human quality control incurred higher costs over the course of the pilot, but proved to be crucial to achieving usable results. Because of this quality control process, digitization was halted so we could address issues with image quality and fiche composition.

Communication with peers in the field is vital to avoid duplication of effort. We were conscious of the need to avoid duplication of effort. The information regarding digitization projects and

availability of materials we received from AGI and other libraries was useful and helped shape the direction and focus of our pilot.

Future Work

Because the quality of the images is poor, we believe image preprocessing is required. OCR technology works best on clearly scanned pages without borders, since most of the content is composed of easily recognized characters. However, the fiche images often contain background content outside of the page itself, which OCR does not deal with gracefully. A multi-pass filter approach, which detects edges, normalizes the image, and removes noise, is an optimal solution. The resulting data analyzed by the OCR program would be much cleaner and easier to process. OCR performs the first step - digitization of the grey literature contained in the fiche. The next step is to transform that data into something searchable and machine readable. Conversion of the metadata into XML would be a good first step towards this goal. Unfortunately the fiche does not contain a good amount of metadata, so metadata creation must be a manual process performed by subject matter experts. The XML can then be used to create a taxonomy to classify the articles and data contained in them. To further augment this, we could create an ontology based upon the taxonomy to make the materials semantically available and searchable.

In the future, we hope to secure additional funding or partner with another organization to continue digitizing the rare and valuable full-text found in the *Antarctic Bibliography* microfiche. Once this grey literature has been digitized, processed, tagged with metadata, and cataloged, the NSF Library intends to collaborate with AGI to make the materials available free of charge to the public on the web. Ultimately, this project will result in increased access to the grey literature recorded in the *Antarctic Bibliography*, which in turn we hope will contribute to greater scientific discoveries, historical revelations, and better education about Antarctica and the Southern Oceans.

Acknowledgements

We respectfully thank Guy Guthridge of the National Science Foundation and Sharon Tahirkheli of the American Geological Institute for sharing their knowledge and experiences with us.

References

1. National Science Foundation (US). Investing in America's Future: Strategic Plan 2006-2011, Arlington, VA: National Science Foundation (US). (2006).

2. Office of Polar Programs. *Operation of the Antarctic Bibliography and the Bibliography on Cold Regions Science and Technology. Program Solicitation NSF 99-94*. Arlington, VA: National Science Foundation (US). (1994).

3. National Science Foundation (US). *Questions and Answers, Program Solicitation NSF 99-94*. Arlington, VA: National Science Foundation (US). 1999 Apr [Cited October 21, 2009] Available from: http://www.nsf.gov/pubs/1999/qa9994/qa9994.doc

4. Office of Polar Programs. Frequently Asked Questions about the U.S. Antarctic Program. Arlington, VA: National Science Foundation (US); 2008 Jul [Cited October 21, 2009]. Available from: http://www.nsf.gov/od/opp/frequent.jsp

5. Brier, F.W. *Air inflated portable maintenance shelter:* Technical note, N-1395. Port Hueneme, CA; U.S. Naval Construction Battalion Center. (1975).

I search

I find

I order

January 2010 – all scientific and technical information available in just 3 CLICKS

Over 35 million records of articles, books, reports and conference proceedings from 1847 up to the present

Refdoc.fr

DISCOVER ALL THE FEATURES AND FUNCTIONS AVAILABLE AT www.refdoc.fr

INIST DIFFUSION
2, ALLÉE DU PARC DE BRABOIS, 54514 VANDŒUVRE-LÈS-NANCY CEDEX, FRANCE → PHONE : +33 (0)3 83 50 46 64

Usage of grey literature in open archives: state of the art and empirical results[1]

Joachim Schöpfel, Charles de Gaulle University of Lille 3
and Hélène Prost, INIST-CNRS, France

Abstract
The purpose of our communication is to present first results from a current research on the development and usage of open archives in France. This study aims to gain empirical insight in usage patterns of freely available scientific items deposited in open archives, especially of non-commercial material, e.g. grey literature, mostly not distributed through other channels. We present a state of the art of published empirical data, standardization, research projects etc., together with a survey on the development and the usage of French open archives, based on open source methods and investigation.
The usage of grey literature in open archives is a recent field of professional and scientific interest. So far, little has been published on usage of open archives, and even less, on usage of deposited grey literature. Nevertheless, there are some promising new initiatives and projects and first empirical data. Our communication will combine review, quantitative and qualitative survey data and case study in order to provide a realistic insight into this emerging field.
Expected results: Empirical data allowing for first comparison between different archives and document types. Awareness on the scientific but also professional and economic interest of these data. A contribution to standardization (recommendations on data production, delivery and analysis).

1. Introduction

Grey literature represents a substantial part of the scientific production (Schöpfel & Farace, 2009). Since the Seventh International Conference on Grey Literature at Nancy in 2006, the GreyNet community intensified its research activities on the impact of the open access movement on the grey literature.

The purpose of this communication is to provide a follow-up study to our 2008 evaluation on the integration of grey literature in French open archives (Schöpfel & Stock, 2009) that described "a landscape in movement", with a significant increase of university institutional repositories supported by the academic consortium COUPERIN.

We considered that "the impact of grey material (…) in open archives is real and will stay", with an overall part of 17% of the deposited items. On the other hand, our survey revealed three major problems:

"(1) Policy statements need improvement. Often, the strategy and positioning of repositories are not explicit or simply missing.

(2) Especially grey items in open archives need improved bibliographic control. Compared to traditional cataloguing standards, metadata for grey material are less specific or again, simply missing. This is a problem for referencing, efficient search strategies and evaluation.

(3) Mostly wanted are detailed usage statistics on access and download of documents and other items in open archives."

The 2009 follow-up study surveys data on the development of the archives, e.g. evolution of deposits, and investigates usage statistics. The GL11 communication provides preliminary results from ongoing empirical and statistical analyses.

The study was funded by the Charles de Gaulle University of Lille 3. Special thanks to the professional team of the academic library of the Lille 1 campus, partner of the research project, and to Chérifa Boukacem-Zeghmouri for her contribution and advice.

2. Methodology

We basically applied the same approach as last year. The 2009 survey includes 150 representative (e.g. registered either with a dedicated platform or as data provider for harvesting) French digital repositories. The different archives were selected through 19 French and international registries of open access repositories or service providers (see appendix), between May and June 2009, and followed a defined set of criteria (located/hosted in France, living archive, size>0).

Each registered archive (URL) was checked; errors (incorrect URLs etc.) and duplicates were eliminated. Information about the remaining archives were incorporated into a spreadsheet with 58 data columns in 5 categories:

1. General (background) information about the archive.
2. Specific information about the archive.
3. Content information.
4. Qualitative data.
5. Comments.

Usage data were collected separately.

3. Results

Our communication will concentrate on three empirical results:

- The development of French open archives, e.g. evolution of size and content.
- The development of grey items in these archives.
- The dissemination and content of usage data.

The following discussion includes a short state of art of international projects of standard usage data in open repositories. The communication ends with some recommendations for usage assessment and gives an overview of future research.

3.1. Development of open archives in France 2008-2009

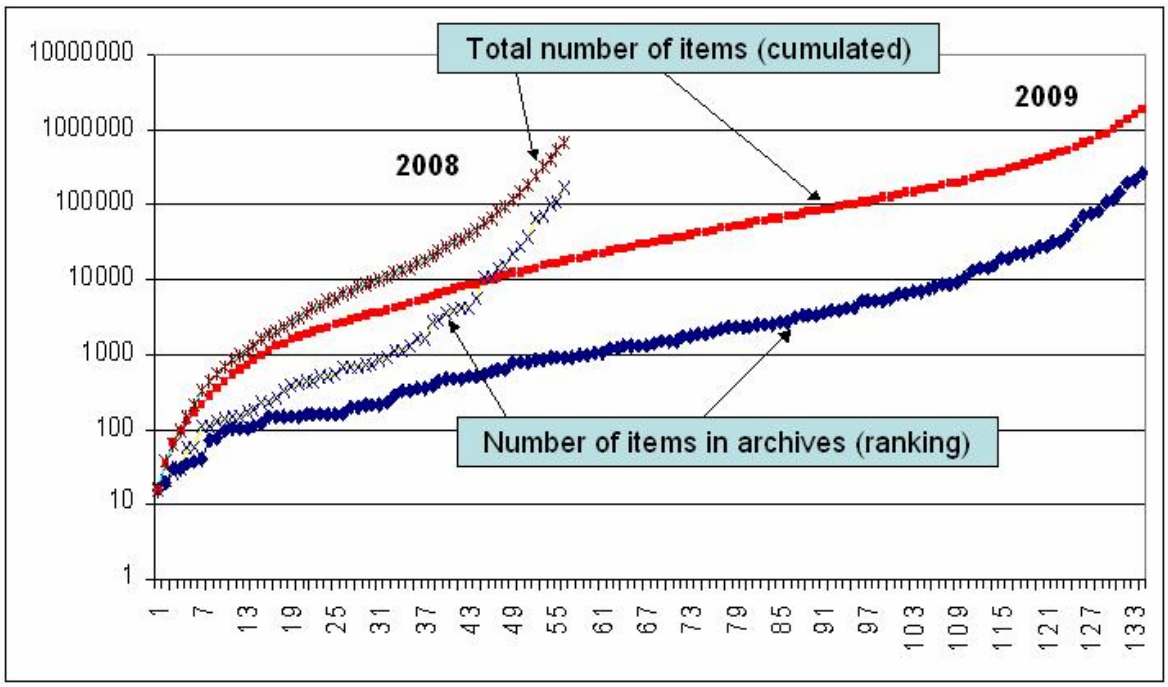


Figure 1: Number of items 2008 and 2009

The total number of referenced open archives in France has increased since last year. While we identified 56 repositories in 2008, their number attained 150 in 2009 (+168%).

In the same time, the total number of deposited items increased from 703,178 to 1,878,520 (see Figure 1). Especially the number of small and medium-sized repositories increased rapidly, with an item number between 100 and 10,000.

Nearly 60% of these archives are institutional repositories, mostly hosted by universities and other HE structures. About 15% are thematic archives (one discipline or one subject), another 15% contain only one category of documents (mostly electronic theses or dissertations, ETDs). The rest are mixed or heritage repositories.

On 103 sites the author or his laboratory can deposit his documents online. Other referenced repositories seem to be rather normal websites, digital libraries or portals, without any characteristics related to the open access movement.

Roughly one third of the documents are journal articles (pre- or post-print), another third are grey material or datasets. The rest are different types of documents, for instance heritage items, or cannot be correctly identified. Following these figures and information found on the web sites, we can estimate that more than 60% of the repositories have some kind of quality control or other validation procedure – significant more than last year.

3.2. Grey literature in French open archives

74% of all referenced repositories contain grey literature. In fact, nearly all institutional repositories and most of the document-specific archives contain grey material.

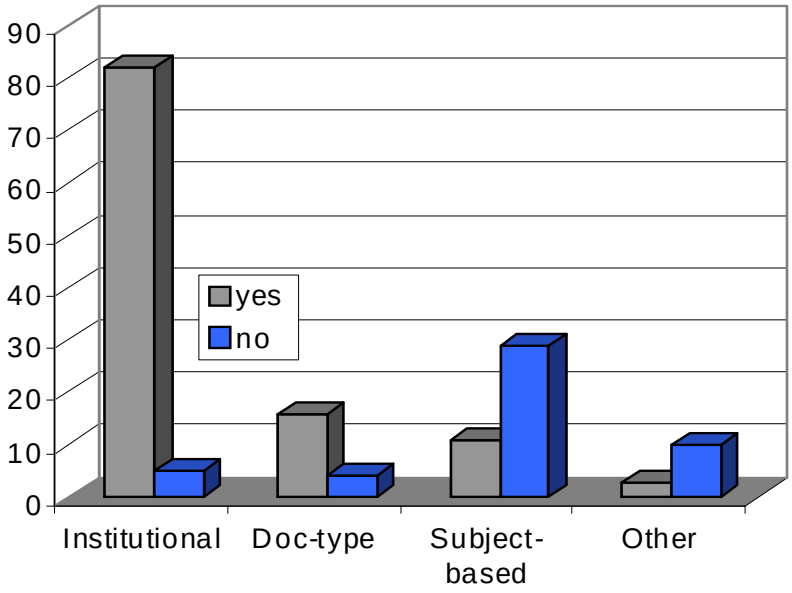


Figure 2: Grey literature in different kinds of archives

A deeper analysis of this material reveals the following figures (Table 1).

Type of grey literature	Number of items
Theses and dissertations	70,488
Reports	36,186
Conferences	157,257
Working papers	6,637
Courseware	2,875
Other	43,308

Table 1: Grey literature in repositories

The most important part of grey literature are conferences (proceedings, communications etc.), followed by theses and dissertations and, at a lower level, by reports. Nevertheless, an important part cannot be correctly identified ("unpublished work" etc.).

The reason is twofold. On the one hand, not all open sites allow for an advanced search on document types. Another reason are missing and/or non standard (non comparable) metadata. Without metadata, the survey becomes tough – one should access each item in order to define or guess its category. Mission impossible for more important repositories.

In fact, nearly 70% of the repositories – a little bit more than in 2008 – provide specific metadata for grey literature. For most of them, this means in particular some specific data for theses (for instance, the university) or for conferences (location and date of the conference).

Figure 3 shows the development 2008-2009 of some more important and well-known French repositories. Except PERSEE – the national repository for the back files of French SS&H journals – all important repositories contain grey literature, and for all we observe a relative more important increase of the part of grey literature than for the overall number of items. The future may show if this is a stable or a transitory evolution.

For instance, HAL the central open archive for academic publishing[2] arrives in 2009 at about 60,000 grey items. The institutional repository of the national centre for agronomic research INRA holds between 40 and 50,000 grey items – significantly more than in our 2008 study.

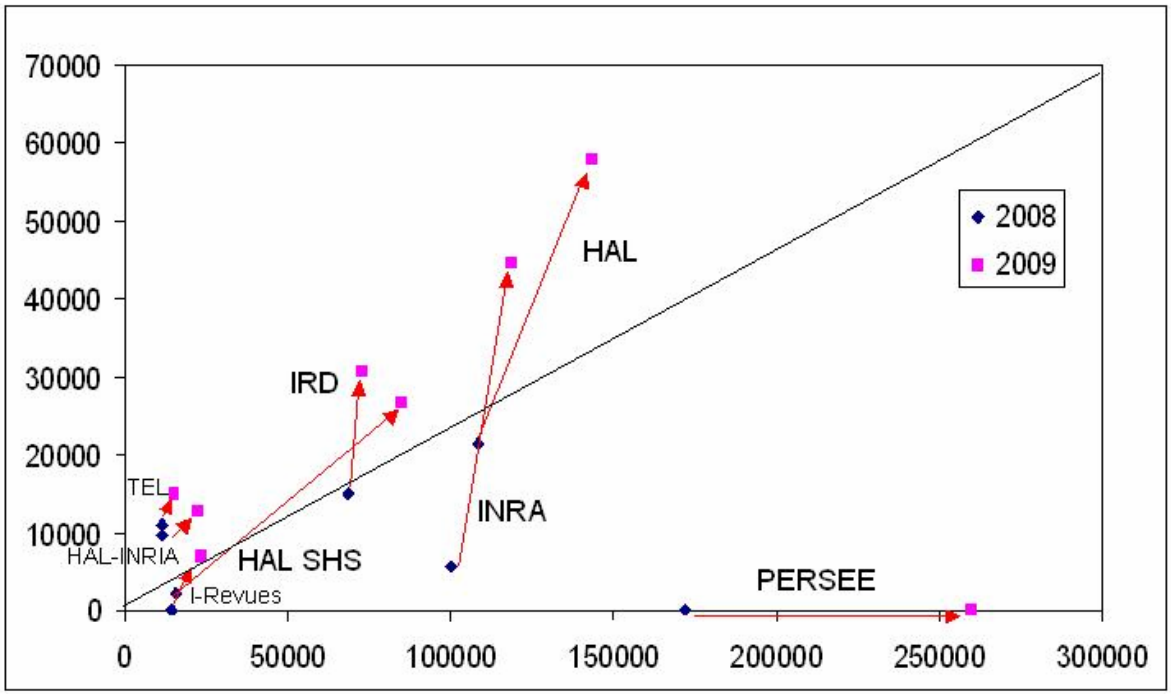


Figure 3: Development of some selected repositories (left side: grey items, below: overall number of items)

Even the CNRS plat-form for open e-journals, I-Revues, began to upload non-commercial material, e.g. conference proceedings.

3.3. Usage statistics

Our survey identified only a small number (7%) of repositories that publish usage statistics. Sometimes "usage" is misinterpreted in terms of upload figures, instead of access and downloads statistics. In the following we provide a short overview on some statistics related to grey literature.

The statistics of the University of Toulouse repository OATAO distinguish between published articles and electronic theses. The difference is significant: the average downloads number of

articles is 49 per item while it is 109 for theses, e.g. the grey (unpublished) material is around 2,2 more often requested than journal articles (Malotaux, 2009).

The IFREMER repository allows for comparison between monthly downloads of articles, reports, conferences and theses (see table 2).

	2007	2009
Articles	10	7
Theses	70	33
Reports	30	10
Conferences	n.d.	9

Table 2: Average monthly downloads (IFREMER repository)

All types of grey documents are more often requested than published articles. And at least for this repository, the most interesting materials for visitors are the electronic theses that can't be accessed elsewhere (Merceur, 2007 and 2009).

The statistics of the University of West Britanny (Brest) confirm that the electronic theses are more often downloaded than other items. 17 of the 25 most requested documents are theses (Bertignac & Gac, 2009).

Other statistics include the list of the requested theses (INP Toulouse) or provide information on activity metrics, information seeking and user characteristics without details on grey literature (ParisTech).

4. Discussion

The 2009 follow-up study encountered the same kind of methodological problems than the 2008 survey (see Schöpfel & Stock, 2009) - the overall number of items in some repositories is difficult to define, and identifying grey content remains a challenge. Also the architecture of the HAL repository (or HAL system) is complex, and one part of deposits is stored in two or three of the HAL archives. Nevertheless, our survey data allows for some more general statements.

4.1. The changing nature of open archives

Open repositories are meant for publications. Yet, 16% of all items in French repositories are metadata (records) without full text, and 25% of the repositories in our survey contain at least 50% simple records. A comparison with US, UK or German repositories shows that this is not particular for French open archives. Paradoxically, the reason for this evolution may be linked to the success of institutional repositories: as they are increasingly integrated into evaluation of institutions and scientists, hosting structures (libraries etc.) started to upload metadata even if the documents, for legal or other reasons, are not available.

Another evolution non-conform with the initial goals of the OA movement is the important part of scientific and/or national heritage items in institutional and other repositories. At the beginning, open repositories were created to foster and speed up direct scientific communication (cf. the role arXiv always plays for the high energy physics community). 30 sites (20%) in our survey hold such kind of content, with some of the most important archives (PERSEE for SS&H journal back files, the French national library's GALLICA and the history of art repository of INHA). All these sites are referenced as open archives. Are they, really? How should we fix the difference, for instance, between HAL with current publications and NUMDAM with back files in Mathematics? In 2009, around 46% of the deposits seem to be heritage items. Should we revise the definition of open repositories?

A third problem is the restricted access to some repositories. Basically, open archives stand for general and unrestricted access to scientific information. In reality, 12% of the repositories

limit access to their content in some way or other. For some sites, visitors have to register; for others, access is restricted for non-institutional user.

4.2. Metadata and usage statistics

The part of repositories with specific GL metadata slightly increased (68%, +2% compared to 2008). But at least 15 open archives with grey materials lack specific metadata. Identifying different types of grey documents is often difficult. Sometimes, the only distinction made is between "published" and "unpublished" papers. In other archives, the GL categories appear to be rather ad hoc categories without any precise or standard definition.

Without metadata, a detailed and in-deep analysis of the repository content is impossible; also, without metadata, usage statistics cannot be linked to the content in terms of document types, institutions (affiliations), disciplines etc.

Nevertheless, the small number of (published) statistics confirms results from international repositories such as RePEc[3] where grey documents – unpublished working papers – are consistently more requested than articles (see figure 3).

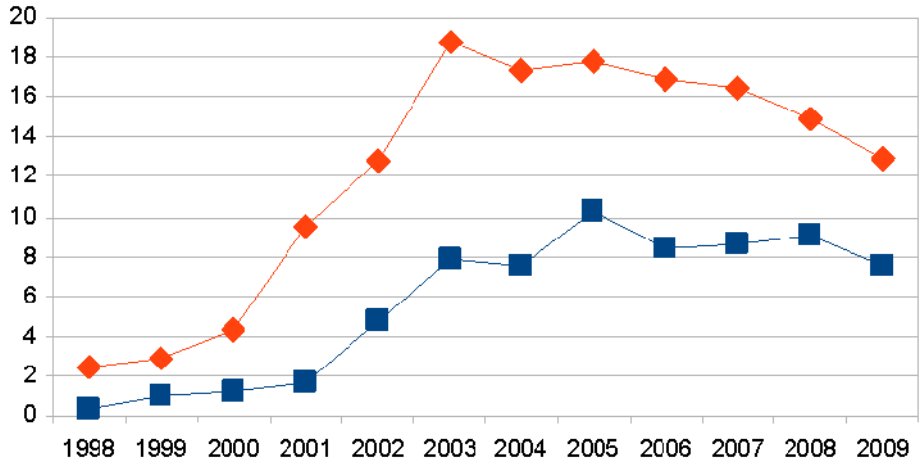


Figure 4: Average annual downloads per document type in RePEc (red: working papers; blue: articles)

Yet, more surveys are needed to confirm this statement, provide more detailed evidence and also help to understand the reason of this difference.

4.3. Recommendations

We already observed that only few repositories made their statistics available. And those who do so use different tools, methods, terms. In some way, this reminds the early period of digital libraries when publishers produced either non-standard or no usage statistics at all.

Nevertheless, as a recent report settled, "researchers want tangible, immediate benefits such as download statistics (…)" (Fry et al., 2009). In the following, based on our own survey and on other projects (JISC PIRUS[4], DINI OA-Statistik[5]) we suggest some recommendations meant to foster standard developments of open repositories.

1. Recipient: Usage statistics should be useful to all groups involved in the functioning of open repositories, e.g. usage data should be made available to authors, users (visitors), and institutions.
2. COUNTER principle: Usage statistics should be defined at different levels, with increasing complexity and a basic minimum level that corresponds to the COUNTER Journal Report 1.
3. Log files analysis: A selection of a minimum data set for a basic log files analysis should be defined. These data should cover the whole range of potential information (visitor,

content, request type, date, unique identifier).
4. Terminology: The usual terms and wordings of usage statistics, log files and open repositories should be clearly defined and if possible, translated in French (access, downloading, visit, request, hit…).
5. Provision: Reports must be provided monthly following the COUNTER rules. Data must be updated within four weeks of the end of the reporting period. All of last calendar year's data and this calendar year's to date must be supplied.
6. Metadata: Usage of full text and metadata (records) must be clearly distinguished.

The usage data should be disseminated in en environment of added value services that take into account the specific needs of the different groups (authors, visitors, institutions).

"Value-added services such as download statistics, email alerts, etc would contribute to the perceived usefulness of repositories and would help them gain popularity in what is an increasingly competitive information landscape." (Fry et al., 2009).

The development of such services should include at least four key elements:[6]

- Modular statistics (collections, document types, time period etc.).
- Summary tables.
- Assistance-help online / FAQ.
- Link with other tools measuring the impact of deposited items (citations, tagging etc.).

These elements would help authors, visitors and institutions to evaluate the impact, popularity and quality of stored content.

5. Conclusion

Our communication presented preliminary results of an ongoing study on the development and usage of French open repositories.

The total number of referenced open archives in France has largely increased since 2008, from 56 to 150 in 2009 (+168%).

74% of all referenced open archives contain grey literature, especially institutional repositories. Yet, an important part of grey items cannot be correctly identified because of missing search options and/or metadata.

Only a small number of repositories publish usage statistics (7%). The available data confirm that grey documents are more often requested than articles.

We added some recommendations for the development of usage statistics in the open repository environment, following in particular the model of the project COUNTER.
In the next months, our research team will apply the log file analytical approach to the institutional repository of the university of Lille 1, IRIS[7] (the former *Grisemine* website, the first "grey" archive in France and presented at the Nancy conference), in order to illustrate the state of the art and the survey results. In the same time, we will elaborate a glossary of terms specific to usage statistics of open repositories. Probably, this part of work will be done together with the DINI team.

6. Bibliography

Bertignac C, Gac D (2009). La voie verte: les archives ouvertes. In: Open Access Week. 19 October 2009. Institut Universitaire Européen de la Mer.

http://www.eur-oceans.eu/WP3.1/OA/LaVoieVerte-Catherine\%20Bertignac.pdf

Fry J, Oppenheim C, Probets S, Creaser C, Greenwood H, Spezi V, et al. (2009). PEER Behavioural Research: Authors and Users vis-à-vis Journals and Repositories. Baseline report. LISU Loughborough University.

http://www.peerproject.eu/fileadmin/media/reports/Final_revision_-_behavioural_baseline_report_-_20_01_10.pdf

Merceur F (2007). Gestion d'une archive et d'un moissonneur, l'exemple de l'IFREMER. In: RPIST 2007, 20 juin 2007.

http://rpist.inist.fr/sites/rpist/IMG/pdf/archimer-2.pdf

Merceur F (2009). Optimiser la visibilité de vos dépôts, au vu des statistiques d'Archimer. In: Open Access Week. 19 octobre 2009. Institut Universitaire Européen de la Mer.

http://www.eur-oceans.eu/WP3.1/OA/OptimisationVisibilite-FredMerceur.pdf

Schöpfel J, Farace DJ (2009). Grey Literature. In: Bates MJ, Maack MN eds. *Encyclopedia of Library and Information Sciences*. 3rd edition. Taylor & Francis.

Schöpfel J, Stock C (2009). Grey literature in French Digital Repositories: A Survey. *The Grey Journal*, 5(3), pp. 147-161.

Malotaux S (2009). OATAO Archive ouverte multi-établissements. Bilan après un an d'existence. In: Journées d'études sur les archives ouvertes. Consortium COUPERIN. 2 et 3 avril 2009.

http://journeesao.wordpress.com/2009/04/19/archives-ouvertes-et-regroupement-d\%E2\%80\%99etablissements-le-cas-des-ecoles-d\%E2\%80\%99enseignement-superieur-toulousaines-sandrine-malotaux/

References

[1] Acknowledgement to C. Boukacem-Zeghmouri for her helpful advice and contribution to data collection and analysis.
[2] HAL ranks second in the list of world repositories *Webometrics* http://repositories.webometrics.info/top400_rep_inst.asp
[3] http://repec.org/
[4] http://www.jisc.ac.uk/whatwedo/programmes/pals3/pirus.aspx
[5] http://www.dini.de/projekte/oa-statistik/
[6] See for instance the metrics of the Public Library of Science PloS at http://article-level-metrics.plos.org/
[7] https://iris.univ-lille1.fr/dspace/

From CNR Annual report to an Institutional repository: Results from a survey

Roberta Ruggieri, Rosa Di Cesare, Daniela Luzi, and Loredana Cerbara
National Research Council, Institute of Research on Population and Social Policies, Italy

Abstract
Considering that the collection policy is one of the most important elements in the development of Institutional Repositories (IR), the general aim of this study is to contribute to the development of a future CNR's IR identifying and proposing GL collections and related metadata. In this paper the results of a qualitative analysis of a sample of GL document types stored in a database containing CNR scientific production are described. The analysis of both the metadata schema and information content of the note fields allowed the identification of database improvements, which are represented by the introduction of fields describing GL characteristics as well as the provision of guidelines that help information providers to correctly insert bibliographic data.

1. Introduction

In 2008 a CNR "OA supporter group", composed mainly of CNR librarians, promoted a survey to acquire a more precise picture of Open Access practices at CNR, one of the largest Italian multidisciplinary research organizations [Luzi, 2008; Di Cesare, 2009]. One of the main results was that the questionnaire respondents identified in a central database, containing CNR scientific production, a potential building block of an IR.

This database derives from the CNR Annual report and contains the description and outputs of its research activities. It is online updated and contents are directly added from the CNR research units involved, which provide descriptions of projects carried out, information on internal and external funding, personnel and equipment involved, collaboration with other national and international institutions, etc. A subset of this information is available at the main CNR webpage (http://www.cnr.it/istituti), where publications are linked to both the research units and to the projects within which they are produced. Moreover, all publications produced by CNR researchers available since 2002 are organised according to a predefined list of document types that include both conventional and non-conventional literature.

The positive attitude towards this central database shown by questionnaire respondents, its comprehensiveness as well as the involvement of researchers and librarians in the process of providing information, motivated us to further investigate its information content and quality. Therefore, our aim is to analyse the bibliographic collections - both conventional and non conventional – reflecting the multidisciplinary CNR nature.
Moreover, CNR is currently developing an Institutional Repository (IR) to diffuse its research production/scientific output. Considering that the collection policy is one of the most important elements in the development of institutional repositories, the general aim of the study is to contribute to the development of the future IR, identifying and proposing GL collections and related metadata. In particular, the analysis intends to verify whether the actual metadata are suitable to clearly identify and correctly describe GL document types. This can contribute both to identification of a standardized metadata schema and to development of guidelines supporting the correct insertion of information.

In this paper we present the results of this study on a sample of selected GL documents produced by a selected CNR research units. Paragraph 2 describes the methods used to select the sample examined and perform the qualitative analysis of the records stored in the database. A general overview of the development of this database is given in paragraph 3, reporting a summary of an interview with the person in charge of its management. The database bibliographic collections are illustrated in paragraph 4, while the results of the qualitative analysis of the selected sample of GL documents is reported in paragraph 5.

2. Methods

2.1. Survey design

The object of our analysis is the information content of the database currently used to collect CNR scientific products. To perform this analysis we used a stratified random sample of CNR research units and documents. In addition we used an in-depth interview with the person

responsible for the management and implementation of this central database, to obtain information on its development as well as its main purposes.

2.2. Sample

The sampling methodology of the survey was a two-stage stratified sample design, in which:

- *in the first stage*, we selected a stratified random sample of 21 units, representing 20% of the universe, composed by 107 CNR research units. At this stage the survey units are Institutes.
- *in the second stage*, we selected a random sample of GL documents proportional to the number of GL documents produced by the selected units on which we applied a qualitative analysis. At this stage the survey units are GL documents.

The criteria followed to choose the sampled research units in the first stage were:

- CNR institutes were identified as survey units, because they are the main providers of the bibliographic information inserted in the central database;
- CNR institutes were selected proportionally to the frequency of the Institutes belonging to each Departments, so that they reflect the CNR's multidisciplinary nature.
- CNR institutes were randomly selected, irrespective of the number of people in charge or of the amount and type of their scientific production;

The criteria followed to choose the GL sampled documents (second stage) in order to carry out a qualitative analysis were:

- Selection of publications produced in the period 2003-2007. This period coincides with scientific assessment made at CNR by a group of external experts, who could access this database. So our hypothesis is that the information providers have focused on the data input.
- Selection of the following GL document types: *Conference papers*; *Oral presentations; Reports; In-house publications.* We excluded the following GL documents, *Project results* and *Project applications,* because of their recent insertion in the database, irregular temporal coverage and limited number of descriptions.

Table 1. shows how we defined our sample. In the first column the total number of CNR institutes is reported, broken down by Department. In the second column we report the number of Institute sampled. The total number of GL documents produced by the selected Institutes is reported in the third column, while in the last one the number of GL document analysed in the pre-test is given.

Table 1. Sampled institute and documents

Departments	*Research units*	*Sampled of research units*	*Total number of GL produced by sampled units*	*Pre-test on GL sampled documents*
Earth & environment	13	3	1985	53
Materials & devices	12	3	1064	34
Molecular design	14	3	1059	43
Cultural identity	15	3	952	57
Medicine	12	2	983	24
ICT	7	1	558	19
Agriculture & food	10	2	1342	38
Advanced manufacturing systems	7	1	415	14
Energy and transportation	6	1	262	8
Cultural heritage	5	1	281	15
Life sciences	6	1	36	6
Total	**107**	**21**	**8937**	**311**

As management of this type of sample is quite complex, we decided to have a pre-test phase where we selected randomly a fixed number of documents for each document type as well as for each selected research units concerning the period examined. The pre-test sample size is

the 311 GL documents. This pre-test is essential to verify the variables used and eventually identify new ones to be applied to the second stage sample, which foresaw the extraction of about 500 documents.

2.3. GL documents qualitative analysis

In order to verify whether the metadata chosen for GL collections were suitable for their identification and bibliographic description, we have analysed for each GL document type:

- The metadata schema adopted in the current database and in particular the predefined mandatory and optional fields;
- The frequency of the use of optional fields made by information providers;
- The information content and frequency of use of the *note field* made by the information providers.

This analysis allows us to evaluate the record's quality in terms of completeness, that is a criterion to measure the fullness of single GL records. Moreover we analysed the information contents of the note field in the hypothesis that this field was compiled when the information providers were not able to find suitable fields to describe specific bibliographic GL characteristics. In our view the *note field* is a "*control key*", information content analysis of which can provide important indications of future improvement in the metadata schema for GL collections.

Data were collected directly from primary source namely the CNR central database. As the database is online updated, the analysis concerns data gathered in September 2009.

2.4. Interview

The interview was held with the person responsible for the *Information Service* appointed to the design and development of CNR information system related both to central administration and integration with territorial research units. An expert of the database under survey was also involved to this face-to-face interview.

The interview focused on the following aspects:

- General information of the central database (called Istituti@cnr) related to its mission and technical features, its interaction with other managerial databases;
- Data model adopted and procedures for data input, ranging from the actors involved in the management of the information content to the analysis of data required;
- Prevision of the development of a future CNR's IR.

3. The CNR database described by the interviewee

The interview provided important background information on the central database under survey, which is briefly reported in this paragraph.
The interview confirmed that this database represents the evolution of pre-existing systems managing CNR scientific publications. It derives from the previous CNR *Annual report,* issued since 2002 to document its scientific activity. This database, currently known as . Istituti@cnr, was created in 2003 together with the revision of CNR website. Its aim is to provide uniform and comprehensive information on all CNR Institutes available at the central CNR website, while each Institute can decide how to deliver this information on its website autonomously.
Moreover, this database was also developed to manage the CNR budget, and more recently its function has been extended to support scientific evaluation activities. The assessment purpose of this database is evident in the information required by input procedures, as we have verified in our survey.

The interview was also focused on other technical aspects. Istituti@cnr software was developed by the internal staff of the Information Service, to be interoperable with the CNR personnel registry, with the research activity planning system and with the module managing the CNR budget.

The database containing CNR scientific products has been constantly updated adding new required services and input modules, for instance a new functionality to insert the full-text of the scientific publication has been recently developed, even though full-texts are not yet available to external users. Moreover, the input is facilitated by links to predefined lists of

options, such as the connection to the ISI-Indexed Journals and related Impact Factor, or to the authors authority file. However, no guidelines related to a correct bibliographic description of the documents have been provided. The interviewed persons are aware of this problematic aspect and stated that for the development of the CNR's IR they are planning to collaborate with librarians, starting from the CNR Central Library. They also agreed that Istituti@cnr is a good starting point for the development of an IR, it being already OAI-PMH compliant. They are currently working on the workflow aspects of the repository, paying particular attention to validation procedures. Concerning the validation of the bibliographic input, they foresee an active role played by librarians. Finally they hoped that CNR would adopt a policy for a mandatory submission of CNR scientific outputs.

4. The bibliographic collections in Istituti@cnr

In order to provide an overview of the bibliographic collections stored in the central database, we have analysed CNR scientific production in the period 2003-2007, broken down into departments and document types (table 2). About 90,000 publications are stored in the database, for the majority of them only bibliographic descriptions are provided.

Table 2. Distribution of document types in CNR central database by Department (2003-2007)

Department	*Total number of publications*	*Journal articles*	*Books and book chapters*	*Conf papers*	*Oral presentations*	*Reports*	*In-house publications*	*Other*
Earth & environment	16757	32,2	8,5	16,6	27,9	10,7	1,4	2,7
Energy and transportation	3529	35,6	2,0	27,7	18,1	13,6	0,3	2,7
Agriculture & food	6963	36,0	6,6	17,2	28,2	6,8	0,9	4,3
Medicine	8207	50,0	3,2	6,1	36,4	2,2	0,3	1,7
Life sciences	1980	67,7	2,8	1,4	25,9	0,9	1,2	0,2
Molecular design	10514	56,9	2,8	9,2	24,8	3,1	0,3	2,7
Materials & devices	15292	64,9	2,9	12,3	13,9	3,0	1,0	2,1
Advanced manifacturing systems	5019	29,2	3,3	19,8	15,0	19,5	1,2	11,9
ICT	7023	30,9	5,4	29,4	9,8	17,5	3,2	3,8
Cultural identity	9262	22,6	24,0	13,5	20,4	10,9	5,3	3,3
Cultural heritage	2154	21,8	18,4	30,8	15,0	7,8	1,5	4,6
Total	***86700***	***42,4***	***7,1***	***15,6***	***21,7***	***8,4***	***1,6***	***3,3***

Considering document types, 42% are journal articles, 7% are books and book chapters. As regard specifically GL collections: 21% are *Oral presentations*, about 16% *Conference papers*, 8.4% *Reports*. The last category also includes working papers, technical reports and so on. More than 1% is *In-house publications* and about 3% includes other types of documents, such as data set, maps, etc. In the table GL collections are represented according to their level of greyness, the darkest ones are the most difficult to describe, as we have verified in the analysis on the GL sample examined.

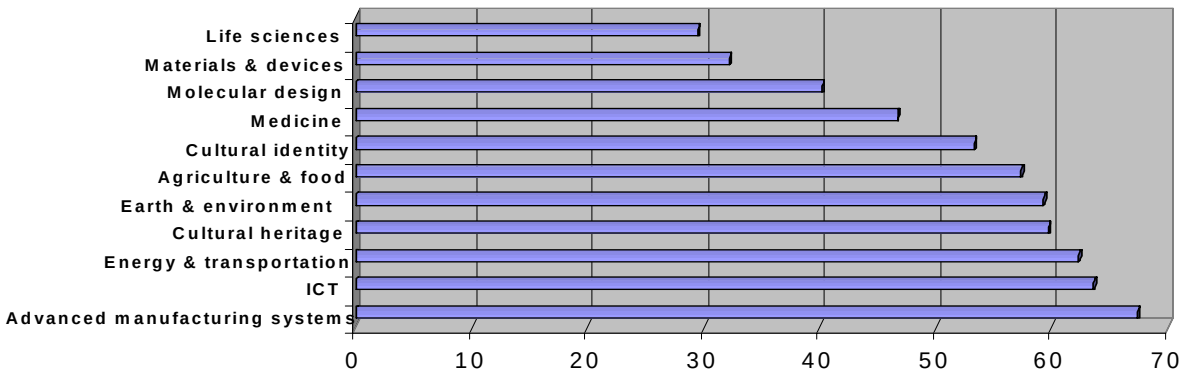


Fig. 1: Percentage of GL documents by Department (2003-2007)

Figure 1 shows the percentage of GL documents broken down by CNR Departments. GL is produced by all CNR Departments with a percentage above 50% for more than half of Departments.
The highest value (67%) is produced by the Department of Advanced manufacturing systems, while the lowest is represented by GL produced by the Department of Life sciences. This research output is mainly connected with the publication models of different scientific communities and is in line with other studies [Katz 1999, Hiks, 2005].

5. Results of qualitative analysis

5.1. The metadata schema

The first step of our qualitative analysis was related to the identification of the metadata schema used to describe GL document types.

Table 3. –Mandatory and optional fields by document type

Fields	Conference papers	Oral presentations	Reports	In-house publications
Mandatory				
Item types	√	√	√	—
Title	√	√	√	√
Author	√	√	√	√
CNR author (ID)	√	√	√	√
Research project	√	√	√	√
Type	—	—	√	√
Scientific editor	—	—	—	√
Publication year	√	√	√	√
Optional				
Author's affiliation	√	√	√	√
Conference title	√	√	—	—
Conference location	√	√	—	—
Notes	√	√	√	√
CNR discipline classification	√	—	√	—
CIVR discipline classification [1]	√	—	√	—
Abstract	√	√	√	√
Further information	√	√	√	√
Key words	√	√	√	√
References	—	—	—	√

[1] Disciplinary classification adopted by the National Agency of Evaluation (CIVR).

Table 3 shows mandatory and optional fields by document type. Besides the common general mandatory fields (such as *Author*, *Title*, *Publication year* etc.), other mandatory metadata present in all GL document types relate the author with the research project he/she is carrying out and with the publication produced within the project. (*CNR Author*, *Research project*). Moreover each CNR author has his/her own unique identifier, and this is fundamental for the construction of authority files. Predefined lists of values are contained in the mandatory fields *Item types* and *Types*. Referring to *Conference papers* and *Oral presentations* these predefined values are reported for evaluation purposes, specifying whether the paper is presented at international or national conferences, if it is an invited paper or if the proceedings are indexed

by ISI. In *Reports* and *In-house publications* these values indentify sub-categories of document types. Given that these collections include a great variety of document types, the meaning of these values sometimes overlaps or is ambiguous and this may cause problems for information providers.

Among optional fields, particular emphasis goes to the identification of publication content. Two different subject categories are used, the first of which has been developed within CNR and reflects the disciplinary fields covered by CNR research activities, while the second is the classification of the National Agency of Evaluation. Moreover, the information provider can also add free key words to describe the publication content.

5.2. Completeness of records

The frequency of use of optional fields was used to analyse the quality of records, in terms of completeness, that is the number of compiled optional fields for each GL document type. For each document type we grouped the optional compiled fields into 4 ranges, from the records that do not have any optional compiled field to the ones that have been compiled using the majority of available optional fields (range 6-9).
Table 4 shows the number of optional fields used by information providers to describe GL documents produced by their research units.

Table 4. Document types by number of compiled fields

Compiled fields	*Conference papers*		*Oral presentations*		*Reports*		*In-house publications*	
	No.	%	No.	%	No.	%	No.	%
0	2	2,3	--	--	14	18,7	10	17,2
1 - 2	23	26,1	54	60,0	27	36,0	31	53,4
3 - 5	51	58,0	36	40,0	30	40,0	17	29,3
6 - 9	12	13,6	--	--	4	5,3	--	--
Total	*88*	*100,0*	*90*	*100,0*	*75*	*100,0*	*58*	*100,0*

The majority of records that have a high level of completeness is concentrated in the range 3-5: 58% *Conference papers*, 40% Oral presentation and *Reports*. *Vice versa Reports* and In-house-publications have the highest percentage of no optional compiled field (respectively 18,7%, 17,2%) and they are also the document types with a high percentage of records in the range of 1-2 optional compiled fields, together with *Oral presentations* (60%). The completeness of optional field compiled pertains only *Conference papers* (13,6%) and *Reports* (5,3%) in a small percentage of records.

5.3. Analysis of the note fields

As mentioned above, Note fields have been used as a *control key* to evaluate whether the actual metadata were adequate to describe a specific GL document type. Generally the note field is used to include information that cannot be inserted in any other field. To analyse in detail their information content we have classified it under two categories:

- *Bibliographic information*, i.e. bibliographic elements that could not be inserted in any other field, but were necessary for a correct description and document identification;
- *Additional information*, i.e. the traditional use of the note field to insert relevant information not including standardized bibliographic elements.

Making this distinction we could identify whether there were *missing* metadata necessary to correctly describe bibliographic characteristics for each GL document types examined.

Table 5 - Note field compiled by document type

Document types	Note field compiled %
Conference papers	26
Oral presentations	20
Reports	52
In-house publications	47

105 out of the 311 records examined contained compiled note fields (33,8%). In particular *Reports* (52%) and *In-house publications (*47%) are the GL document types, which have the highest percentage of filled in note field (tab. 5).

Table 6. Use of note field by content information type:

Note field content	%
Bibliographic information	71
Additional information	29

Table 6 shows the result of the analysis of the note field content classified in two categories. The majority of records (71%) uses note field to provide bibliographic information.

5.3.1. Use of note fields in Conference papers

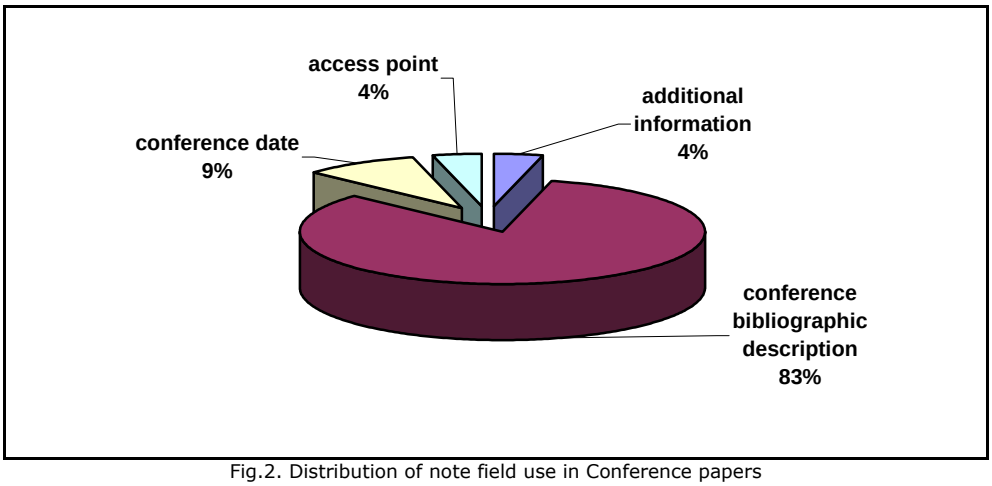


Fig.2. Distribution of note field use in Conference papers

26% of the analysed records describing *Conference papers* contained a compiled field note (fig. 2). The majority (83%) of them used the note field to provide conference bibliographic description related to the conference proceedings where the paper was presented, such as scientific editor, publisher, year of publication, page numbers. 9% of note fields are used to insert conference dates, for which a separate field was not foreseen. This field is useful especially when the conference date and the publication year do not coincide and should be mandatory information together with that related to the *Conference title* and *location* that are currently optional. Only 4% of record uses the note field to insert additional information.

5.3.2. Use of note fields in Oral presentations

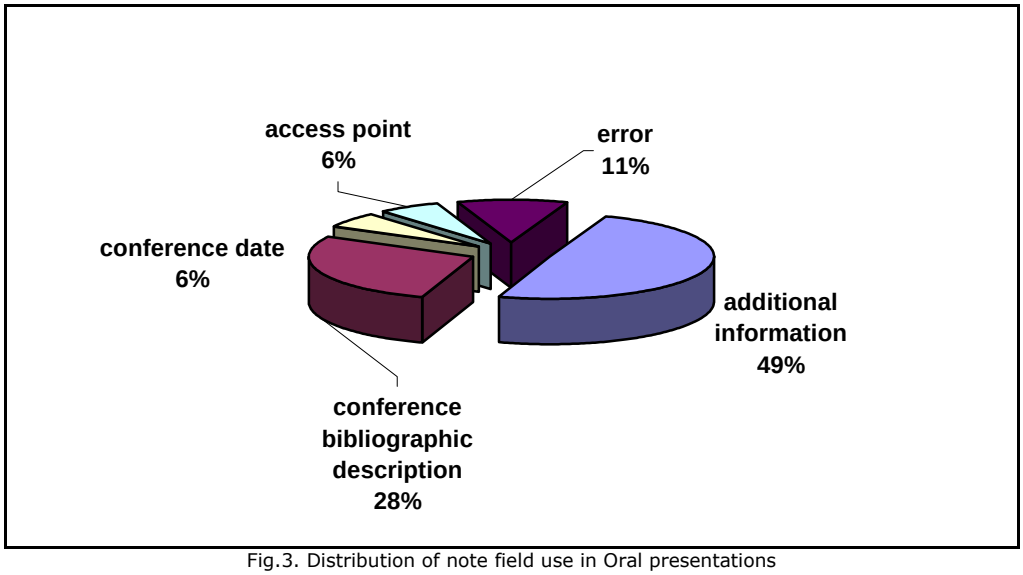


Fig.3. Distribution of note field use in Oral presentations

Compared to *Conference papers*, a smaller percentage (20%) of records describing oral presentation has a compiled note field. Further the percentage distribution is quite different from the previous one: only 28% of the note field is used to insert bibliographic description,

while 49% of the note field is used to provide additional information, such as, PPT slides, video, exhibitions. Considering that this category includes a variety of document types, ranging from posters to conference papers available only on conference websites, it would be useful to provide a more precise definition of this category.

5.3.3. Use of note fields in Reports

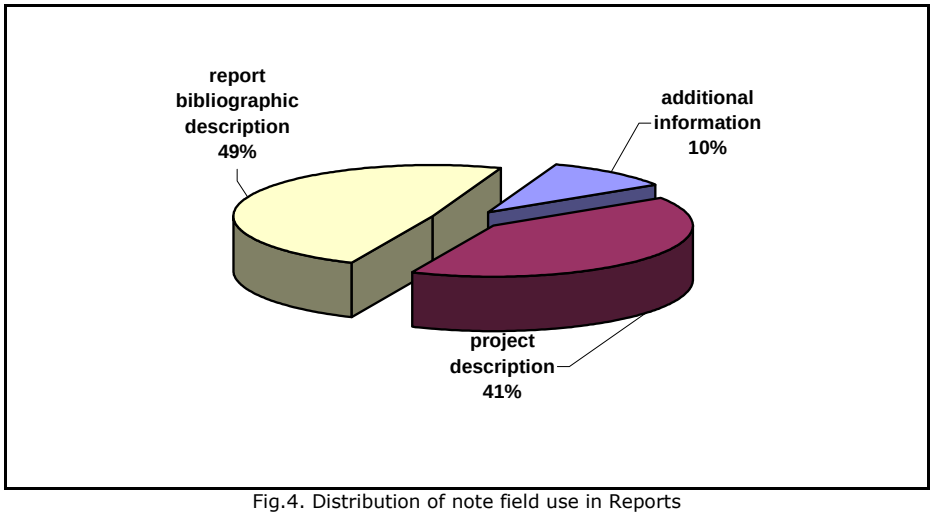


Fig.4. Distribution of note field use in Reports

52% of records describing *Reports* contains a compiled field note. Almost 50% of them are used to give a correctly report bibliographic description and 41% are used to provide information on the project during which reports are produced. The bibliographic information often pertains to important report bibliographic elements such as the report number and corporate source and they are often provided in a standardised form. This homogeneous description may depend on librarians providing the input to this database.

5.3.4. Use of note fields in In-house-publications

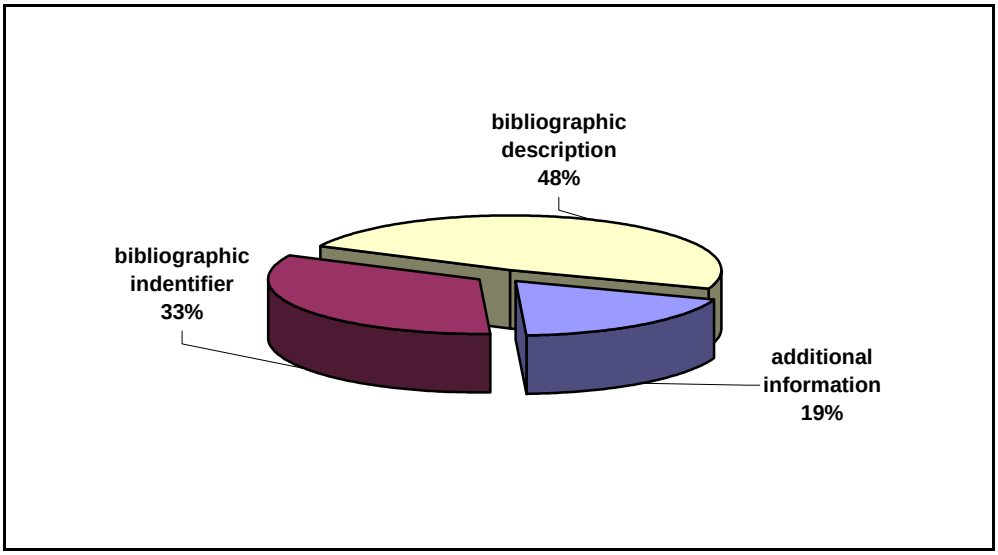


Fig.5 Distribution of note field use in In-house-publications

This category groups all the scientific results directly published by CNR, ranging from monographs to serials, catalogues and databases. Moreover this category also includes excerpts from journals, and editorials. The broad definition of this category may create some confusion on the document type to insert in this category.
Given that a limited number of both mandatory and optional fields are foreseen to describe a great variety of document types, almost half of the records analysed, 47%, contains a compiled field note.

Most of them are used to provide bibliographic descriptions, 48%, while 33% of them contains bibliographic identifiers such as ISSN, ISBD. In this case too this information may come from an expert hand.

5.3.5. Missing fields

The analysis of the contents of the note fields allowed us to identify fields that could be added to the current metadata schema in order to better describe and identify GL document types. The results of this analysis are summarized in table 7. *Conference papers* and *Oral presentations* should first consider the conference title and location as mandatory fields, also adding information on the conference date. Moreover, a more precise description on the conference proceedings is needed providing information on the scientific editor and publisher. Considering *Reports* it is important to add a field containing the report number and series title, as well as a field containing information about project title and number and contract number. The category In-house-publications should be more precisely defined, identifying a limited group of document types which reflect the CNR editorial production as well as the related bibliographic field to correctly describe and identify them.

Table 7. Missing fields identified by document types

Document types	Missing fields
Conference papers	▪ Conference date ▪ Conference proceedings bibliographic description (title, editor, publisher)
Oral presentations	▪ Conference date ▪ Conference proceedings bibliographic description (title, editor, publisher)
Reports	▪ Report series and Report number ▪ Project description (project title, project number, contract number)
In-house publications	▪ Specific bibliographic description fields related to document types

6. Conclusions

The analysis of the database that currently collects CNR scientific production was based on a selected sample of records produced by a representative number of CNR institutes. Its metadata schema that was progressively updated over the last few years reflects the initial purpose of this database, designed above all for managerial and evaluation puposes as a derivation of the CNR Annual report. The main characteristics of this database are its integration with other internal databases, the evident connection of authors with research projects descriptions and publications produced, the special emphasis given on data related to evaluation (peer-reviewed, ISI publications, invited paper, etc.).

According to other studies [Griscom 2006, Palmer 2008] one of the parameters to evaluate an IR is the *critical mass* of documents stored and diffused together with the quality of bibliographic collections. Our analysis shows that this database currently contains about 90.000 documents that are a good starting point to build IR information content.
The weak point of the database is connected with the lack of metadata for a correct description and identification of GL documents, which represent an important part of the collections stored therein. This was the result of our qualitative analysis focused on the information content of note field compiled in the selected number of GL documents. This field was compiled more frequently when the collections become *more grey* and contained important bibliographic elements that could not be inserted in any other field of the related metadata schema. Through this analysis we have identified fields that could be added to the current metadata schema in order to better describe and identify GL document types improving the quality of GL bibliographic collections.

Acknowledgments

We would like to thank Dr**.** Maurizio Lancia responsible of the CNR Information Service and his collaborator Angelo Olivieri, for agreeing to the interview and taking the time to answer to our questions.

References

Bui Yen, Park Jung-ran (2006). An assessment of metadata quality: A case study of the National Science Digital Library Metadata Repository. URL:http://www.cais-acsi.ca/proceedings/2006/bui_2006.pdf

Ki-Tat Lam, Diana L.H. Chan (2007). Building an institutional repository: sharing experiences at the HKUST library. *OCLC Systems & Services*, 23 (3), p. 310-323.

Cees de Blaaij (2004). Grey literature from invisibility to visibility. *Publishing research quarterly*, 20:1, p. 70-76

Di Cesare Rosa, Luzi Daniela, Ruggieri Roberta, Cerbara Loredana (2009). Per un Institutional repository del Consiglio Nazionale delle Ricerche. Indagine sulle esperienze Open Access nelle strutture scientifiche dell'Ente. *IRPPS Working papers* no. 27. URL: http://www.irpps.cnr.it/sito/download/wp27_09.pdf

Ferreira Miguel, Ana Alice Baptista, Eloy Rodrigues, Ricardo Saraiva (2008). Carrots and sticks: some ideas on how to create a successful Institutional repository. *D-Lib Magazine*, January-February.

Griscom Richard, Hassen Marjorie, Steiner Mary D., Kerbel Sandra (2006) *Content recruitment and development: a proactive approach to building an Institutional repository*. EDUCASE conference in Dallas, 11 October 2006. URL:http://repository.upenn.edu/cgi/viewcontent.cgi?article=1011&context=library_papers

Hiks Diana (2005). The four literature of social sciences. Journal of management of social sciences, 1 (1), p. 1-20

Katz J. Sylvan (1999). Bibliometric indicators and the Social Sciences. University of Sussex. URL: http://www.sussex.ac.uk/Users/sylvank/pubs/ESRC.pdf

Lambert Simon, Matthewz Brian, Jones Catherine (2005). Grey literature, institutional repositories and the organizational context. In: 7. International Conference on Grey Literature (GL7), Nancy, France, 5-6 December 2005

Luzi Daniela, Di Cesare Rosa, Ruggieri Roberta, Cerbara Loredana (2009). Towards and Institutional repository of the Italian Research Council: a survey on Open access experiences. *International journal on grey literature.* 5 (1), p.35-47.

Lynch Clifford A., Lippincott Joan K. (2005). Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine*, 11(9), p.1-9.

Moen William E., Stewart Erin L., McClure Charles R. (1997).The Role of content analysis in evaluating metadata for the U.S. Government Information Locator Service (GILS): Results from an exploratory study. URL:http://www.unt.edu/wmoen/publications/GILSMDContentAnalysis.htm

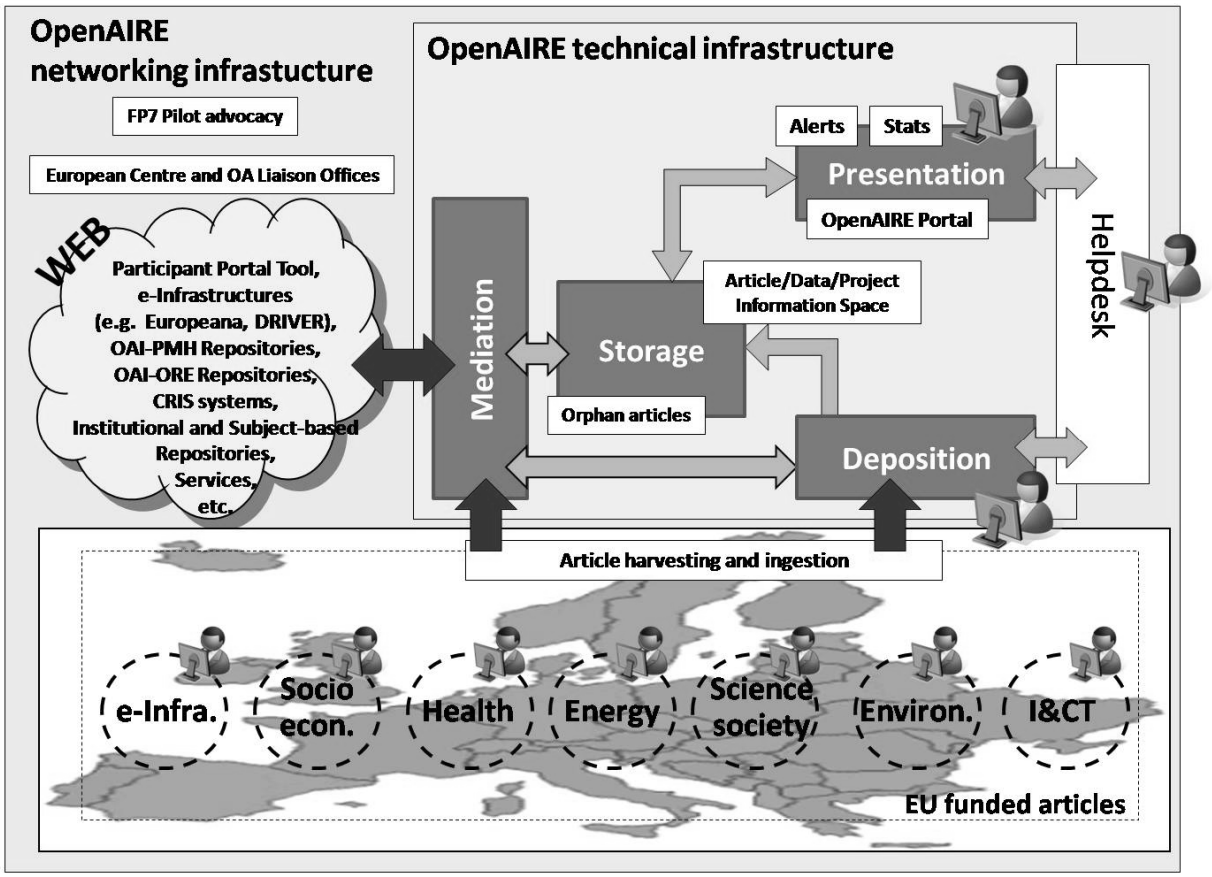
Palmer Carole L., Teffeau Lauren C., Newton Mark P. (2008). Strategies for Institutional repository Development: a case study of three evolving initiatives. Purdue libraries. : URL: http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1131&context=lib_research

Suber Peter A. (2008). A study of Institutional repository holdings by Academic discipline. *D-Lib Magazine*, 14(11/12). Yeomans Joanne (2006). CERN'S open access E-print coverage in 2006: three quarters full and counting. *High energy physics libraries webzine*, Issue 12/March 2006. URL:http://library.web.cern.ch/library/Webzine/12/papers/2/

The OpenAIRE Project
Open Access Infrastructure for Research in Europe

Stefania Biagioni, Donatella Castelli, and Paolo Manghi
Institute of Information Science and Technologies, ISTI-CNR, Italy

OpenAIRE (Open Access Infrastructure for Research in Europe) [1] will deliver (*i*) *technological infrastructure* for the identification, collection (by deposition and harvesting), access and monitoring of FP7 and ERC funded articles and (*ii*) a *networking infrastructure* for the uptake of the EC Open Access's mandate and OpenAIRE's system across authors and institution in Europe. Dissemination will be achieved by means of the establishment and operation of the OpenAIRE European Helpdesk System and a network of National correspondent.
In particular, the project targets the following objectives: (*i*) it will offer a special repository, called "the orphan repository", for the ingestion of metadata and document of the articles whose authors do not have an institutional or subject-based/thematic repository of reference (*ii*) it will enable deposition and OAI-PMH harvesting policies of metadata of the articles whose authors have a repository of reference, (*iii*) it connect all collected article metadata with EU project contracts metadata, (*iv*) it will monitor the system use to obtain statistically-significant trends about projects, articles and authors, and (*v*) it will provide the the OpenAIRE portal (www.openaire.eu) through which all collected and inferred data will be freely accessible worldwide.
Thematically, the project will focus on peer-reviewed publications (primarily, journal articles in final or pre-print form, but also conference articles, when considered important) in at least the seven disciplines highlighted in the Open Access Pilot in FP7 [2] (energy, environment, health, cognitive systems-interaction-robotics, electronic infrastructures, science in society, and socioeconomic sciences-humanities) and on research datasets in a subset of them. Geographically, however, it will have a definitive "European footprint" by covering the European Union in its entirety, engaging people and scientific repositories in almost all 27 member states and beyond. The technological infrastructure built by the project will be based on state-of-the-art software services of the D-NET package developed within the DRIVER and DRIVER-II projects [3] and the Invenio digital repository software developed at CERN [4]. These will be further enhanced and complemented with services developed within OpenAIRE to address critical requirements and issues that arise in the target environment and require further investigation. For example, monitoring tools and statistics services will infer relevant information and statistics on FP7 and ERC funded research from article, research data, project metadata and relationships between them. Further, OpenAIRE will work with several subject communities to explore the state of the art of research datasets management and their combination with research publications.



Project information

- Title: Open Access Infrastructure for Research in Europe;
- Project n°: RI-246686;
- Project type: CPCSA;
- Start date: 1st of December 2009;
- Duration: 36 months;
- Funding: 4,17 Million;
- Coordinator: Michael Hatzopoulos , University of Athens, Greece;
- Scientific coordinator: Norbert Lossau, University of Goettingen Library, Germany;
- Technical coordinator: Donatella Castelli , ISTI - National Research Council, Italy;

With the collaboration of 27 National Open Access Liaison Offices covering all European member states.

References

1. http://www.openaire.eu/index.php
2. ftp://ftp.cordis.europa.eu/pub/fp7/docs/open-access-pilot_en.pdf
3. http://www.driver-repository.eu/
4. http://cdsware.cern.ch/invenio/index.html

Open Access to Grey Literature on e-Infrastructures: the BELIEF Project Digital Library

Stafania Biagioni, Donatella Castelli, and Franco Zoppi
Institute of Information Science and Technologies, ISTI-CNR, Italy

Abstract

The BELIEF (Bringing Europe's eLectronic Infrastructures to Expanding Frontiers) Project is a Coordination Action funded by the European Commission in the context of the FP6 and FP7 Programmes. It aims to create a platform where e-Infrastructures providers and users can collaborate and exchange knowledge. This will help ensuring that e-Infrastructures are deployed and effectively used worldwide, filling the gap separating the Research Infrastructure providers from the users, and thus contribute to the emergence of a competitive knowledge-based economy.

To create this synergy among multi-disciplinary Research Infrastructure communities, BELIEF created a one-stop-shop for e-Infrastructures communities providing a Community Portal http://www.beliefproject.org/ and a Digital Library (DL) http://belief-dl.research-infrastructures.eu/ with a huge number of e-Infrastructures open access documents. Moreover, it has organised events including brainstorming, networking workshops and international conferences and publications, since BELIEF's values are firmly rooted in international cooperation with the emerging economies, particularly in Latin America, India and South Africa.

The BELIEF DL – implemented on top of the OpenDLib Software System – offers uniform access to multimedia documentation and especially to grey literature (e.g. presentations, videos, technical reports, manuals, on-line tutorials, etc.), providing continuously updated information on e-Infrastructures-related projects, initiatives and events. The contents are harvested from different sources, such as projects web sites, repositories and databases. The DL provides services to support the submission, description, searching, browsing, access, preservation and visualization of these multimedia documents. Although designed to meet the needs of a specific scientific community, the technology adopted by BELIEF can be easily adapted to meet the information and collaborative needs of other user communities.

The BELIEF e-Infrastructures Community ranges now on more than 80 projects, initiatives and organisations inside and outside Europe, and the DL offers nearly 15.000 documents harvested from their repositories and websites.

The BELIEF Consortium is composed by Brunel University (UK), CNR-ISTI (Italy), ERNET (India), Escola Politécnica da Universidade de Sao Paolo (Brasil), Meraka Institute (South Africa), Metaware SpA (Italy), and the National Kapodistrian University of Athens (Greece).

1 Introduction

The BELIEF (Bringing Europe's eLectronic Infrastructures to Expanding Frontiers) Project aims to create an effective open workspace where e-Infrastructures providers and users can collaborate and exchange knowledge, ensuring the development and adoption of e-Infrastructures on a worldwide scale. The BELIEF DL play a key part in the project, bringing a range of benefits to e-Infrastructures stakeholders across the globe by facilitating the exchange of knowledge and experiences through a single and easily accessible tool. The BELIEF Project arose from the awareness that a gap existed between Research Infrastructure providers and users. In order to bridge this gap, a complete and common source of information on e-Infrastructures was needed, both for users demanding provision and resources and for providers intending to extend their user base and develop their systems. The BELIEF DL responds to this demand by providing users with documentation matching their search criteria accurately and according to their interests and professional profile. This paper focuses on the implementation of this key component.

General outcomes of the design phase are reported in [7]. In the following (Section 2) the main characteristics of the DL are described. General concepts regarding the implementation of the DL are outlined (Paragraph 2.1). These general concepts represent the logical link between basic requirements and the components of the implemented solution (Paragraph 2.2). Then (Paragraph 2.3) the organization of the User Interface is introduced.
Section 3 summarizes the most relevant usage data as per the statistics gathered during the whole projects lifetime, from 2006 onward.

This paper won't cover all of the relevant topics of the implementation work, for those please refer to [4]. Hereafter are summarized the ones covered in this paper (Section 4):

- Definition of the metadata structure implemented by the DL.

- Definition and implementation of protocols and tools for the submission of metadata and documents to the DL.
- OAI-PMH Compliance.
- Integration with external portals.

Conclusions and results achieved are finally outlined in Section 5.

2 The Digital Library

The DL was designed to serve the needs of e-Infrastructures research and industrial users that want to keep up to date with existing projects and the latest developments in e-Infrastructures. The DL offers its user community advanced services to uniformly access multimedia documents such as technical reports, presentations, videos, manuals, on-line tutorials, etc. These documents contain the very latest details on e-Infrastructures related projects, initiatives and events.

The material maintained in the DL is harvested from different sources, such as web sites, repositories and databases of e-Infrastructures Projects, Initiatives and Organisations. The DL organises the harvested information according to the information needs of the user communities rather than according to its physical format, structuring and distribution on diverse sources. This means that it is capable of providing users with multiple virtual views of the existing documentation. To this end, an extremely accurate collection and analysis of the requirements of potential users was made before the DL was created.

The DL provides services to support the submission, description, searching, browsing, retrieval, access, preservation and visualization of multimedia documents. Users can define the information space which they want to search/browse upon in terms of collections (i.e. sets of documents) selected from those managed by the DL. Collections can be created interactively, based on the archives the documents are to be selected from. Different search/browse options are offered: Google-like or fielded (with fields selected from a variety of known metadata formats), with or without relevance feedback. Users can search/browse any information associated with digital documents and their parts.

2.1 Preliminary Design - Methodology

A design phase based on a deep analysis of the Project's target audience needs was carried out with a selected set of Research Infrastructure Entities (Projects, Organisations and Initiatives). The methodology adopted for requirements collection was basically structured in the following activities:

- Drawing up a Questionnaire to collect requirements on functional requirements, documents to be managed and their related metadata (descriptive data).
- Identifying proper interfaces within Entities, contacting them, obtaining and discussing requirements.
- Matching similar requirements from diverse Entities.
- Verifying the quality of the metadata and making the semantics and use of metadata conform.
- Verifying the quality of documents, analysing the types of documents used by diverse Entities with different semantics and making them conform.
- Drawing up a Memorandum of Understanding to be signed with each of the Entities collaborating in subsequent releases to ensure clear, comprehensive and effective interaction and collaboration.

2.2 DL's General Concepts

This section introduces the main concepts characterising the DL from the perspective of its end-users. These concepts represent the logical link between the basic users' requirements and the components of the solution implemented, as follows:

Users of the BELIEF **DL** access the DL's **Resources** according to their **Role** and to the **Access Rights** they hold on **Collections**. These are grouped in user-specific **Information Spaces**. A **User Profile** is then associated with each User, recording his/her characteristics in terms of Role and Access Rights.

Where:

Users Are the actors entitled to interact with the DL. Different types of Users can be provided.

DL From a high-level perspective, a DL is an organisation capable of collecting, creating, accessing and managing Resources.

Resources The resources that can be accessed by Users, depending on their Role and Rights. Diverse types of Resources can be provided (**Content** and **Services**)
Resources of type Content are gathered from **Information Sources**.

Information Source This term is used to indicate a repositoriy belonging to an Entity (Project, Organisation or Initiative) from which metadata and documents are collected to form the DL. Information Sources are characterised by the type, structure and semantics of the metadata and documents they contain.

Role A function within the context of an organisation (in our case the BELIEF DL) with some associated semantics regarding the authority and responsibility conferred to the user. Different types of Roles can be provided (**Consumers**, **Providers** and **Librarians**).

Access Rights The actions allowed on Resources (Create, Modify, Delete, Search/Browse, Use/Perform, etc.).

Collection A set of Resources of type Content defined according to characterisation criteria. Collections represent the mechanism to organise Content in order to provide focused views of the Information Space, and allow DL end-users to access to thematic parts of the whole Information Space.

Information Space Corresponds to the **set of Collections of the DL** that each user can access. It defines the scope of Content against which the Browse/Search operation are executed. The Information Space can be dinamically modified by the user to restrict/enlarge such scope.

User Profile Contains information about the user. It may be viewed as a set of characteristics with associated values. Each piece of user information alone and in conjunction with others determines system behaviour towards the user.

2.3 General Characteristics of the Implemented Solution

The BELIEF DL is created as an instance of the OpenDLib [4] Software System, thus being capable of offering the following services to its users:

- Creation, submission, search, browse, access, and preservation of multimedia documents.
- Definition of the their personal Information Space which they want to search/browse upon in terms of collections selected from those managed by the DL. Collections can be interactively created defining:
  - the condition that is satisfied by the members of the collection;
  - which archives the documents are to be selected from.
- Different search/browse options: Google-like or fielded (with metadata elements selected from a variety of known metadata formats). Users can search/browse any information associated with digital objects and their parts. As a result of their search/browse operations, users obtain a set of result pages with the list of digital objects that satisfy their request. By clicking on an object, users can access any of its multiple manifestations. In particular, they can select the one that is compatible with the software installed on the computer they are working with.
- Full compliant Open Access via the OAI-PMH protocol.

In the implementation of the DL particular effort has been devoted to the following issues:

- Implementation of harvesters supporting standard formats and specific conventions on Information Sources side. Presently, the following standard protocols and coding formats are supported by the harvesting tools of the DL:
  - Metadata encoding formats: DC, DCQ (recommended), MARC, UNIMARC, MARC21, MARCXML.
  - Metadata harvesting protocols: OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting (recommended) or any API call returning an XML file containing metadata encoded in one of the above mentioned formats.
  - File formats: XML (recommended), RSS.
- Implementation of "document models" and of a web-based interface allowing users to easily submit metadata and documents to the DL using those models.
- Implementation of a set of APIs to be easily used by programmers to interface basic functions of the DL.
- Harmonization of concepts and practices (e.g. use of metadata and terms and of different types of document) to be benefited by the whole Community in facilitating knowledge communication and document exchange.

- Definition and/or homogenisation of vocabularies used both for metadata and document content description, categorization and search.

The benefits coming from the use of the DL are twofold:

- On the Information Provider's side:
  - e-Infrastructures projects' documents can be accessed by a wider users' community.
  - This wider audience is also a more targeted audience. The DL provides users with the most accurate selection for their information needs, in a way that bridges them directly with the projects' material.
  - It gives projects' audience a one-stop-shop to find the precise documentation that they require about the projects, thus it is a valuable target to send inquirers to, should they be searching quickly for projects' documentation.
  - Documents are accessible via different formats.
  - The DL supports document preservation.
  - Projects can submit their documents directly to the DL. BELIEF offers a submission interface and a document hosting service. To ensure that any new documents can be easily collected by the DL a standard method is provided to describe them.
- On the Information Consumer's side:
  - While accessing a website and browsing a project's material, users get material from that project/website only, using the DL they get the relevant information from many projects' sites in a single shot. This give them a wider range of information on e-Infrastructures, how they work, how to use them, etc..
  - User can search documents in a uniform way, accessing via a unique interface documents produced by different providers.
  - The material accessible is extremely well focused on the users' needs. This is not like Google which may often come up with nothing or thousands of results or it yields material that is not relevant or up to date.

2.4 The User Interface and its usage

The DL's User Interface (UI) has been designed to reflect the most recent advances in UI usability:

- The look and feel let users have an extremely comfortable access to functionality and content.
- The overall navigation structure has been designed to minimise the number of clicks the users need to access any content.
- All most common functions can be easily accessed via one-click commands.
- Most of the relevant information related to a document are shown in the same window.
- A complete On-line User Guide has been implemented by means of a friendly Wiki capability.
  - The overall work area of the DL (the *DL Desktop*) is clearly organised into three sections [2]:
- On the left side is the "Community" section. According to each specific account permissions – after having executed a Login – a number of functions are accessible here: "News", "Personal Profile", "Users & Groups Management", "Documents Management", and "Information Space Management".
- On the right side is the "Content Access" section, where the content access functions (Browse, Search) are accessible. A "News" section is also shown, presenting relevant information about the DL from different points of view (events, user services, technical, etc.).
- In the middle is the "Information Space" section of the Desktop where available Collections are listed and the content resulting from Browse and Search operations is shown.

The following figures show respectively the overall look of the UI, the definition of a search criteria and the way for accessing the actual content of a document – once it has been retrieved via a Browse/Search operation. Please refer to [3] and [6] for details, or access directly [2].

Figure 1 – The DL Desktop organization

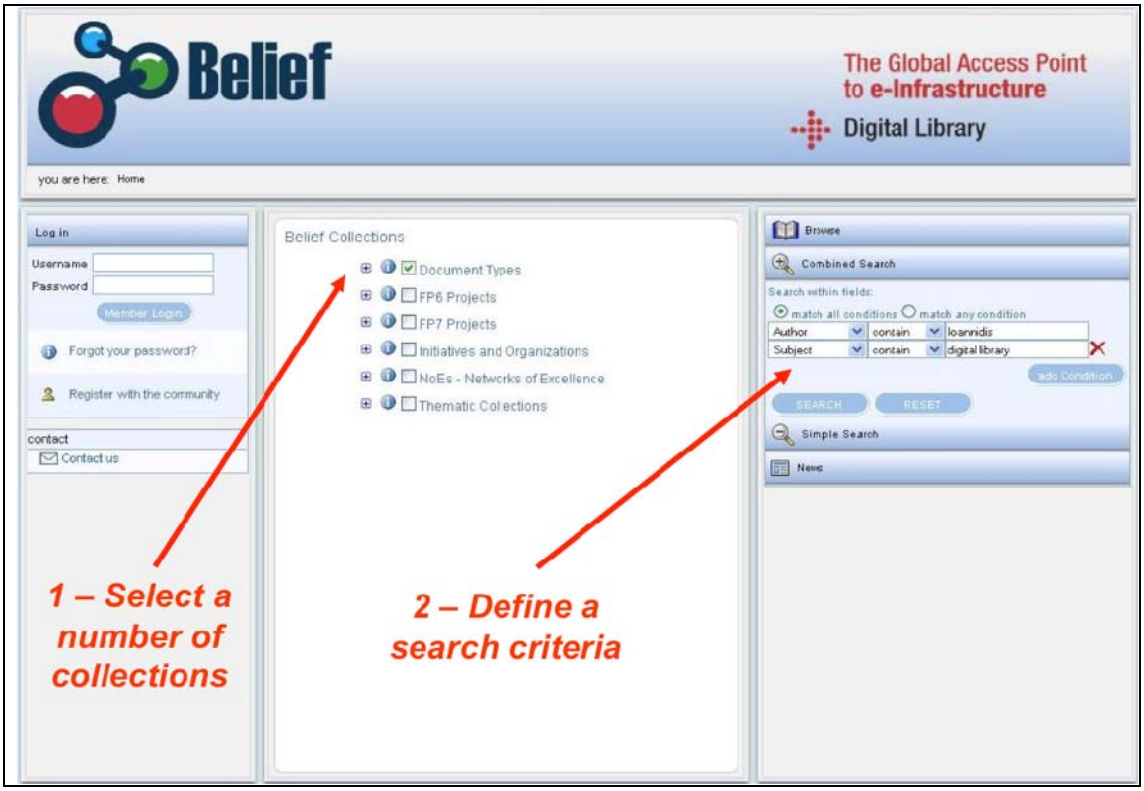


Figure 2 – Searching content

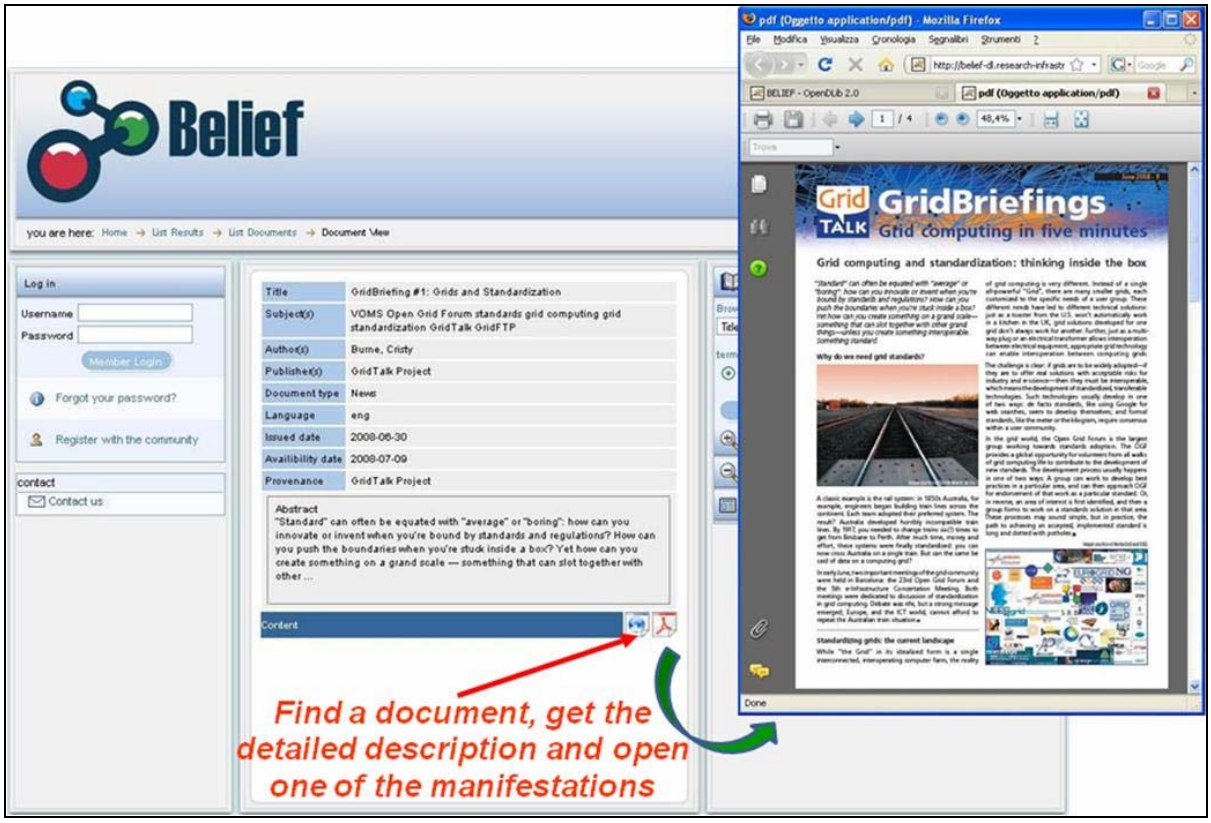


Figure 3 – Accessing content

3 BELIEF Events and Community's DL growth and usage

BELIEF successfully organised a series of events which broke new ground in the arena of Grid-empowered infrastructures and e-Infrastructures. These events brought together Grid & e-Infrastructures experts and technology developers, IT innovators from both enterprise and research, decision-makers, and scientific policy-makers. Participants came together to share ideas and knowledge, discuss how technological challenges could be tackled, strengthen alliances between business and research, and help unleash the potential of e-Infrastructures.

All of these international events have helped to setup a synergetic network of relationships and to valuably increase the content of the Digital Library with material submitted by all Community's Entities. Thanks to the outcomes of these events and to the established synergies and relationships, a significant growth of the Community has been achieved, largely exceeding the original planning of the project.

The rich and differentiated content offered by the DL is reported by the following chart, with a clear prevalence of material produced by Conference and Technical Meetings, then Presentations, Articles, Training Material, Deliverables, Technical Documents, etc..

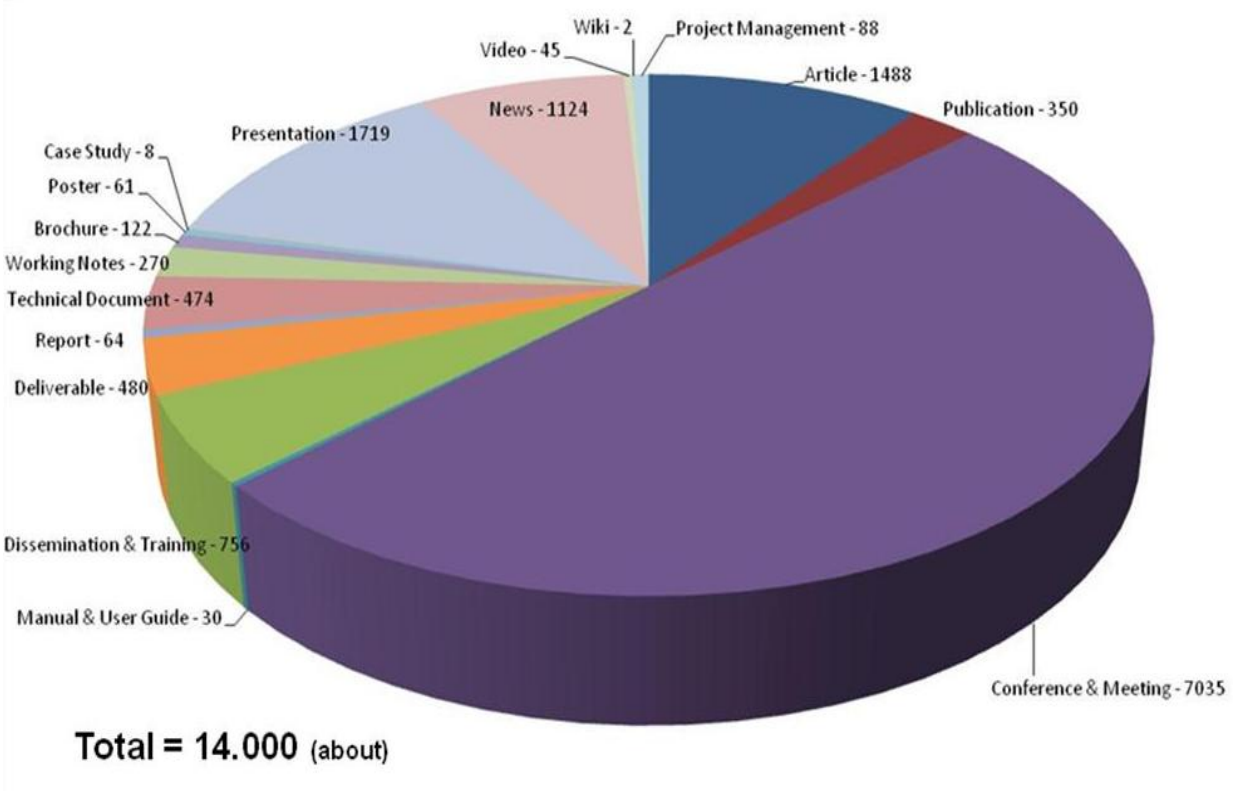


Figure 4 – DL's content type

It's meaningful to point out the trend of accesses as logged by the DL's statistical tools (see following figure). The chart clearly shows a number of peaks related – on one side – to the aforementioned events and – on the other side – to key dates corresponding to milestones in projects' lifetime (reviews, call for proposals, etc.). The decrease in the last two months is related to the particular conjunction of the year's period and of most of projects' lifecycle period.

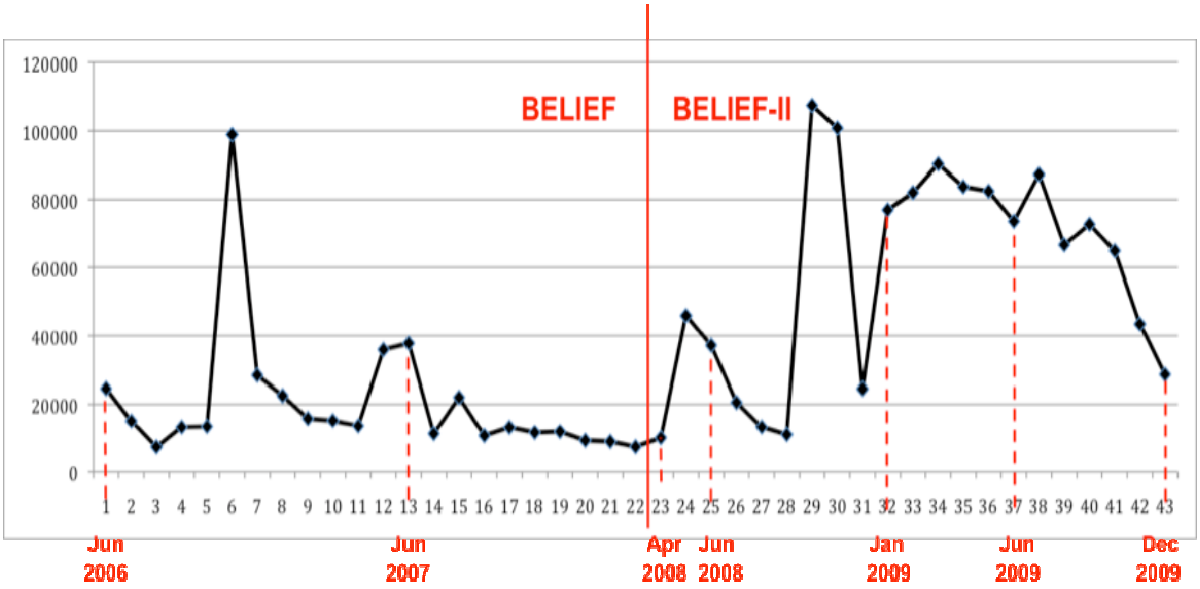


Figure 5 – DL hits per month from June 2006 to December 2009

It's worthwhile also mentioning that both the average and the maximum number of hits registered a growth of one order of magnitude from the first phase of the project (marked as "BELIEF" in the previous figure) to the second one (marked as BELIEF-II).

The detailed analysis of the users' provenance evidenced that among the **Top Sites** (excluding generic name-servers, and accesses coming from known EU projects and projects' partner sites) from which the DL is accessed the following should be mentioned:

- Google, with the GoogleBot web spider
- Yahoo, with its Web Crawler

- A number of telecoms' portals via their search services, e.g. Fastweb, Telecom Italia, Vodafone, On Telecoms, etc.
- A number of search engines, e.g. MSN Search, Cuil, Dotbot, etc.

It's worth notice that these data, joined to the info related to the provenance of accesses, highlight the wide spreading and variety of consumers achieved by the DL.

Finally, the **Top Operations** as registered by the statistical tools are the following:

- Query (simple)
- Browse
- OAI access
- Query (complex)
- Submission of new content

These data manifestly show that the most frequent operations on the DL are queries performed on both all Metadata Elements and Full Text, followed by generic browse of the content and by accesses from sites/repositories making use of the OAI/PMH protocol to query the DL. Query focused on specific Metadata Elements and the insertion of new content in the DL follow, as expected to be.

4 Technical insights

Albeit referring to [4] for a complete documentation of the technical characteristics of the OpenDLib Software System, this section briefly summarises the most relevant aspects of the BELIEF implementation, namely:

- Definition of the metadata structure implemented by the DL.
- Definition and implementation of protocols and tools for the submission of metadata and documents to the DL.
- Integration with external portals.

4.1 DL Metadata

With OpenDLib, resources can be catalogued with multiple metadata formats. The BELIEF DL uses Dublin Core Qualified (DCQ) encoding for the purpose of interoperability since DCQ enables the enhanced sharing of information between Information Sources adopting different coding with no loss of semantics [8].

A total of 17 DCQ metadata are currently supported by the BELIEF DL (The metadata in bold have to be considered mandatory for an effective classification in the DL): **Title**, **Creator**, **Subject**, **Description**, **Publisher**, Contributor, **Date**, **Type**, **Status**, **Format**, **Identifier**, **Source**, **Language**, Relation, Coverage, Rights, and **Provenance**.

The detailed description of the semantics and usage of such metadata elements is given in [5]. Implemented or suggested controlled vocabularies are introduced for Type, Status, Format, and Language, whilst the definition of controlled vocabularies has been undertaken for Creator and Subject and will be carried out along with the implementation of an "Authority File Control Service" to support librarians and administrators in such arduous task.

4.2 Submission of Metadata and Documents

Different methods for submitting documents and metadata to the DL are supported:

- On-line submission via Document Models (i.e. Web forms).
- This can be done either using the native DL facility (accessible for authorized users through the "Community" section of the DL Desktop) or via integration with external portals (e.g. BELIEF itself and D4Science, as shown in the following section).
- In the former, the document models and the underlying metadata structure are designed and implemented by the DL's Librarian. Different pre-defined models are supplied to cope with the "standard user" needs. In addition, a "Free Model" is supplied, which can be modified by skilled users according to their specific needs.
- In the latter, the models and the metadata are designed in cooperation with the administrators of the external portals, to exactly fit with their requirements.
- Harvesting from existing Information Sources, implemented by specific modules and interfaces.
- To this end, a number of different harvester has been implemented for repositories supporting any kind of programmatic access. As to the formats, both DC/DCQ and different "flavors" of MARC have been used, whilst both OAI-PMH and a number of proprietary interfaces are

supported to implement the communication protocol. In some relevant case, even HTML parsers have been implemented to access repositories not supporting any of the above.

- Batch submission via an XML schema based file.
- Finally, bulk load from repositories not supporting any on-line access facility has been implemented via batch processing of XML files. These are based on an XML schema supplied by the DL's Librarian and fine-tuned with the administrators of those repositories.

4.3 OAI-PMH Compliance

The DL features an OAI-PMH [9] fully compliant implementation which let the DL provide effective tools to support the Principles of the European Commission's Communication of Scientific Information [10] and the subsequent Competitiveness Council Conclusions on Scientific Information in the Digital Age [11].

These tools will also promote widespread adoption of an Open Access Policy that will lead towards global and seamless dissemination of publicly-funded research results (publications and data), as set out by the ERC Scientific Council Guidelines for Open Access [12] and the Open Access Pilot [13] launched by the European Commission.

To apply all of these, the integration of the DL within the DRIVER Infrastructure [14] was undertaken, implementing the full compliance with the DRIVER Guidelines 2.0 released by the DRIVER Consortium. DRIVER is the well known European data infrastructure connecting hundreds of digital repositories of institutions and research organisations. Such integration produced a first implementation which is now going to be refined with a close interaction with the DRIVER Consortium.

4.4 Integration with External Portals

An advanced and powerful characteristic of the BELIEF DL is its capability of being integrated in an external portal, that is the portal implemented by another project, willing to strictly interact with the DL to let this behave like a component/service of the portal itself.

Such integration can be achieved via either the use of the APIs provided by the underlying OpenDLib System, or by incapsulating specific code modules – especially developed for the BELIEF DL – within the external portal.

The former approach was implemented in the BELIEF Portal, which interact with the DL by means of APIs like "List Authorities" (i.e. the Entities authorized to submit content), "Submit Document", etc.. The latter was used in the D4Science project's portal, since this required the DL to run in a sort of "sand box" within the Portal itself.

The following figures show the implementation of the D4Science project, where the authentication of the user and the submission of a document to the DL are embedded in the portal. In a similar way a "plug-in-search" capability was integrated in the portal to implement a direct access to the BELIEF DL content.

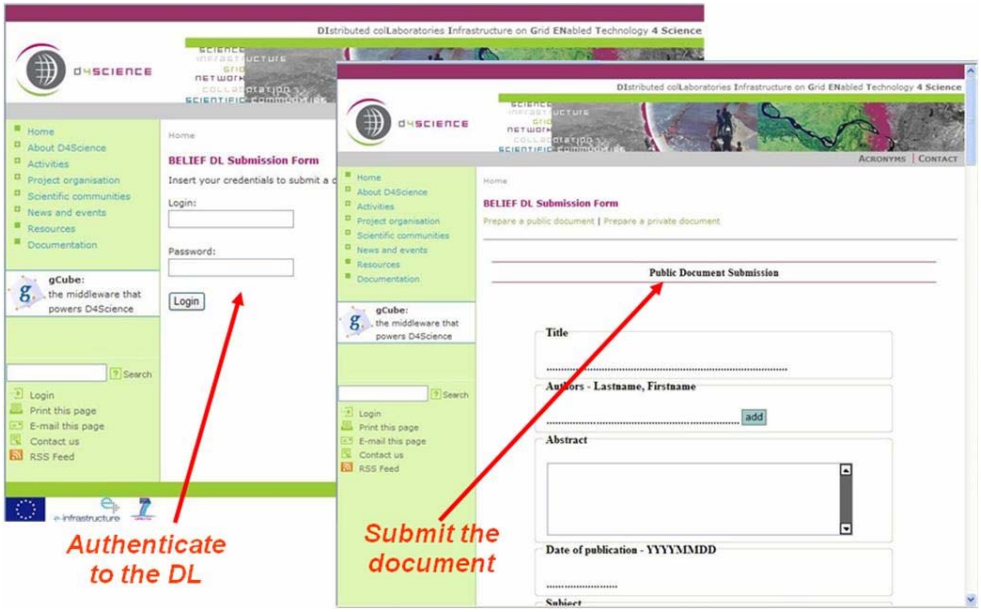


Figure 6 – DL integration with D4Science (Authentication and Submission)

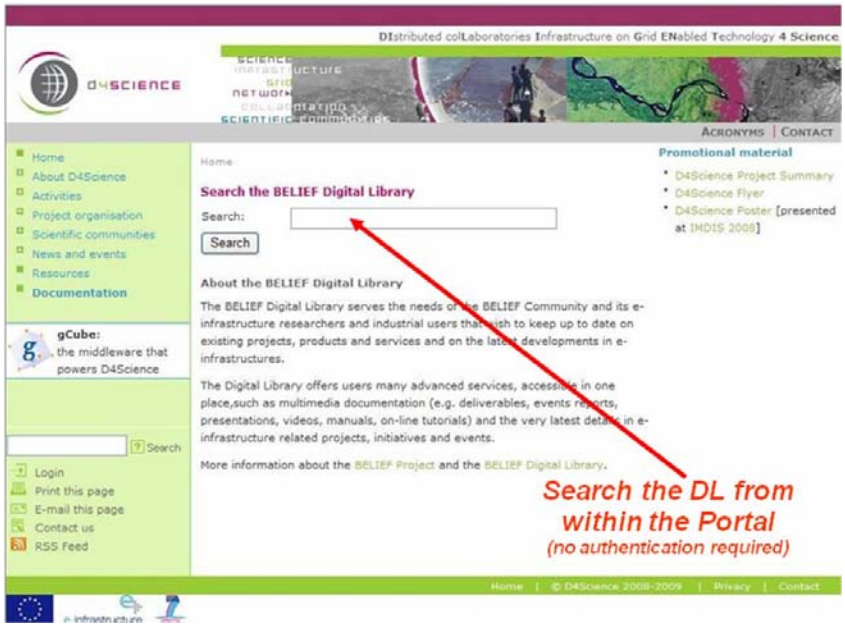


Figure 7 – DL integration with D4Science (Search)

5 Conclusions and Future Work

The effective use of a Digital Library within the scope of the BELIEF Coordination Action was shown. Specific implementation issues were highlighted, as well the conceptual aspects characterising this application context.

Focus was put on harmonisation of metadata coming from different Information Sources and on harvesting rules, protocols and formats as well as on specific harvesting tools implemented to cope with the diverse characteristics of those Sources. Compliance to open standards as well openness of the DL's architecture was introduced, with special emphasis on the DRIVER and D4Science examples respectively.

Finally, importance of a proper and controlled usage of metadata elements and values was discussed. Implemented or suggested controlled vocabularies were introduced, whilst the implementation of an "Authority File Control Service" was envisaged. Its detailed design and implementation will be one of the main goal of the project.

References

[1] BELIEF Project http://www.beliefproject.org
[2] BELIEF Digital Library http://belief-dl.research-infrastructures.eu/
[3] BELIEF Digital Library Wiki User Guide https://userguide.wiki.belief.research-infrastructures.eu/
[4] OpenDLib website http://opendlib.research-infrastructures.eu/
[5] Zoppi F., Biagini F. *BELIEF-II D3.4.1 DL - Release Notes*. Bringing Europe's eLectronic Infrastructures to Expanding Frontiers - Phase II (BELIEF-II) - FP7-Infrastructures-2007-2 - 223759. Deliverable BELIEF-II D3.4.1, 2008
[6] Zoppi F., Baldacci M. B., Biagini F. *BELIEF-II D3.3.1 DL - User Guide*. Bringing Europe's eLectronic Infrastructures to Expanding Frontiers - Phase II (BELIEF-II) - FP7-Infrastructures-2007-2 - 223759. Deliverable BELIEF-II D3.3.1, 2008.
[7] Zoppi F., Artini M., Baldacci M. B., Biagini F. - *BELIEF - WP1D5.4 - Bridging Information Consumers & Providers*. BELIEF (Bringing Europe's eLectronic Infrastructures to Expanding Frontiers) - FP6-2004-Infastructures-6 - 026500. Deliverable BELIEF WP1D5.4, 2008.
[8] The Dublin Core Metadata Initiative Open Forum http://www.dublincore.org/
[9] The Open Archives Initiative Protocol for Metadata Harvesting http://www.openarchives.org/OAI/openarchivesprotocol.html
[10] Principles of the European Commission's Communication of Scientific Information http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_en.pdf
[11] Competitiveness Council Conclusions on Scientific Information in the Digital Age http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/intm/97236.pdf
[12] ERC Scientific Council Guidelines for Open Access http://erc.europa.eu/pdf/ScC_Guidelines_Open_Access_revised_Dec07_FINAL.pdf
[13] Open Access Pilot ftp://ftp.cordis.europa.eu/pub/fp7/docs/open-access-pilot_en.pdf
[14] DRIVER - Digital Repository Infrastructure Vision for European Research http://www.driver-repository.eu/ http://www.driver-community.eu/

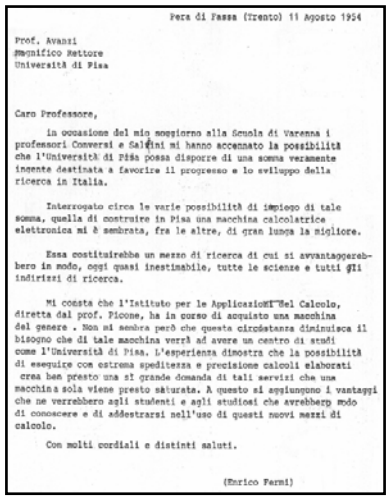
A Hidden Treasure on Computer Science Pre-History in Pisa: the CSCE Collection

Stefania Biagioni and Silvia Giannini, ISTI-CNR,
Institute for the Science and Technologies of Information, National Research Council, Italy

Abstract
The collection of the Center for the Study of Electronic Computers (CSCE) contains the scientific documentation produced by the staff and associate members of the Center. The CSCE was established by the University of Pisa with the specific purpose of designing and building the first electronic computer entirely built in Italy: the CEP, standing for "Pisan Electronic Computer". In recent times we have carried out the analysis, the description and the digitization of these historical documents, stored in paper copy at our library, and turned them into an Open Access digital archive freely available at http://csce.isti.cnr.it/.

The most important historical events

The history of the CSCE started at the beginning of the '50s, when the municipal and provincial administrations of Pisa, Lucca and Livorno granted the University of Pisa a significant amount of money (150 million liras) to be used for the realization of an important scientific device. The original idea was to build an electrosynchrotron. The studies about the construction of this electrosynchrotron began in February 1953 at the Institute of Physics of the University of Pisa, and were about to finish when a bid by the University of Rome, one of the project partners, to have the project transferred to Frascati, near Rome, was accepted; this transfer meant that the amount originally granted to the University of Pisa became available for a new project.



Pera di Fassa (Trento) 11 Agosto 1954

Prof. Avanzi
Magnifico Rettore
Università di Pisa

Caro Professore,

in occasione del mio soggiorno alla Scuola di Varenna i professori Conversi e Salvini mi hanno accennato la possibilità che l'Università di Pisa possa disporre di una somma veramente ingente destinata a favorire il progresso e lo sviluppo della ricerca in Italia.

Interrogato circa le varie possibilità di impiego di tale somma, quella di costruire in Pisa una macchina calcolatrice elettronica mi è sembrata, fra le altre, di gran lunga la migliore.

Essa costituirebbe un mezzo di ricerca di cui si avvantaggerebbero in modo, oggi quasi inestimabile, tutte le scienze e tutti gli indirizzi di ricerca.

Mi consta che l'Istituto per le Applicazioni del Calcolo, diretta dal prof. Picone, ha in corso di acquisto una macchina del genere. Non mi sembra però che questa circostanza diminuisca il bisogno che di tale macchina verrà ad avere un centro di studi come l'Università di Pisa. L'esperienza dimostra che la possibilità di eseguire con estrema speditezza e precisione calcoli elaborati crea ben presto una sì grande domanda di tali servizi che una macchina sola viene presto saturata. A questo si aggiungono i vantaggi che ne verrebbero agli studenti e agli studiosi che avrebbero modo di conoscere e di addestrarsi nell'uso di questi nuovi mezzi di calcolo.

Con molti cordiali e distinti saluti.

(Enrico Fermi)

In 1954, Marcello Conversi, director of the Institute of Physics, along with other researchers from the same institute, approached the Nobel laureate Enrico Fermi at the Varenna Summer International School of Physics "A. Volta" to seek advice on how best to use the unused grant. Enrico Fermi suggested to the Rector of the University of Pisa that the amount should be used to realize an electronic computer, the first entirely built in Italy. At that time, in Italy there were only two electronic computers, both purchased from abroad: the CRC102A, made in the US and installed in Milan at the Institute of General Electrotechnics in 1954, and the MARK I STAR, made in the UK and installed in Rome at IAC (standing for Institute for Calculus Applications) in 1955. In the opinion of Fermi, the hypothesis "*to build an electronic computer in Pisa - seemed by far the best among all the others. It would be a research tool that would benefit all the sciences and all research in an almost inestimable way*.[1]"

The recommendation of Fermi raised some doubts because the academic community claimed that the building of electronic computers was an exclusive job of the industry, and that the money would thus best be employed by buying a computer rather than by building it.

At the end, the choice fell on the route that was longer and more difficult, albeit more captivating and rich of satisfactions: building a computer. The CSCE was established in 1955 and the project started. The enterprise was financed also by the National Institute of Nuclear Physics, by the National Committee for Nuclear Research, by the National Research Council and by the society Olivetti, which participated with its own staff.

Before starting the real construction of the CEP, standing for "Pisan Electronic Computer", the executive group of CSCE decided to build a reduced-functionality prototype ("Reduced Machine" – RM). The RM was completed in 1957. Its main features were:

- word length of 18 bits
- magnetic core memory of 1024 words (1K)
- fixed-point arithmetic
- 32 machine instructions
- 70,000 additions or 500 multiplications per second
- photoelectric tape reader as the input device and teletype as the output

Despite the simple and limited characteristics of the RM, this was used in scientific calculi of considerable complexity.

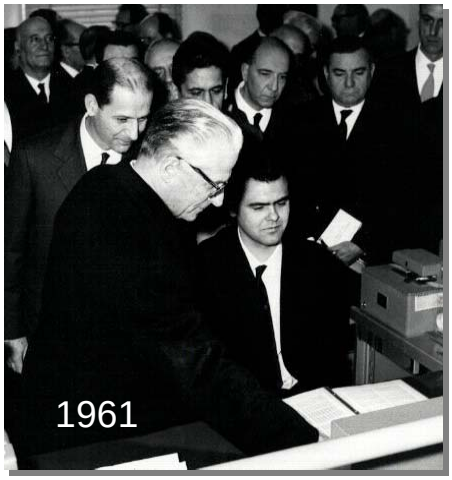
So, from a purely historical point of view, *the RM and not the CEP was the first electronic computer entirely built in Italy*.

With the RM in operation, the construction of the full-blown machine started. It was completed at the end of 1960, and it inherited from the RM about half of its component parts. However, the CEP had decidedly better performance:

- word length of 36 bits
- arithmetic in fixed point and floating point
- single and double precision
- 128 instructions and 220 pseudo-instructions
- instructions of fixed length, with length equal to a word
- 8192 words (8K) of magnetic core memory
- 70,000 additions or 7,000 multiplications per second
- data entry via photoelectric reader and exit drills with tape, teleprinter and parallel printer.

The CEP was inaugurated on November 13, 1961 by the President of the Italian Republic, Giovanni Gronchi. Six years had passed from the birth of the CSCE; in these years, the number of computers installed in Italy had grown from 2 to 22. The CEP remained in operation for seven years, offering its computing services for a minimum of 2000 hours per year up to over 4000.

The main users were researchers from the National Institute of Nuclear Physics and the Institute of Physical Chemistry at the University of Pisa. It is quite likely that, had the CEP been built before, it would have had a greater scientific value. However, the studies and the experiences connected to the building of the CEP had produced a more important and durable result: the creation of a group of experts in the emergent field of Computer Science.



In 1962, the University of Pisa and the CNR signed a convention, on the basis of which the CSCE became a "national center of interest" of the CNR. In 1964 the institute CNUCE, standing for National University Center of Electronic Calculus, was born with the aim to supply scientific calculus service. In 1968 the CSCE became the Institute for Information Processing of the National Research Council (IEI-CNR), with the aim to promote and carry out computer science research.

The close interactions and the common activities of both institutes led to their merging, in 2000, in to the present Institute for the Science and Technologies of Information "Alessandro Faedo" (ISTI-CNR).

Prof. Alessandro Faedo, Emeritus Professor of Mathematical Analysis of the University of Pisa, member of the Managing Committee of CSCE since 1954 to 1958, and Rector of the University of Pisa since 1959 to 1976, created the first Curriculum in Computer Science at the University of Pisa in 1969. Indeed, this research activity, at CSCE before and at IEI later, had contributed to establishing, in the academic community, a new scientific field such as Computer Science that, at that time, was not yet considered a discipline in its own. Currently, ISTI-CNR is devoted to basic and applied research in Computer Science.

The analysis of the collection

On the occasion of the Congress organized in order to celebrate the 50th anniversary of the CEP, held in Pisa on last June, we have carried out the analysis, the description and the digitization of the scientific documentation produced by the staff and associate members of the CSCE, which were stored in paper copy at our library. The cataloguing work started with the supervision of all the materials. The analysis of the CSCE documents has shown that the set of documents grew spontaneously and sedimented over the years, with the only aim to record the progress of CSCE studies and activities. Only later these characteristics, typical of

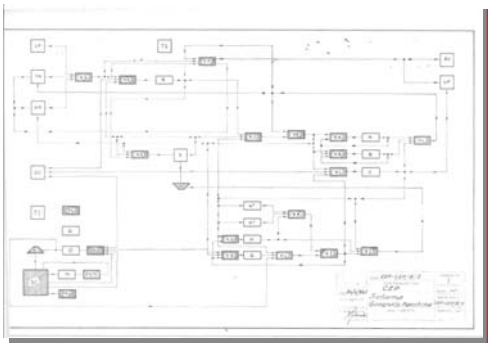
archive materials, appear to have been lost because of the increasing prevalence of a different purpose: giving to the documents the nature of a collection. Indeed, the identification codes associated to the print copies of the internal memos show that these codes were attributed *a posteriori*, and not in the year in which the notes were actually written. The two key features (naturalness and originality) that characterize the "archival bound" were thus lost. Because of the impossibility to trace or to follow the original and natural order of the papers, and because of the numerous rearrangements that the material has undergone over time, we have reorganized all the documents using a bibliography edited by CSCE and printed in 1969, called "Documentation 2".
This choice was also supported by the almost exact correspondence between the numbering in Documentation 2 and the identification codes associated to the print copies.

Right from the first recognition that we carried out, we realized that the material had been previously divided into two series, first and second, and that a further division had been made between published material, internal memos, and miscellaneous documents.
Series I consists of 42 works published between 1957 and 1965, and 147 internal memos written between 1955 and 1963 including individual descriptions of the sub-programs and the machine pseudo-instructions (61 papers).
Series II consists of 128 published works for the period 1962-1968 and 72 internal memos for the period 1963-1968.
The analytical description of the contents of the two series is presented in the bibliographic sections of the http://csce.isti.cnr.it/ site.

The archiving procedures

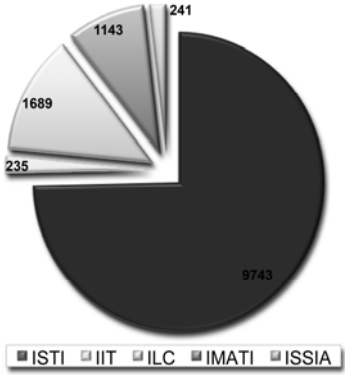
We have produced the bibliographic descriptions of all the documents by using the PUMA publication management system. We have assigned each record to the specific type provided by the system and according to its contents. The identification code for each document is assigned automatically by the system, and is composed of the name of the collection, the code corresponding to the document type, the year, and the progressive sequence number within the year (e.g., /cnr.csce/1958-A2-001). In the field "Document Number / Original Code" (e.g., CSCE-P-2-1958 (Series I)) we have recorded the original identification codes and the indication as whether the work belongs to Series I or Series II in the new record.

The collection consists in 389 documents and is composed of publications (national and international journal and conference papers), internal memos, and technical drawings related to the CEP (logical schemes, schemes of electronic circuits, mechanical diagrams)[2]. The other types of documents included in the database are all classified, by the Center, as internal memos, but this category presents other subtypes of documents such as Theses and / or Project reports. We have chosen to assign all these subtypes to the type "Internal Memos" also to Theses and to other types of documents, in order to respect the original intention of the CSCE librarians, and we have provided specific information relating to the specific content in the "Description" field of PUMA.



When the work is anonymous or the intellectual responsibility is uncertain, given the mandatory nature of the field "Authors" provided by the system, we have assigned the collective name CSCE. Each record contains an abstract, if it is available in the work or if the cataloguers have been able to extrapolate useful information to fill this field.
We have used the "Notes" field to include any news about the documents considered important for the completeness of description but not associable with any of the specific fields provided by PUMA.

All the documents, irrespectively of type, can be freely accessed at http://csce.isti.cnr.it/ because any contractual obligations with publishers are now extinct according to the Italian norms on copyright.
We have conducted the semantic indexing by analyzing all sources normally provided for such activities, so as to identify the basic contents of the documents. We have represented the contents of a document by assigning to it one or more free textual descriptors and one or more textual descriptors derived from the classification system in use at our library, the ACM Computing Classification System (1998), together with their respective class codes. Search and retrieval of documents by content can be obtained by entering the class code and / or the word in the "Subject" field of the system.
To each bibliographic description we have associated a file obtained from the digitization of the paper copy kept in archive. The main difficulties encountered in this process have been due, above all, to the physical conditions of the historical documents, in particular of the internal memos: faded ink, torn paper, very large format of the technical schemes. The digitization work, thus, has been long and complex, and has sometimes required several scanning passes. Sometimes it has also been necessary to rework the file with specific software that allows editing of the resulting files obtained by the acquisition, until the result was considered satisfactory. We have tried to enhance and simplify, above all, the readability of internal memos, but this choice has often produced the loss of the original format.

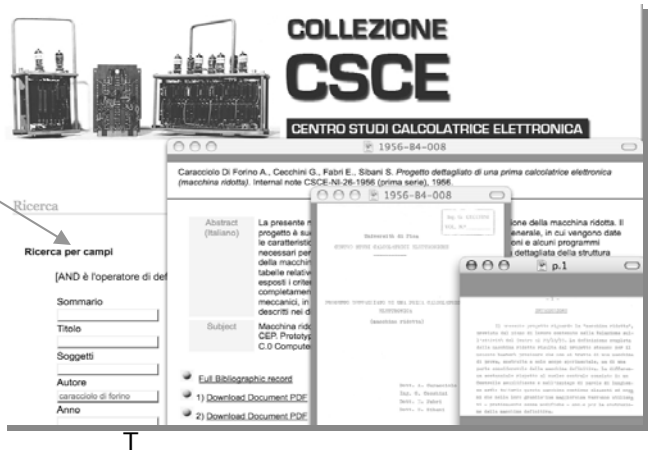


The computerized management systems
The Open Access archive CSCE has been created and is managed by the systems Puma and MetaPub (both developed by the ISTI Library). PUMA is a software infrastructure, user-focused and service-oriented. It presently manages 32 CNR institutional repositories that contain about 22.000 documents covering different disciplines. Repositories and collections are growing daily. The system functionalities are oriented to meet the requirements of CNR researchers by facilitating their self-archiving, ensuring the preservation of their documents, providing world wide easy web access to their papers and by permitting them to manipulate the stored content to fulfil scientific and administrative issues. PUMA offers a common user interface, both in English and Italian, to search, browse and self-archiving.
Special functions are provided by the additional software module MetaPub that uses repository content to deliver different bibliographyc and amministrative services. MetaPub also opens PUMA repositories to the OAI world by implementing the OAI-PMH protocol (the full text harvesting is allowed under authorization).
The CSCE collection is part of a larger project, the cataloguing and digitization of the heritage of the CNR institutes in the Information and Communication Technology (ICT) area.

The CSCE Collection Web site
The CSCE collection is accessible from its own web site: http://csce.isti.cnr.it, from PUMA interfaces: http://puma.isti.cnr.it, from MetaPub interfaces: http://leonardo.isti.cnr.it/metapub

The Web site provides general information about the collection.

.

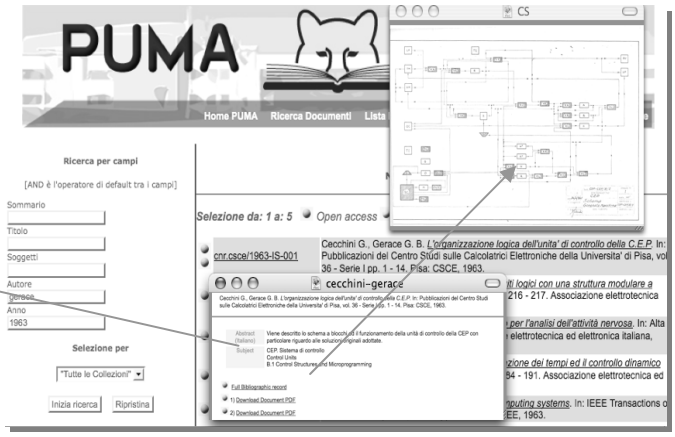


The search and browsing tools available from the CSCE site are endowed with advanced search functionality.

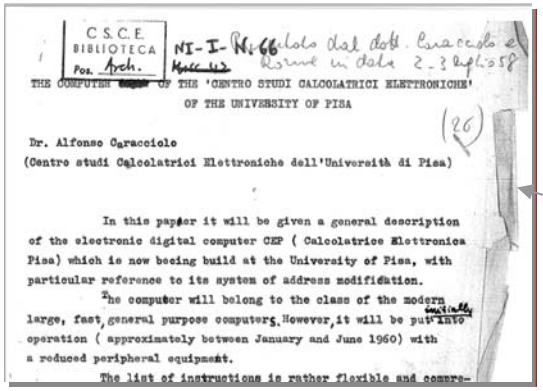
T

The PUMA Interface

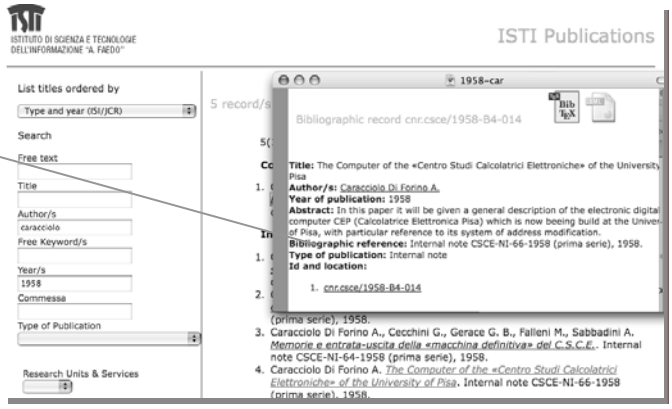
.



The selected document has two associated files: a document and a scheme.



The MetaPUB Interface



The MetaPUB system allows downloading from the PUMA repository.

Conclusions

We created this open access digital collection in order to make freely available what we like to think of as "a little hidden treasure" of historical value. A further purpose of ours was to "celebrate" the CSCE and its original and beautiful adventure as a sort of milestone in the history of our institute, since the CSCE represents the origins of our institute and the origins of Computer Science in Pisa and in Italy. Finally we wanted to pay homage to an enterprise that, in our opinion, has a great scientific and human value.

The CEP project was a symbolic effort of reconstruction and recovery by those who invested in that work, during a period of Italian history in which the wounds of war were not yet completely healed. It was a symbol of organizational capacity, a demonstration of skill by the graduates and technicians that, all Italians, planned and built a calculator that was the object of admiration throughout Europe[3]

References

- Franco Denoth. *CEP (Calcolatrice Elettronica Pisana). Dal CSCE all'IEI*. In: Convegno Internazionale sulla Storia e Preistoria del Calcolo Automatico e dell'Informatica, Siena, 10-11-12 settembre 1991. Milano: AICA, 1991, p. 105-115.
- Giuseppe De Marco. *La calcolatrice elettronica pisana: le origini dell'informatica in Italia*. Tesi di laurea in Scienze dell'Informazione. Pisa, Istituto di Elaborazione dell'Informazione del C.N.R., giugno 1996.
- De Marco G., Mainetto G., Pisani S., Savino P. *The early computers of Italy*. In: IEEE Annals of the History of Computing', 21, n. 4 (1999), 28-36. IEEE, 1999.
- Biagioni S., Carlesi C., Romano G. A., Giannini S., Maggi R. *PUMA & MetaPub: open access to Italian CNR repositories in the perspective of the European Digital Repository Infrastructure*. In: 9th International Conference on Grey Literature. GL9 (Antwerp, BE, 10-11 December 2007). Abstract. Dominic Farace (ed.). (GL Proceedings). TextRelease, 2007.

Acknowledgments
All aspects of the history of CEP and CSCE are drawn from (De Marco, 1996).

References

[1] translated from *G. De Marco. La calcolatrice elettronica pisana: le origini dell'informatica in Italia. Pisa, IEI-CNR, giugno 1996, All. 1: Letter typewritten of Enrico Fermi*

[2] We have also found some works presented in the bibliography Documentation 2 as "CSCE Publication" without no reference to a real and occurred publication. In these case we have assigned the documents to the type "IS-Institutional Series".

[3] translated from *G. De Marco. La calcolatrice elettronica pisana: le origini dell'informatica in Italia. Pisa, IEI-CNR, giugno 1996, p.94*

Dissemination of Information on the Results of R&D through the JAEA Library

Kiyoshi Ikeda, Mayuki Gonda, Takashi Nozawa, Keizo Itabashi, JAEA, Japan Atomic Energy Agency
Naomi Ebisawa, TOSS, Total Support Systems Corporation, Japan

Abstract
This paper describes the new systems of circulating grey literatures in the fields of nuclear science and technology in Japan. The JAEA Library has been making the efforts to ensure that JAEA Reports (i.e. JAEA-Research etc.) are accessible via the Internet. The JAEA Library website incorporates two types of systems for use disseminating R&D results including JAEA Reports. JAEA Abstracts is a list of the results of current R&D achieved by our staff. JOPSS is the abbreviation of JAEA Originated Papers Searching System. The full-texts (PDF format) of JAEA Reports are also available through the above mentioned methods. JAEA Abstracts have been available in both Japanese and English, whereas JOPSS was only available in Japanese. Recently, a new search interface and retrieval algorithms have been developed for JOPSS. JOPSS has also been made available in English with the aim of making our R&D results more convenient to users throughout the world.

1. Introduction

Use of the Internet is an effective way of disseminating grey literature. The Japan Atomic Energy Agency (JAEA) makes R&D results in the fields of nuclear science and technology, including JAEA Reports (i.e. JAEA-Research etc) available via the Internet, with the JAEA Library being responsible for the task.
The JAEA Library website provides two types of systems: JAEA Originated Papers Searching System (JOPSS) and JAEA Abstracts, for current information on research results brought by JAEA staff.
This case study focuses on the activity of JAEA Library with JOPSS as the main purpose of it was to understand exactly how the JAEA Library contributes to the dissemination and organization of grey literature. In view of the future plan with JOPSS, the data contained in JOPSS was then analyzed.

2. JAEA

JAEA is a comprehensive research and development organization that aims at the practical use of nuclear energy and is involved in basic and applied research through to the establishment of nuclear fuel cycles. Its largest site is in Tokai-mura where the central library is located.

2.1. JAEA R&D Results

Every year JAEA publishes more than 300 JAEA Reports, which can might be the grey literatures, and also submits more than 1,700 papers to academic journals. The JAEA Library also makes the efforts to provide in information on the result of R&D, including the full-texts of JAEA Reports, available via the Internet. The following table describes the results of JAEA's R&D accumulated by the JAEA Library.

Table1. Statistics of accumulated R&D results

As of Dec.2009

	Reports	Papers	Remarks
Japan Atomic Energy Agency (JAEA)	1,040	5,639	Oct.2005~
Japan Atomic Energy Research Institute (JAERI)	7,810	23,857	Jun.1956 ~Sep.2005
Japan Nuclear Cycle Development Institute (JNC) Power Reactor and Nuclear Fuel Development Corporation (PNC)	11,723	4,581	(JNC) Oct.1998 ~ Sep.2005 (PNC) Oct.1967 ~Sep.1998
Total	20,573	34,077	

The Japan Atomic Energy Research Institute (JAERI) and Japan Nuclear Cycle Development Institute (JNC, but originally known as PNC) merged as JAEA in October 2005. The history of JAEA is still rather short but JAERI, JNC and PNC have both released the results of R&D for a total of about 50 years. Part of the R&D results includes, about 20,000 JAEA Reports which are grey literatures, with their full-texts being available on the JAEA Library website.

Fig.1. Breakdown by research fields

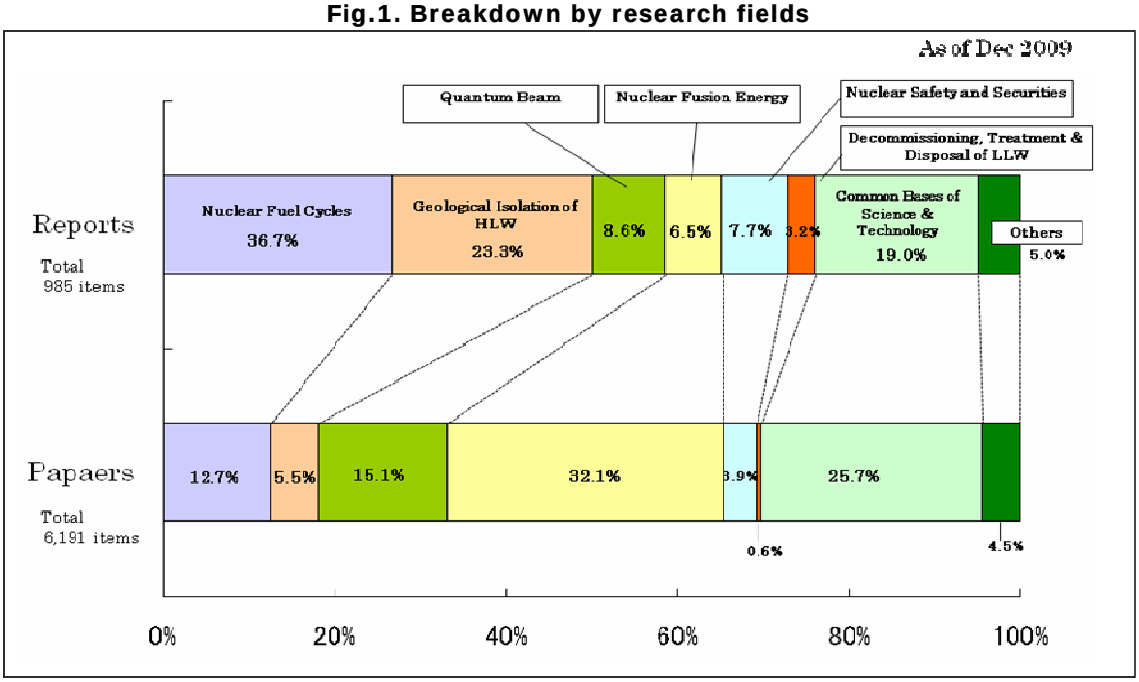


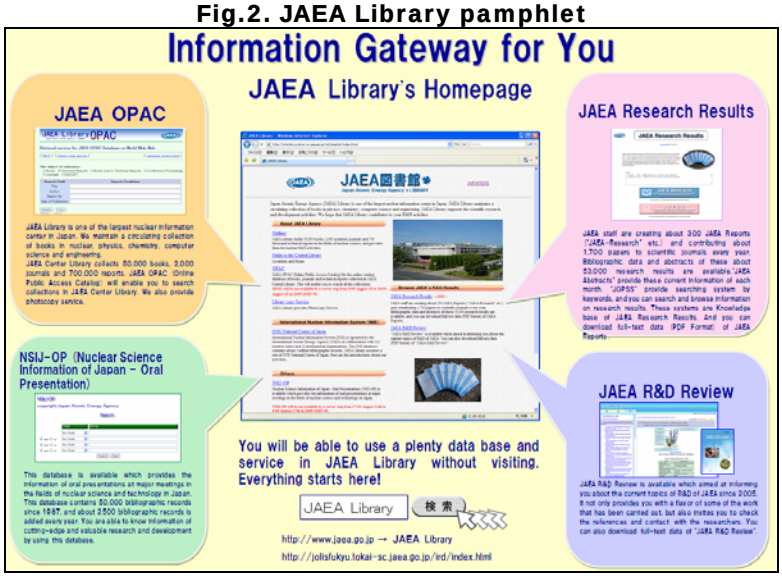
Figure 1 provides a breakdown by research fields of JAEA Reports and papers submitted by the JAEA staff. For example, 36.7% of the reports and 12.7% of the papers were published on nuclear fuel cycles, which include fast reactors. The methods of presentation depend on the research field.

2.2. Features of the JAEA Library

The JAEA Library is one of the largest nuclear information centers in Japan. One of the missions of the JAEA Library is disseminate the results of R&D achieved by the staff of JAEA, along with providing a library of nuclear related information (Figure 2).

More concrete the JAEA Library is involved in: (1) Collecting and providing materials as a library and nuclear information center (2.3. JAEA Library), (2) Publishing technical reports that originated in the results of JAEA's R&D and making them available both throughout Japan and abroad (2.4. Publication of JAEA Reports), (3) Collecting and providing information in the fields of nuclear science from all over the world (2.5. International Cooperation).

Fig.2. JAEA Library pamphlet



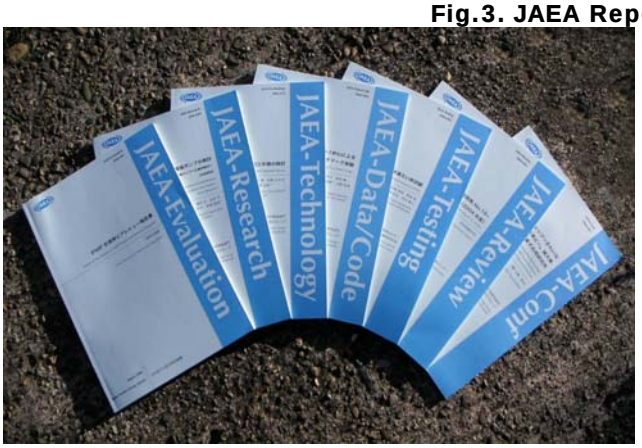
2.3. The JAEA Library

The JAEA Library includes 150,000 books, 2,000 academic journals and 1.1 million technical reports in the fields of nuclear science as the largest nuclear information center in Japan. The primary subject areas cover nuclear technology, chemistry, physics, material science, computer science, engineering and more.
The library is committed to innovation and has continued to increase desktop access to the library service through electronic journals, online databases, web resource guides, and electronic request forms. The library staff also provides specialized information services and maintains collections at 13 JAEA sites.

2.4. Publication of JAEA Reports

JAEA Reports can be divided into seven types of report codes (report prefixes) according to context (Figure 3). The advantages of JAEA Reports include that (1) R&D results can be described in detail because there is no limit on the number of pages, and (2) the results of R&D can be published with their copyrights retained. Conversely however, in the case of papers submitted to journals the number of pages is limited, and the authors also have to transfer the copyright to the publishers.

Fig.3. JAEA Reports



JAEA-Research
-Research Reports
JAEA-Technology
-Technical Reports
JAEA-Data/Code
-Numerical Data, Computer Codes and Database
JAEA-Testing
-Manuals and Experimental Results
JAEA-Evaluation
-Institutional Evaluations, Research Evaluations and Project Evaluations
JAEA-Review
-Research Review, Annual Reports and Theses, etc.
JAEA-Conf
-Conference Proceedings

2.5. International Cooperation

The International Nuclear Information System (INIS) is an informational system that involves the field of peaceful uses of nuclear sciences and technology. INIS is operated by IAEA (International Atomic Energy Agency) in collaboration with 122 member states and 24 international organizations, with the INIS database containing approximately 3 billion bibliographic records. The JAEA Library assumed the role of being the INIS National Center for Japan in 1970. The JAEA Library prepares bibliographic records on nuclear related literature published in Japan, and then submits them to the INIS database.
The International Nuclear Library Network (INLN) was initiated by the AECL Library and IAEA Library in support of global nuclear knowledge management initiatives and to promote international cooperation between nuclear libraries. The JAEA Library is also a participant in INLN.

3. JOPSS and JAEA Abstracts

Circulating grey literature is usually rather difficult but the JAEA Library has been making the efforts to achieve accessibility to JAEA Reports (i.e. JAEA-Research etc.) via the Internet. JAEA Library makes the results of R&D available using the two methods. JAEA Abstracts provide current information on the results of research achieved by JAEA Staff. JOPSS provides a keyword search system that can be used to search and browse information on the results of research achieved by JAEA staff. Full-text data (PDF format) on JAEA Reports (JAEA-Research etc.) can be dowonloaded through either of methods.
Information on the results of JAEA R&D is available from the JAEA Library website (Figure 4): http://jolisfukyu.tokai-sc.jaea.go.jp/ird/english/index.html

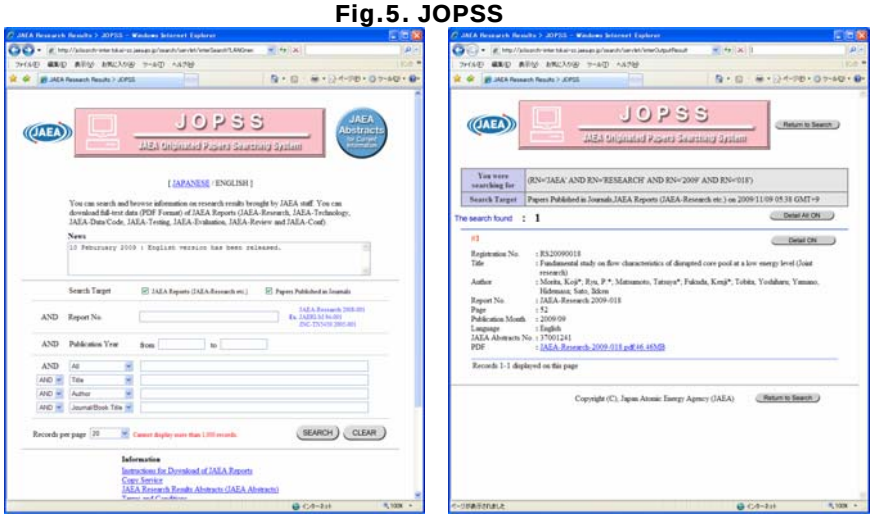
Fig.4. JAEA Library website

3.1 JOPSS

JOPSS is an abbreviation of JAEA Originated Papers Searching System. All the results of JAEA R&D accumulated over the past fifty years can be searched. JAEA Reports has the hyperlink to the full-texts.

A new search interface and retrieval algorithms for JOPSS were developed. The main features of that development include "Pull-down menu options for searching items" and a "Google like search style". Originally JOPSS was only available in Japanese. Recently however, JOPSS has also now been made available in English with the aim of making the results of R&D more globally convenient (Figure 5).

Fig.5. JOPSS



3.2 JAEA Abstracts

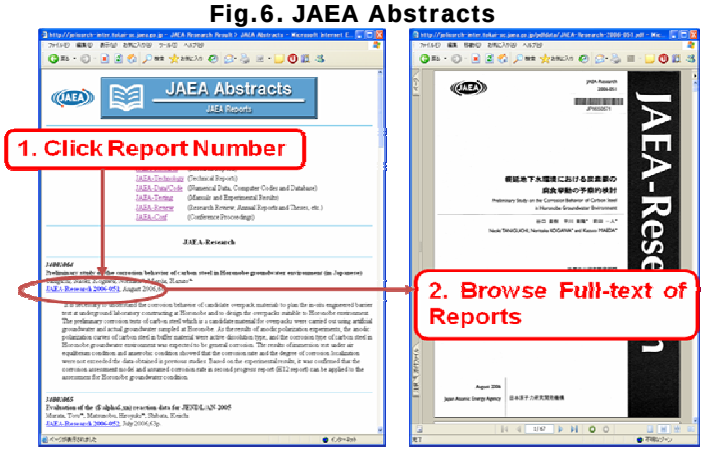
JAEA Abstracts list current R&D results, and is updated monthly. Bibliographic data on JAEA Reports and papers submitted to journals by JAEA staff, including abstracts, is made available through use it. JAEA Reports data includes hyperlinks to their full-texts (Figure 6).

JAEA Abstracts include the following features:

- Current list of the research of achieved by JAEA staff
- Hyperlinks to the full-texts of JAEA Reports
- Monthly update and collections
- Increased reference and citation

JAEA Abstracts are available in both Japanese and English. The full-text (PDF format) of JAEA Reports can also be obtained using.

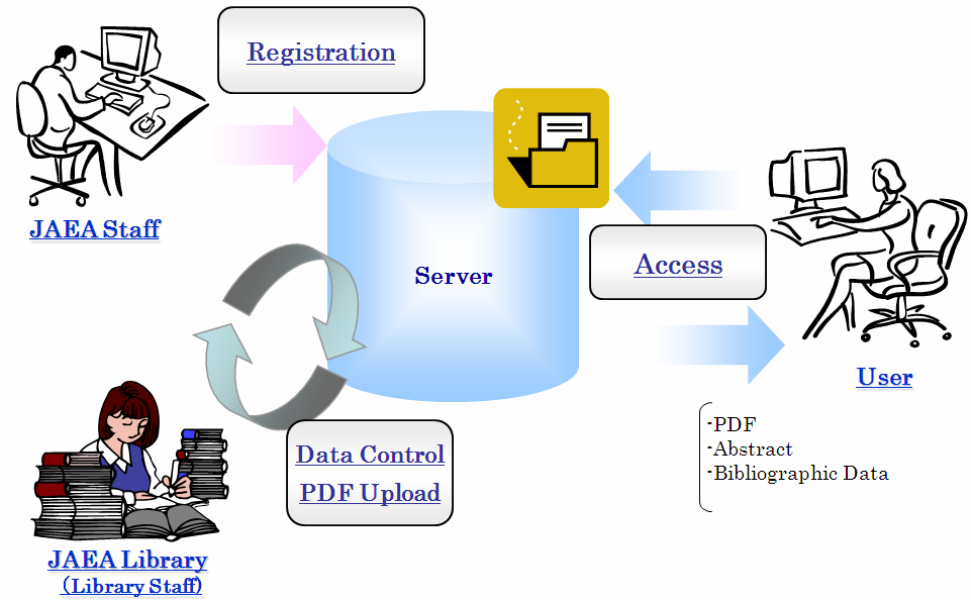
Fig.6. JAEA Abstracts



3.3 Web based applications

The method used to process the information will be described here. When JAEA staff wish to submit a paper, they have to first register/input its bibliographic data to our application system after, being approved by their directors. Our Library Staff control any bibliographic data input by JAEA Staff (Figure 7).

Fig.7. Web based application systems



For example the system incorporates authorities such as the author's name and journal title. After a paper is published the librarians confirm the bibliographic data using reprints, and make any necessary addition or corrections.

This JAEA rule is observed in ensuring accurate information is consistently provided. We consider the system to be the first step in providing an institutional repository, but there are still issues to be faced in completing the system.

4. Results and Discussion

4.1. Japanese language problem

The results of research are generally presented in two steps. The first step can involve oral presentations being made at conferences, meetings or seminars held by the society concerned. The second step involves papers being submitted to academic journals technical reports, or books. In Japan, at present still, these activities are done in Japanese in many cases.

Fig.8. JAEA Reports by language

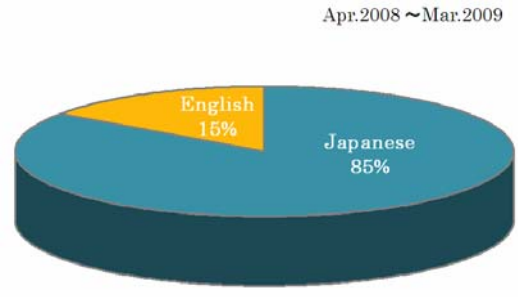


Fig.9. Journal articles by language

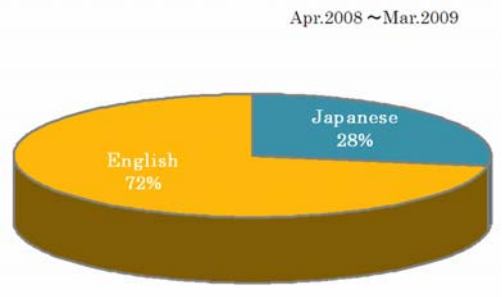
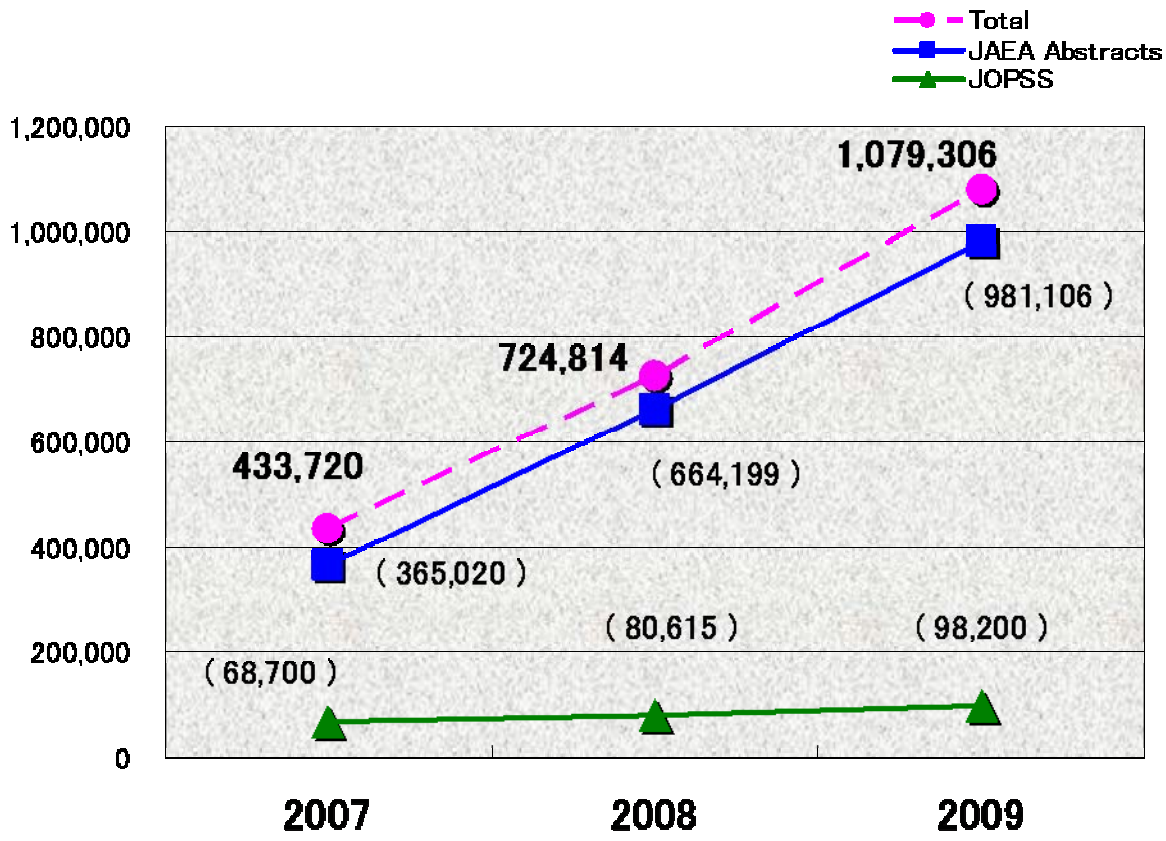


Figure 8 provide breakdown of the languages used in the JAEA Reports. 85% of JAEA Reports are published in Japanese, some of which can be considered to be grey literature. To improve the accessibility of these reports, English title and abstracts are requested to be added to all JAEA Reports. Furthermore, 72% of articles submitted to academic journals are in English (Figure 9). This language barrier will changed with grey literature in the future.

4.2. Increasing access from just JAEA

Figure 10 reveal the change in total accesses of JAEA Abstracts and JOPSS from others than just JAEA. It is obvious that the numbers of times they are accessed has rapidly increased.

Fig.10. Access statistics from other than JAEA



We only publish 100 copies of report and only provide printed reports to organizations or libraries closely related to nuclear science. However, a large number of people actually read the full-texts. Table 2 provides the numbers of times to the full-texts of JAEA Reports are accessed. As can be seen a certain specific report was actually downloaded more than 40,000 times.

Table 2. Top 10 of download count at JAEA Reports

Total download count: 1,685,803 items
Oct.2005～Dec.2009

No.	Report No.	Download Count	Title	Publication Date
1	JAEA-Research 2006-042	40,401	Feasibility study on commercialized fast reactor cycle systems technical study report of phase Ⅱ 1; Fast reactor plant systems	2006/06
2	JAEA-Research 2006-043	31,853	Feasibility study on commercialized fast reactor cycle systems technical study report of phase Ⅱ 2; Nuclear fuel cycle systems	2006/06
3	JAEA-Review 2008-009	24,051	Fast reactor cycle technology development; "FaCT Seminar, R&D for Key Technology of National Importance" conference report and collection of documents	2008/03
4	JAEA-Review 2008-037	23,235	Handbook on process and chemistry of nuclear fuel reprocessing, 2	2008/10
5	JAEA-Research 2006-044	21,410	Feasibility study on commercialization of fast reactor cycle systems technical study report of phase Ⅱ 3; Synthetic evaluation for FR cycle	2006/06
6	JAEA-Research 2007-044	20,443	Horonobe Underground Research Laboratory project synthesis of phase I investigations 2001-2005 volume "Geoscientific Research"	2007/03
7	JAEA-Technology 2008-009	20,422	Safety analysis for severe accidents anticipated during truck transport of uranium fresh fuel and raw material (Contract research)	2008/03
8	JAEA-Evaluation 2008-001	19,025	Assessment report of research and development activities; Activity "Advanced Science Research" (Interim report)	2008/08
9	JAEA-Evaluation 2006-002	18,744	Feasibility study on commercialized fast reator cycle systems; Phase Ⅱ final report	2006/07
10	JAEA-Research 2007-047	18,731	Research on the state-of-the-art of accident consequence analysis method for non-reactor nuclear facilities, 1	2007/06

4.3. Future plans

We have three future plans. Firstly, for good circulation of R&D results, we will promote writing articles in English to JAEA staff. Writing articles in English is very useful tool in research community for communicating or exchanging information.
Secondly, we will have links to e-journal sites from JOPSS and JAEA Abstracts to enhance access to full-text. Because JAEA Reports is available the full-text but full-texts of the journal articles are not available in JOPSS and JAEA Abstracts.
Thirdly, we will bring our bibliographic data into compliance with OAI-PMH (International standard) for the creation of a complete institutional repository.

5. Conclusion

This paper presented the attempt to disseminate information on R&D Results based on JOPSS and JAEA Abstracts through the activities of the JAEA Library. It is said that grey literatures is only read by limited number of people or that it is does not have very good information circulation. JAEA Library is making an effort to achieve greater through use of the Internet technology, while retaining the advantages of JAEA Reports.
The development of JOPSS will hopefully aid our efforts in contributing to the dissemination of the results of R&D and result in the better circulation of grey literatures in the fields of nuclear science and technology.

References

- Gonda, M., Ikeda, K., and Ebisawa, N. (2008). "JAEA Abstracts" and "JOPSS"; Pioneer of institutional repository (in Japanese). Senmon Toshokan, (228), p.26-32.
- Gonda M., Ikeda K., Ebisawa, N., and Nozawa, T.(2008). Management System; aiming to user-friendly system for foreign researchers (in Japanese). Preprints of the 5th Information Professional Symposium, p.125-129.
- Ikeda, K., and Nozawa, T.(2007). Dissemination of JAEA R&D Reports via the Internet, Conference Papers & Powerpoints on CD-ROM of 9th International Conference on Grey Literature, Antwerp, Belgium, 10-11 December 2007, TextRelease, 2008.
- Ohshima, T., Nagaya, S., and Yonezawa, M. (2009) .Follow-up study of oral presentations at 2005 annual meeting of atomic energy society of Japan (in Japanese). Proceedings of 2009 Fall Meetings of the Atomic Energy Society of Japan, 23-25 March 2009.Tokyo. p.399.

Grey Literature and Computational Linguistics: From Paper To Net

Claudia Marzi, Gabriella Pardelli, and Manuela Sassi
Istituto di Linguistica Computazionale; Consiglio Nazionale delle Ricerche, CNR, Italy

Abstract
The advent and exponential development of the World Wide Web has led to an increasing availability of unstructured knowledge and distributed information sources, meeting general public requirements that are hardly addressed by other more traditional information channels. This trend has concurrently raised a considerable interest in the application of Computational Linguistics (CL) methodologies to document access and retrieval, as they offer the unprecedented opportunity to make the subjective, user-centred information demands of Net citizens meet the ever changing and heterogeneous information flow of the web.
Over the last five years, more and more Italian Universities have introduced CL courses into their Humanities curricula, making available on-line teaching materials, tutorials and language engineering software that appear to supply the lack of offer from traditional Italian publishing houses. In this paper, we consider in some detail the role played by this type of Grey Literature in bringing up a wider and increasingly more aware community of web users in Italy.

Keywords: Computational Linguistics, Grey Literature, Web-based information

1. Introduction

1.1 Computational Linguistics and Language understanding

Computational Linguistics (CL) and Natural Language Processing (NLP) have profoundly changed the way we look at human language as a subject of scientific inquiry, shifting emphasis from abstract knowledge to real usage (Manning & Schütze 1999).
Understanding language requires the ability to master a heterogeneous system of manifold skills, based on the processing of complex information structures in context (Bybee & Hopper 2001, Jackendoff 2002). In turn, these structures may vary, depending on the speaker's communicative intentions and purposes (Barsalou 1999). In this specific sense, knowledge of language can no longer be decoupled from "doing things with words" (e.g. reading, learning, recalling, guessing, judging etc.). The application of computer technologies to issues of language understanding and production epitomizes such a profound change of perspective in the most exemplary way. CL concerns itself with the empirical testing of language models and has contributed, over the last 15 years, to shedding considerable light on the interplay between linguistic abilities and general cognitive functions such as inference, classification and learning (Mitchell 1993, Guarino 1998). Awareness of these issues can have a tremendous impact on the general-public daily demands for real-time, goal-oriented and personalized information paving the way to a new generation of web users.
To date, the vast majority of information available on the web is conveyed through a flood of largely unstructured and uninterpreted text material, ranging from large digital archives to blogs, electronic newswires and dedicated web pages. Flexible, intelligent access to such a digital haystack is a logical precondition to its very existence: information needles are non existing if the haystack is not structured, disseminated and made available to respond, in real time, to the personalized needs, queries and ever changing goals of daily web users (Lawrence & Lee 1999).

1.2 Language, ontology and web contents

Gaining intelligent access to web-based contents presupposes, among other things i) a formal representation (ontology) of the knowledge areas of interest, ii) an algorithm for classifying a new text item according to known ontological categories, iii) an automated procedure to relate terms and complex relations among terms to knowledge categories, iv) ways for automatically indexing contents on the basis of the terms and relations they contain, v) ways for querying a document repository by terms, relations and concepts (Lancaster 2003, Buitelaar, Cimiano & Magnini 2005).
Clearly, non professional web users can profit from ontology-based and language-based technologies for document access and management. Although considerable progress has been made in this area, however, the search for relevant web-based information is still either a) channelled through the strictures of largely pre-defined, context-free, fully-interpreted document ontologies, or b) limited to simple text queries, containing key words and relatively unstructured word patterns connected through few logical operators. Both modes of access have serious limitations.

Document ontologies are formal representation tools for sharing, categorizing, normalizing and indexing knowledge. They are typically highly structured, explicit and symbolic, but do not necessarily reflect the personal viewpoint of a web user, which is often driven by goal-oriented needs and may be unaware of encyclopaedic knowledge.
On the other hand, projecting text queries onto the web is often based on pattern-matching techniques, searching for surface text strings in a few limited variants. This is very efficient and effective for many purposes, but cannot go beyond surface text information. For example, pattern-matching cannot distinguish between the different uses of the English homonym "bank".
More advanced algorithms for intelligent document access are currently narrowing down the gap between string-based and concept-based queries of web-based content. These algorithms call for a tighter integration between ontological and linguistic knowledge: language interpretation certainly presupposes knowledge about the world, but much world knowledge is acquired through language.
Since language and ontological knowledge are two complementary ways of representing information, future generation search engines will face the challenge of dealing with flexible and open information, which is structured but goal-directed, explicit but personalized. In this scenario, web users will rely more and more often on language queries to look for information, but they will be supported by more structured and shared ways for content indexing and access.

There is another deep sense in which current progress in language technologies can help to narrow down the persisting gap between the demand for personalized information needles and the unlimited offer of unstructured haystacks on the web. Awareness of issues of automated text interpretation and document understanding may represent an altogether different way of narrowing this gap. Web users that understand NLP algorithms are certainly in a good position to use this understanding for better queries. For example, it is often possible to avoid unwanted ambiguities in a query by simply providing it with more context. Adding the word 'money' to a query containing 'bank' is useful to convey the intended sense of financial institution. In other cases, the use of synonyms in the query can broaden up the range of retrieved documents to a significant extent. We contend that grey literature can play an important role in increasing the general public awareness of these issues. In the remainder of this paper we investigate the potential impact of web-based dissemination of Italian CL literature onto a new generation of non professional web users.

2. Teaching Computational Linguistics in Italy

2.1 CL and the Web

The increasing popularity of Computational Linguistics, over the last ten years, as a knowledge area of general interest has gone hand in hand with the exponentially growing need to access web-based information by daily web users. This is also witnessed by the number of courses on CL that are nowadays active in the curricula of most Italian universities, and also meets a non specialist demand for a more aware use of language as a vehicle for querying and accessing web-based information.
Such an upsurge of interest, however, has not been accompanied by a comparable growth in the number of Italian books devoted to the topic, as traditional publishing houses are typically conservative and rather reluctant to explore new scientific domains and nonetheless so far capable of addressing the general public need for a more aware use of current technologies only to a very limited extent.
This fact, together with the persisting difficulty to read and understand the English technical literature by Italian undergraduate students, has created an acute demand for accessible Italian grey literature on the topic. To date, the amount of unpublished recommended reading material for taking CL university exams largely exceeds the white literature available on NLP issues.

2.2 Grey CL literature

To make it up for comparative shortage of white CL literature in Italian, CL courses have sprouted dedicated web sites providing tutorials, exercises, power-point presentations, language engineering software and other teaching materials. The increasing availability of on-line materials on language technologies provides also a potential access to a wider community of users, by offering introductory information for a better understanding of i) aspects of computer architecture and functioning, ii) issues of digital text encoding and document representation, iii) aspects of text browsing with personalized search patterns, iv) issues of document mark-up and classification, v) fundamentals of document content indexing, vi) introductory notions on database management systems.

The trend is setting up high standards for web-based publishing and provide as well an important meeting point between academic information providers and non-academic information consumers, by modifying the general public attitude towards computer-based information access and by prompting more personalized ways of accessing web-based information; as a more intelligent and personalized use of the most popular search-engines ties in with a deeper understanding of how texts are automatically searched and indexed and how text information is internally represented and eventually made available by computers.
A new generation of self-taught web users is getting more and more aware of the potentials and limits of the current language-based technology, in these novel scenarios for information access based on integration of language-specific abilities and language-specific cognitive functions.
When people understand better and better what sort of queries make sense in what contexts and why, they will most likely use this knowledge to avoid the strictures of precompiled information repositories and ontologies and to exploit the web in more personalized and creative ways.

3. Case study

3.1 "Informatica Umanistica" university course in Pavia

The university course "Informatica Umanistica" at the Faculty of Humanities in Pavia for the academic year 2008-2009 has been presented on the Faculty website together with the whole teaching materials. A course overview describes the goals of the course and suggests some prerequisites. The online tutorials offer i) a full set of course slides (in power point extension), ii) full set of teaching material offered during the course, iii) on-line exercises, iv) downloadable documents, v) links to websites of interest and downloadable software, vi) access to on-line tools for Italian text processing.
On-line availability of teaching materials provide not only the possibility of remote on-line access to actual course materials, but also appear to supply the lack of offer from traditional Italian publishing houses. The reference in the course program to only 5 chapters of *"Testo e Computer" Lenci, Montemagni, Pirrelli – Carocci Editor (2005)* is a clear evidence of the difficulty of Italian publishing houses of addressing the general public need for a more aware use of current technologies, and of their traditional and conservative attitude.

3.2 "Electronic Archive of Carlo Emilio Gadda's Works"

As a second case study we counted the website accesses to the whole literary production of Gadda (1893-1973) [http://www.ilc.cnr.it/CEG/]. See figure 1.
The creation of the "Electronic Archive of Carlo Emilio Gadda's Works" was started in 1994, when publisher Garzanti made their published text materials (Gadda, 1988-93), already prepared for photo-composition and for study purposes only, available to the ILC.
The first version of the Corpus, running in DBT 3.0, was presented on November 14, 1997 (the 104th anniversary of Gadda's birthday) at the CNR in Rome.
While the community of Gadda scholars is starting to demand something that cannot be distributed but is available for consultation at ILC - concordances -, the project started to: (i) produce lexicographic tool standards by means of an automatic processing system; (ii) apply innovative methodologies onto sample subsystems.
The web site advertised the history of the project and made paper publications available online. Linguistic resources available:

{1)} Hapax Legomena Inverse Index
{2)} L'accentazione in Gadda [Accentuation in Gadda]
{3)} Un primo censimento di termini gaddiani [A first census of Gaddian terms]
{4)} Il latino in Gadda [The use of Latin in Gadda]
{5)} Annotazioni su composti in -cola [Annotations on compounds terminating in -cola]
a) Concordances by form of La cognizione del dolore
b) Concordances by form of Pasticciaccio
c) Complete concordances of the question mark (contexts with right-hand arrangement by ?)
d) Complete concordances of the question mark (contexts with left-hand arrangement by ?)
e) Complete concordances of the exclamation mark (contexts with left-hand arrangement by !)
f) Co-occurrences of Giornale di guerra e prigionia
g) Index Locorum of Latin forms in Gadda
h) Latin forms in Horace and Gadda - Comparison table
i) Comparisons between the two versions of Pasticciaccio
j) Gaddian iterations.

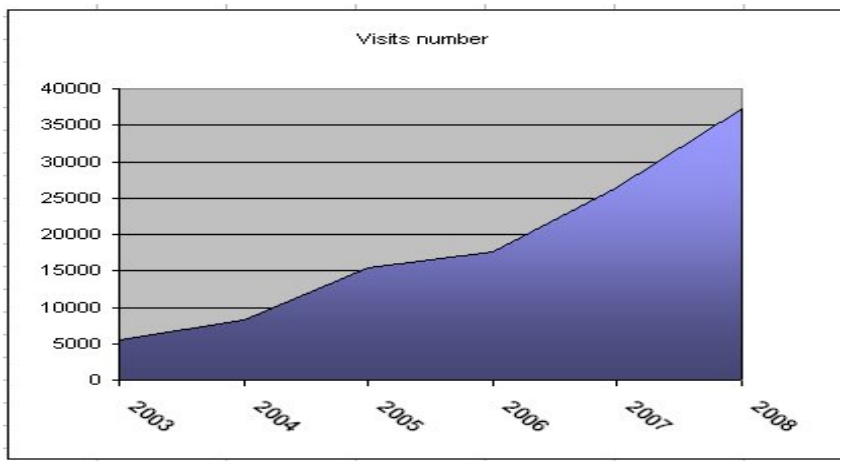


Figure 1 Visit numbers in year sequence

4. Concluding Remarks

On-line availability of tutorials and teaching materials on the web pages of Computational Linguistics courses in Italy is playing a two-fold role: a) supply a persisting lack of offer of Italian manuals from traditional publishers; b) provide a jumping-off point to general public internet users for a more aware and selective access to the web haystack, based on integration of personalized language-driven queries and precompiled information repositories and ontologies. In the near future, this is bound to change habits, demands and expectations of a new generation of self-taught web users, capable of exploiting the web in creative ways and pacing the development of novel technologies for information access.

References

- Barsalou, L. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences*, 22, 577–609.
- Buitelaar, P., Cimiano, P., and B. Magnini (2005). Ontology Learning from Text: an Overview. In Buitelaar et al. (eds.), Ontology *Learning from Text: Methods, Evaluation and Applications* (Volume 123 Frontiers in Artificial Intelligence and Applications): 3–12.
- Bybee, Joan, Hopper, Paul (eds.) (2001). *Frequency and the Emergence of Linguistic Structure*. John Benjamins.
- Guarino, N. (1998). Formal Ontology in Information Systems. *Proceedings of FOIS'98*. Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.
- Kilgarriff A., Grefenstette G. (2003). Introduction to the Special Issue on the Web as Corpus, *Computational Linguistics*, 23 (3), 333-47.
- F. W. Lancaster (2003). *Indexing and abstracting in theory and practise*. Third edition. London, Facet.
- Lawrence S., Lee Giles C. (1999). Accessibility of Information on the Web, *Nature*, 400, 107-9.
- Manning, Chris, Schütze, Hinrich (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press US
- Mitchell, Tom M. (1995). *Machine Learning*. McGraw Hill US.
- Ramos de Carvalho E.M. (2001). Grey Literature and its contribution to knowledge Society. *67th IFLA Council and General Conference*, 1-11.

OpenSIGLE - Crossroads for Libraries, Research and Educational Institutions in the field of Grey Literature

Dominic J. Farace and Jerry Frantzen, GreyNet, Netherlands
Christiane Stock and Nathalie Henrot, INIST-CNRS, France
Joachim Schöpfel, University of Lille, France

Abstract
This work is based on a paper[1] presented at the Tenth International Conference on Grey Literature (GL10) in which GreyNet's collections of conference preprints were made accessible via the OpenSIGLE Repository. OpenSIGLE offers a unique distribution channel for European grey literature with roots dating back a quarter century. In the first part of this paper, the experience of INIST as service provider and GreyNet as data provider will be discussed including recent developments.
Later in this paper, the draft of a project proposal called for in the final session of that conference will be elaborated. The proposal seeks to explore the capacity required for the OpenSIGLE Repository to develop in multilateral and international cooperation in support of European research infrastructures committed to the open access of grey literature collections and resources. Emphasis is placed on the involvement of libraries, research centers, and institutions of higher education, as well as, requirements for a grey literature network service to sustain further development, exploitation, and promotion of the OpenSIGLE Repository.

From SIGLE to OpenSIGLE: A Progress Report
SIGLE (System for Information on Grey Literature in Europe) was a unique multidisciplinary database dedicated to grey literature. Up to 15 European partners participated in SIGLE, mostly national libraries or libraries aligned to well-known research institutes. Their principal goals were the centralized collection of scientific and technical reports, theses and other grey material and to facilitate access to these documents through an engagement for document delivery or loan. Created in 1980 and produced from 1984 onwards by EAGLE (European Association for Grey Literature Exploitation), the database was last available through STN International and on CD-ROM via Silverplatter/Ovid until it became dormant in 2005. INIST then decided to make the data publicly available on an open access platform. Details of the migration from SIGLE to OpenSIGLE have been presented at the GL8 Conference[2] held in December 2006 (Schöpfel 2007). And in December 2007, the OpenSIGLE website[3] went live.
This paper further discusses three related issues dealing with OpenSIGLE: (1) usage statistics covering two years of access to the repository, (2) a bilateral cooperative agreement with GreyNet, the Grey Literature Network Service, and (3) a project proposal exploring the capacity required for the OpenSIGLE Repository to develop in multilateral and international cooperation.

OpenSIGLE Traffic Report
Usage information for a database is at all times interesting for the producer of the information. In this case an additional incentive was the fact that OpenSIGLE records, which migrated from the SIGLE database, had not been updated since 2005. Would then the move to an open access environment be at all "useful" for the grey literature community?
The usage analysis is based on data obtained through phpMyVisites, an open source software for website statistics that works with a javascript image call. Only completely uploaded pages are counted and robots are excluded. The following data provide only a part of the information that can be obtained through phpMyVisites. Other statistics based on server logs might however provide even higher figures.
The first figure shows that the number of visits as well as the number of page views has increased steadily since the opening of the website in 2007. A first peak was reached in July 2008 following a press campaign in the middle of the French holidays. The result is both surprising and rewarding since visits usually decrease during summer months.
The usage of OpenSIGLE continues to increase and has more than tripled between August 2008 and August 2009 in terms of page views and number of visits, where the average duration of a visit is 100 seconds. Visits where only a single page is viewed represent a stable 50% average of the traffic to the site. These users accessed the database after searching via Google or Google Scholar. While in other cases, users may carry out extensive searches and view hundreds of web pages.

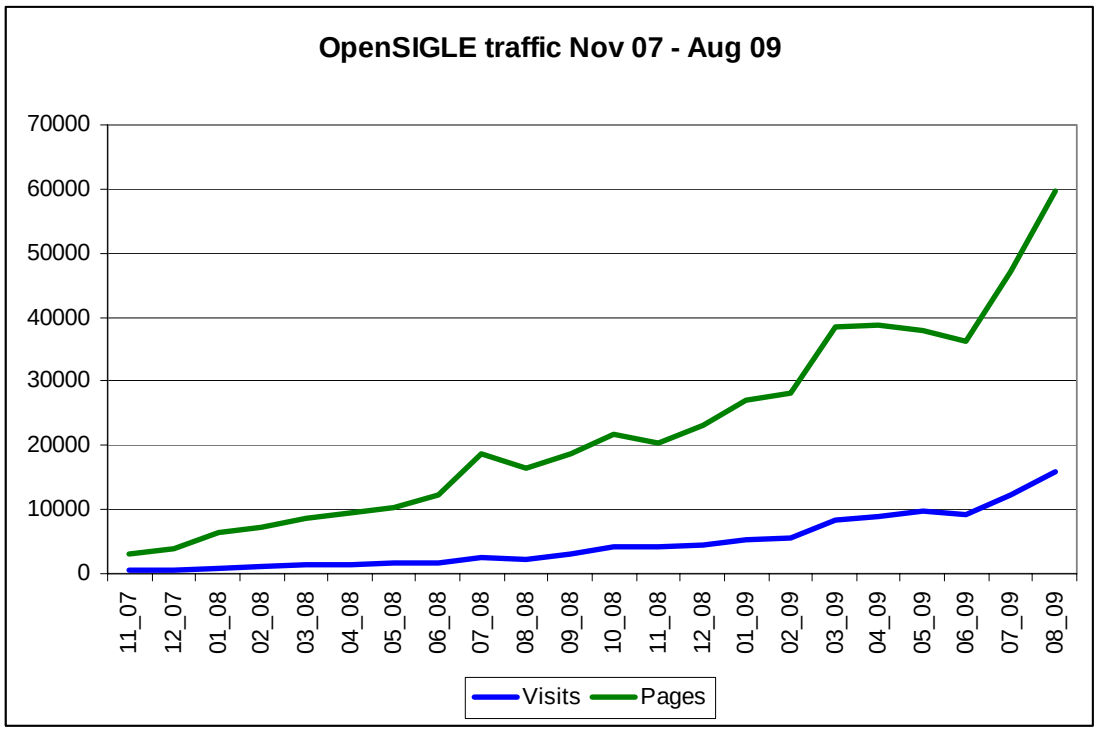


Figure 1: OpenSIGLE traffic report – number of visits and pages viewed

Geographic Origin of Visitors

The software used allows us to monitor the origin of visitors for the top ten countries each month. The sum of 22 months worth of data shows the United Kingdom in the lead, closely followed by the United States. A second group is formed by Germany, France, Italy and Spain. Countries in the long tail may not appear on a given monthly top ten listing. It is obvious that OpenSIGLE users come not only from the former EAGLE countries, but also from the United States, Canada, and since recently China and Australia. Clearly an indication that European grey literature presents an interest worldwide.

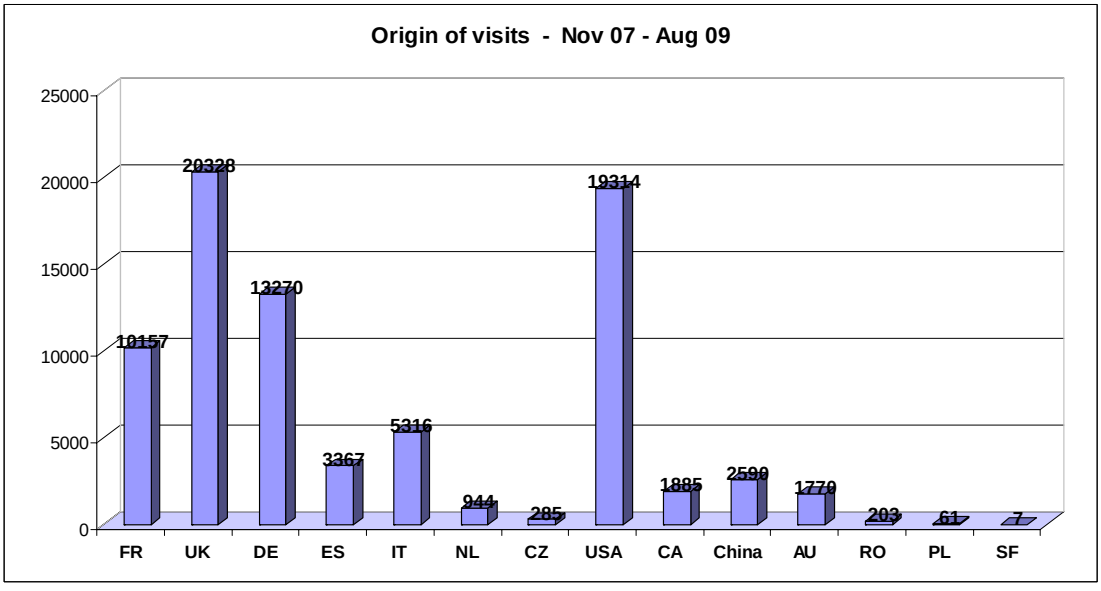


Figure 2: Origin of visitors to OpenSIGLE

Usage and Feedback

Compared to other INIST websites and e-resources, statistics show that 16% to 19% of the users come from North America. OpenSIGLE is in third place among users from this continent preceded by the English version of INIST's institutional website[4] and IndicaSciences[5] - an INIST product dedicated to research evaluation and indicators. INIST websites geared to a French speaking audience receive an average of 7% of the visits from North America.

The analysis of web links as well as feedback through incoming messages reveal that OpenSIGLE is often used in the biomedical and public health sectors. However, at present, statistics do not allow us to go into further detail regarding scientific domains.
During the course of 2008, several requests were received from former users of the STN or Ovid versions of the SIGLE database dealing with complex search strategies. Such questions required another look into the limits of the Jakarta Lucene search engine implemented within DSpace, especially with regard to the length of the search query. It was discovered that Lucene allows for more possibilities than mentioned in the help provided by DSpace. Besides inquiries involving search strategies, users were also interested in the download and export features of OpenSIGLE.
One critical view of OpenSIGLE found on a blog[6], mentions the absence of links to the full text of documents. Of course this is understandable given the fact that it was one of the very reasons why the SIGLE database was discontinued.

Promotional Activities
Before the official announcement of the launch of OpenSIGLE, the project was presented at a DSpace meeting focused on the exchange of experiences among its users (Grésillaud and Stock, October 2007) [7]. Shortly afterwards, and as a result of that meeting, visitors from Spain and Italy were observed on the OpenSIGLE website. In December 2007, INIST also focused attention on OpenSIGLE during the GL9 Information Walk-Thru at the Ninth International Conference on Grey Literature in Antwerp, Belgium[8].
In May 2008, a short presentation for the French public was given at I-expo (IT conference and exhibit) in Paris. And in July, INIST sent a press release to national and international lists and agencies *i.e*. Information World Review and Research Information. This no doubt resulted in the above mentioned peak of visits in the middle of summer. A brief message about OpenSIGLE was placed simultaneously on the French and international homepages of INIST. Since "news items" are normally less frequent during summer months, the message remained for a longer period of time on these WebPages.
Today OpenSIGLE is indexed by Google and Google Scholar and included in the bookmarks of national libraries and research institutes. Following the creation of the WorldWideScience Alliance and website[9] in June 2008, INIST (a partner in this Alliance) proposed to integrate OpenSIGLE into the WorldWideScience portal. This was realized in September 2008. And, in the web statistics that following month WWS.org appeared as forth partner site for visitors accessing OpenSIGLE through a website with GreyNet.org[10] following closely behind. Overall, these different promotional activities have had a positive impact on the use and branding of OpenSIGLE.

GreyNet, On the Background and Forefront of OpenSIGLE
In this section, the relationship between GreyNet and the former EAGLE Association including its SIGLE database will be addressed. This will then be followed by a conscious positioning of GreyNet in the newfound OpenSIGLE Repository with INIST as its Service Provider.
In 1992, EAGLE agreed to act as main sponsor for the launch of the International Conference Series on Grey Literature first held in the Amsterdam RAI in December 1993. GreyNet was at that time a newly established network service – driven on two fronts: (1) to promote the field of grey literature and the work of organizations involved in this branch of information the world over, and (2) to stimulate research on grey literature and make the results available both in print and digital (electronic) formats. EAGLE participated as sponsor and/or program committee member in the first five Conferences in the GL-Series.
In early 2005 GreyNet was invited as an observer to the final EAGLE Board meeting at FIZ Karlsruhe upon which the EAGLE Association formally voted to be dissolved. It was at that same meeting that the initial draft of an OpenSIGLE proposal[11] was presented by Dr. Joachim Schöpfel , last in line of EAGLE Presidents.
In the two ensuing years (2005-2007), INIST worked unilaterally on OpenSIGLE, which could then be described as a caretaker repository. In the autumn of 2007, once OpenSIGLE had become operational, GreyNet met with colleagues at INIST to hammer out an agreement that on the one hand would make GreyNet OAI-compliant and on the other hand would expand INIST's role in OpenSIGLE from solely a caretaker to an external service provider. To this end, GreyNet's conference based collections would provide an example of OpenSIGLE's potential for other data providers in the grey literature community.

GreyNet's Collections in OpenSIGLE
In December 2008, five years of research issuing from the GL Conference Series had been uploaded in the OpenSIGLE Repository. The bilateral contact between INIST as service provider and GreyNet as data provider was successful in customizing a metadata record for the enriched publication of conference preprints and the subsequent migration of GreyNet's collections to an

open access environment. The bilateral agreement likewise holds for future conferences in the GL-Series, continuing with GL10 records onward.
Retrospective input of the initial four conferences in the GL-Series (1993-1999) would of course make GreyNet's collections comprehensive in OpenSIGLE. To this end, in January 2009, GreyNet purchased from Emerald Group Publishing – former MCB University Press – the rights to allow the full-text papers from the earlier four conferences in the GL-Series to be made available in the OpenSIGLE Repository. This step was not only applauded by the open access community[12], but it also suggests other possibilities to retrieve content controlled by commercial publishers[13]. GreyNet proceeded with the production of metadata records, while INIST took on the work of scanning and creating image files for the retrospective records. In October 2009, half of the retrospective input had been achieved.

GreyNet's Potential for OpenSIGLE
The initial reaction from the grey literature community to GreyNet's alliance with OpenSIGLE has been positive; however, due to the brief timeframe in which GreyNet's collections are actually available in the OpenSIGLE Repository, it is too early to provide substantial user statistics. While GreyNet has been receiving monthly reports from INIST generated via OpenSIGLE, GreyNet is looking for other ways to compile use and user statistics via its own channels. In this way, there would be separate data issuing from INIST as service provider and GreyNet as data provider that would allow for comparisons and provide grounds for decision making in the future.
In September 2008, an OpenSIGLE webpage was added to the GreyNet website with hyperlinks to its conference collections already in the repository; and in January 2009 that webpage became a main page on GreyNet's website. Not only did the number of visits to the webpage double in the first half of 2009, but it now also allows for the addition of sub-pages used for promotional and instructive purposes.
The Grey Literature Network Service feels that it has even more to offer OpenSIGLE than its conference collections. Going back to 1992, when GreyNet was first launched, one of its primary goals was to promote the field of grey literature and the work of organizations involved in this branch of information. What EAGLE was to SIGLE, GreyNet could be to OpenSIGLE and more. GreyNet operates internationally and maintains a full-time established network service specializing in grey literature with information products and resources both in print and electronic formats. GreyNet has for the past seven years (2003-2009) often together with colleagues from INIST carried out research projects involving citation analysis, surveys, interviews, as well as standard review of the literature. Over the past years (1992-2009), GreyNet has developed channels for promotional outreach as well as a modest publishing arm. More recently, GreyNet has set up a program of training and instruction in the field of grey literature, which could also be linked to OpenSIGLE. These and other such initiatives would no doubt serve and support future developments in the OpenSIGLE Repository.

OpenSIGLE Project Proposal, A Feasibility Study
What began unilaterally with the vision and determination of INIST and what has recently been expanded in bilateral cooperation with GreyNet has yet even greater potential for the international grey literature community. GreyNet together with INIST are committed to drafting a project proposal. This proposal will explore the capacity required for the OpenSIGLE Repository to further develop in multilateral and international cooperation in support of European research infrastructures committed to open access of their grey literature collections and resources. One in which special emphasis is geared to libraries, research centers, and institutions of higher education.

Project Lead-Time
Both INIST and GreyNet have put forth a number considerations and recommendations based on their recent experience with the OpenSIGLE Repository. An inventory of issues and recommendations were collated and will be used in the development of work packages in the implementation phase of the project. Some of the issues include: closing time gaps in record entries, linking to full-text documents as well as plus links to datasets and software, integrating OpenSIGLE in other networks and portals, streamlining the SIGLE Classification scheme, etc.

Project Consortium
Based on the main objective of the proposed project and in relation to the issues that would have to be dealt with in order to achieve this objective, project partners and external advisors need to be identified and brought together in a consortium for the duration of the project. To achieve optimal results, the number of stakeholders in the project would be limited. In the diagram below, the content as well as management base of the project is visualized. However,

the names of the prospective organizations, who would be carrying out the projects' roles and tasks are masked here until final confirmation.

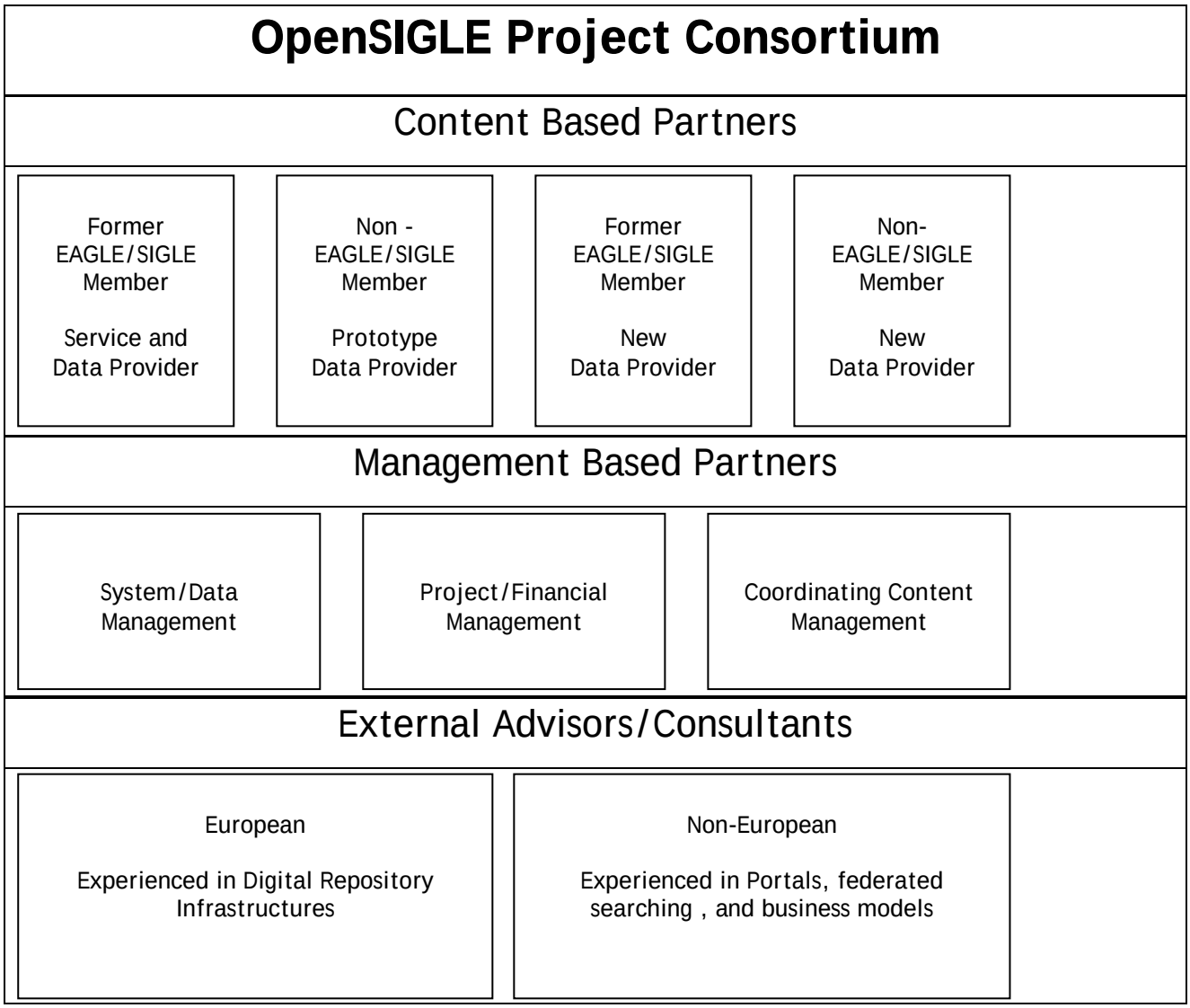


Figure 3: OpenSIGLE Project Consortium

Expected Results and Impact of the Project

This project has its roots in a European framework of cooperation among longstanding infrastructures including national libraries, research centers, and networked services. The outcome of this project would support and strengthen policy development for infrastructures in the field of grey literature, where open access to their collections and other knowledge based resources stand central. The OpenSIGLE Repository with its technical know-how would be sustained by a coordinating infrastructure in the advancement of European cross-disciplinary research well beyond its geographical borders. A draft of this project proposal will be presented during a Panel Session at the Eleventh International Conference on Grey Literature[14] held in the Library of Congress in Washington D.C on 14-15 December 2009 The panel members will take the opportunity to discuss the project proposal in order to illicit feedback from the international grey literature community, raise public awareness to the OpenSIGLE Repository, and solicit leads for further project funding.

Appendix: Timeline SIGLE – OpenSIGLE

2009
GreyNet's Retrospective Collections 1997-1999 in OpenSIGLE; GL11 – 11th Eleventh International Conference on Grey Literature in Washington D.C. USA with INIST as Co-Sponsor
2008
OpenSIGLE Repository included in WWS.org, World-Wide Science Portal; GreyNet's Collections 2003-2007 accessible via OpenSIGLE Repository; GL10 – 10th Tenth International Conference on Grey Literature in Amsterdam with INIST as Co-Sponsor
2007
OpenSIGLE operational as caretaker repository;
Bilateral agreement INIST (Service provider) – GreyNet (Data Provider); GL9 – 9th International Conference on Grey Literature in Antwerp, Belgium with INIST and the EU as Co-Sponsors
2006
SIGLE migration to OpenSIGLE with INIST as Service Provider; GL8 – 8th International Conference on Grey Literature in New Orleans with INIST as Co-Sponsor
2005
EAGLE Association dissolved and SIGLE Database dormant; INIST proposal for OpenSIGLE;
GL7 – 7th International Conference on Grey Literature in Nancy, France with INIST as Host and EU as Co-Sponsor
2004
GL6 – 6th International Conference on Grey Literature in New York NY with INIST as Co-Sponsor
2003
GreyNet re-launched; GL5 – 5th International Conference on Grey Literature in Amsterdam with INIST as Co-Sponsor and EAGLE on Program Committee
2000
GreyNet discontinued
1999
GL'99 – 4th International Conference on Grey Literature in Washington D.C. USA
1997
GL'97 – 3rd International Conference on Grey Literature in Luxembourg with EC as Host and EAGLE on Program Committee
1995
GL'95 – 2nd International Conference on Grey Literature in Washington D.C. USA;
1993
GL'93 – 1st International Conference on Grey Literature in Amsterdam, EAGLE Sponsor; Bi-Annual GL-Conference Series 1993-1999 with EAGLE on the Program Committee
1992
GreyNet founded
1985
EAGLE Association was founded as producer of the SIGLE Database
1980
SIGLE established as database and document delivery system; Later exploited via STN and Blaise Hosts as well as SilverPlatter CD-ROM
1978
York Seminar on Grey Literature hosted by British Library and the EC

References

[1] Farace, D.J., J. Frantzen, C. Stock, N. Henrot, and J. Schöpfel (2009), OpenSIGLE, Home to GreyNet's Research Community and its Grey Literature Collections: Initial Results and a Project Proposal. – In: The Grey Journal : An International journal on Grey Literature, vol. 5, no 1, Spring 2009. ISSN 1574-1796

[2] Schöpfel, J., C. Stock, and N. Henrot (2007), From SIGLE to OpenSIGLE and beyond: An in-depth look at Resource Migration in the European Context. – In: The Grey Journal : An International journal on Grey Literature, vol. 3, no 1, Spring 2007. ISSN 1574-1796

[3] OpenSIGLE - System for Information on Grey Literature in Europe, http://opensigle.inist.fr/

[4] English version of INIST's institutional website, http://international.inist.fr

[5] INIST product dedicated to research evaluation and indicators, http://indicasciences.inist.fr

[6] Critical view of OpenSIGLE found on a blog, http://healthinformaticist.wordpress.com/2008/08/28/does-opensigle-exist-for-its-own-sake/

[7] Grésillaud, S., and C. Stock (October, 2007), DSpace at INIST-CNRS: one platform, different usages and resulting specific needs/problems. Paper presented at DSpace User Group Meeting 2007, Food and Agriculture Organization of the United Nations, Rome, Italy. Available at http://www.aepic.it/conf/viewabstract.php?id=208&cf=11

[8] Grey Foundations in Information Landscape (2007), Ninth International Conference on Grey Literature, 10-11 December 2007 in Antwerp, Belgium. – GL9 Conference Program and Abstracts. – ISBN 978-90-77484-09-8

[9] WorldWideScience.org, the global science gateway, http://worldwidescience.org/

[10] GreyNet, Grey Literature Network Service, http://www.greynet.org/

[11] Schöpfel, J. (2006), MetaGrey Europe, A Proposal in the Aftermath of EAGLE-SIGLE. – In: GL7 Conference Proceedings, pp. 34-39. – ISBN 90-77484-06-X

[12] Posting by Peter Suber on January 29 (2009), http://www.earlham.edu/~peters/fos/2009/01/greynet-buys-rights-to-deposit-papers.html

[13] Posting by Heather Morrison on August 6 (2009), http://www.connotea.org/comments/uri/92b11113ecf827be19a369f21e81161b

[14] GL11 Program and Conference Bureau, http://www.textrelease.com/

Appendices

Author Information	130-132
List of Participating Organizations	133-134
GL12 Conference Announcement	135
Index to Authors	137
GL11 Publication Order Form	138

Author Information

Asserson, Anne **54**
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 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) cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS is now the custodian of the CERIF model eurocris.org. Anne Asserson is a member of the Best Practice Task Group. anne.asserson@fa.uib.no

Biagioni, Stefania **92, 94,104**
Graduated in Italian Language and Literature at the University of Pisa and specialized in data processing. She is currently a member of the research staff at the Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI), an institute of the Italian National Research Council (CNR) located in Pisa. She is head librarian and member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She has been the responsible of ERCIM Technical Reference Digital Library (ETRDL) and currently of the PUMA (PUblication MAnagement) & MetaPub, a service oriented and user focused infrastructure for institutional and thematic Open Access repositories looking at the DRIVER vision, http://puma.isti.cnr.it. She has coauthored a number of publications dealing with digital libraries. Her activities include integration of grey literature into library collections and web access to the library's digital resources, including electronic journals and databases. She is a member of GreyNet since 2005. Email: stefania.biagioni@isti.cnr.it

Blaaij, Cees de **37**
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 GL on the Internet. Email: cdbl@zebi.nl

Carroll, Bonnie C. **28**
President of Information International Associates, Inc. (IIa), supports government and industry in managing information as a strategic resource. She is Secretariat Director of CENDI, the federal scientific and technical information (STI) managers' group and Consultant to USGS, supporting the development of the National Biological Information Infrastructure and the OSTP Biodiversity Informatics Working Group to promote interagency coordination of national and international biodiversity informatics initiatives. For over 3 decades, she has participated in research and development projects including a major assessment of STI in the U.S. for the National Science Foundation; the development of a topography of STI systems for the Library of Congress; and the development of a system to provide input to the DOE's Energy Science and Technology Database. She helped develop a National Information Strategy for the Kingdom of Jordan and has done special studies for the International Atomic Energy Agency, UNESCO and the World Bank. Before starting IIa, Ms. Carroll worked at DOE's Office of Scientific and Technical Information as Director of Program Development and International Activities and Deputy and Acting Assistant Manager for Information Services. She worked with contract research companies, Oak Ridge National Laboratory, and Cornell University. She served as President of the American Society for Information Science &Technology and as Chair of the American Association for the Advancement of Science Section on Information, Computing and Communications. Ms. Carroll is currently the US National Representation to International CODATA under the auspices of the International Committee for Science (ICSU) and is on the editorial board for the journal Information Services and Use. She has an MS from Columbia University and a BA from Cornell University. Ms. Carroll has been active in gray literature research and management dating back to the early 1990s. Email: bcarroll@infointl.com

Castelli, Donatella **92, 94**
Joined the CNR-ISTI in 1987. Since then she has participated actively in several EU and Nationally funded projects on Digital Libraries and Research Infrastructures. Among them: "ERCIM Digital Library-DELOS", DELOS ECHO, SCHOLNET (as scientific co-ordinator), CYCLADES, Open Archives Forum, DLib Competence Center, DILIGENT (as scientific co-ordinator) and DRIVER. Currently she is the Scientific Co-ordinator of the D4Science and D4Science-II projects. She is also involved in the BELIEF-II, DRIVER-II, OpenAIRE, DL.org and GRDI2020 FP7 projects. Her research interests include digital libraries content modelling, interoperability and architectures. Email: donatella.castelli@isti.cnr.it

Crowe, June **28**
Senior Researcher at Information International Associates, Inc. (IIa). She received her AMLS from the University of Michigan, Ann Arbor and her M.Ed. in geographic education from the University of Georgia, Athens. She has extensive experience in the management and operations of library services across government, public, academic, and special libraries. At IIa she performs open source research on a variety of topics and manages research projects involving web-based open source document identification, collection, and processing. Her primary interests are open source information in Grey Literature, repositories, and open source intelligence tools. Email: jcrowe@iiaweb.com

Di Cesare, Rosa **82**
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. Email: biblio.irpps@irpps.cnr.it

Fairclough, Ian **47**
Cataloging & Metadata Librarian at George Mason University. A native of Liverpool, England, in 1975 he moved to Toronto to study musicology and since 1977 has lived all across the United States. He has served in numerous positions in a wide range of libraries -

academic, public, and special; large, medium, and small. His M.L.S. degree is from Queens College, CUNY. From 2001 to 2008 he lived in Marion, Ohio and worked on the River Valley Collection repository of grey literature. Email: ifairclo@gmu.edu

Farace, Dominic **122**
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-seven years. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences, 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; Instructor of Grey Literature via UNO Distance Education, and Editor of TGJ, The Grey Journal. Email: info@textrelease.com

Gheen, Tina **68**
Library Director at the National Science Foundation (NSF) in Arlington, Virginia, where she manages library operations, projects, and staff. She is particularly interested in making the GL in the NSF Library collection machine readable and searchable so that it may be leveraged for future scientific work. She received her MLS from the University of North Texas and her BS in Biology from Southwest Texas State University. She is currently a representative of CENDI, an interagency working group of senior scientific and technical information managers from 13 U.S. federal agencies; FLICC, the Federal Library and Information Center Committee; the FEDLINK Advisory Council; and the Naval Research Laboratory Alliance. Email: tgheen@nsf.gov

Giannini, Silvia **104**
Graduated and specialized in library sciences. Since 1987 she has been working in Pisa at the Institute for the Science and Technologies of Information "A. Faedo" of the Italian National Council of Research (ISTI-CNR) as a librarian. She is a member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She is responsible of the library automation software "Libero" in use at the CNR Research Area in Pisa and coordinates the bibliographic and managing activities of the ISTI library team. She cooperates in the design and development of the PUMA (PUblication MAnagement) & MetaPub, an infrastructure software for institutional and thematic Open Access repositories of published and grey literature produced by CNR. silvia.giannini@isti.cnr.it

Henrot, Nathalie **122**
Graduated in History, then in Information Sciences from the University of Tours in 1988. She has been working for the INIST-CNRS for seventeen years, more specifically at the Monographs & Grey Literature Section from 1993, for congress proceedings acquisition. She is now the user administrator in the OpenSIGLE project.
Email: henrotn@inist.fr

Ikeda, Kiyoshi **111**
Librarian at the Central Library of Japan Atomic Energy Agency (JAEA). He joined JAERI (former JAEA) in 2004, and had been working for acquisition of library materials (domestic/foreign books, journals and technical reports). Since 2007, he is in charge of editing work at "JAEA Reports" and "JAEA R&D Review" which are grey literatures. He is also a member of editorial board of the Japan Special Library Association (JSLA) since 2008. He holds a degree in library and information science (M.A.) from the Graduate School of University of Library and Information Science (in Japan).
Email: ikeda.kiyoshi@jaea.go.jp

Jeffery, Keith G. **54**
Director, IT and International Strategy of STFC (Science and Technology Facilities Council), based at Rutherford Appleton Laboratory in UK. Previously he was 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. STFC 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 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

Lin, Yongtao **19**
Librarian at Tom Baker Cancer Knowledge Center, Health Information Network, University of Calgary. Yongtao is responsible for managing the Tom Baker Cancer Knowledge Centre, where she provides library services to clinicians, staff, patients and families. Prior to this position, she worked as a hospital librarian in Nova Scotia, after graduating from the School of Information Management from Dalhousie University in 2004. She also has a Bachelor of Education and taught in the UK. Her research interests include evidence-based clinical practice, grey literature, problem-based learning, and health information literacy. Email: yolin@ucalgary.ca

Lines, Michael **59**
Law librarian at the University of Victoria Diana M. Priestly Law Library in British Columbia, Canada. He has a background in classical and medieval history, and has strong interests in archival studies, legal history, and promoting access to information. Email: mlines@uvic.ca

Luzi, Daniela **82**
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

Manghi, Paolo **92**
Research Fellow at the NMIS lab of the Istituto di Scienza e Tecnologie dell'Informazione (ISTI), Consiglio Nazionale delle Ricerche, Pisa, Italy. His interests are in the fields of Data Models for Digital Libraries, Types for Compound Objects in Digital Repositories, Digital Library Systems and Services, Service-Oriented Infrastructures for Digital Libraries, Database Systems, Type systems for XML languages, Query languages for XML data, XML P2P database systems. Member of the core expert group of Europeana and of the DL.org workgroup on content interoperability. Email: Paolo.Manghi@ISTI.CNR.IT

Marzi, Claudia **118**
Graduated in Foreign Languages and Literature at Pisa University in 1998, with the dissertation "The power of words: language creativity in Edgar Allan Poe's narrative". Since 2001 she has been working at the CNR Institute for Computational Linguistics in Pisa. She contributed to many national (funded by Italian Ministry of Research) and international (funded by European Commission) projects. Recently she was principal organizer of the European Science Foundation Exploratory Workshop "Words in action: Interdisciplinary approaches to understanding word processing and storage" (Pisa, Oct. 2009) E-mail: claudia.marzi@ilc.cnr.it

Oxnam, Maliaca **65**
Science-engineering librarian at the University of Arizona Libraries (UAL) since 2000. Her leadership roles include leading several strategic projects for UAL, including the Strategic Long Range Planning for the Libraries. Maliaca has also served as a co-PI on the development of the Geotechnical Rock & Water Resources Digital Library (GROW) and as a key participant on Einstein's Protégé's": A Heuristic approach to Bridging the Gap between Faculty Expectations and Student Preparedness funded respectively by the NSF and the Department of Education. Maliaca has served as vice-chair/chair-elect of the ALA ACRL Science & Technology Section and serves as chair of the Technical Report Archive & Image Library (TRAIL) Taskforce on a joint project for the Greater Western Library Alliance and Center for Research Libraries aimed at enabling and increasing access to US scientific grey literature. E-mail: maliaca@u.arizona.edu

Pardelli, Gabriella **118**
Graduated in Letters at the University of Pisa in 1980 and has been working at the Institute for Computational Linguistics (ILC) of the National Research Council in Pisa since 1984. She has been active in the creation of bibliographical databases for Natural Language Processing (NLP) and Digital Library in the Humanities. Other interests regard terminology and History of Human Language Technology. She is responsible for the Library of the ILC and the collection called "Antonio Zampolli Fund". Has participated in many research projects and has worked on the creation of bibliographical resources in the field of language technologies. She has presented many works at different national and international conferences and congresses. gabriella.pardelli@ilc.cnr.it

Pavlov, Leonid P. **13**
Graduated from Moscow Physical-Engineering Institute, Dipl. Eng. in computer systems. He is a Candidate of Sciences in informatics; and since 1976 is employed with the Scientific and Technical Information Centre of Russia (VNTIC) as Deputy Director. Main works in information systems, scientific and technical information, and grey literature. Email : pavlov@vntic.org.ru

Prost, Hélène **74**
Works since 1995 as a librarian at INIST-CNRS and is specialised in the evaluation of collections and document supply. Actually she is preparing a new SS&H database as a part of the FRANCIS database. She obtained a Master in History in 1991 and a Master in Scientific Information in 1993. She published on statistical and bibliometric analysis of information. Email: prost@inist.fr

Sassi, Manuela **118**
Graduated in Foreign Languages and Literature at Pisa University, 110/110 cum laude. Since 1974 she has been working in Pisa at the Institute for Computational Linguistics of the National Research Council. Her interests and experiences range from linguistic to textual data processing and in providing linguistic resources on-line. She has been responsible for many national projects and has participated in numerous international projects. Email: manuela.sassi@ilc.cnr.it

Schöpfel, Joachim **74, 122**
Joachim Schöpfel obtained his Ph.D. in psychology from the Hamburg University in 1992. During his studies in psychology, he participated in research on bilingual children of Turkish immigrants in Hamburg, of the German minority in Denmark, and in a French-German High School in Versailles, France. From 1991 to 2008, he worked at the French Institute for Scientific and Technical Information (INIST-CNRS) in different positions in database production and library management, at last as head of the e-publishing and document supply department. During the same time, he was lecturer at the University of Nancy. At present, he is senior lecturer in information and communication sciences at the Charles de Gaulle University of Lille 3. He published on GL, document delivery, digital libraries, scientific publishing, usage statistics and professional development.
Email: joachim.schopfel@univ-lille3.fr

Stock, Christiane **122**
Christiane Stock is the Head of the Monographs and Grey Literature service at INIST, in charge of the repositories LARA (reports), mémSIC (master's theses in information sciences) and OpenSIGLE. Member of the Technical Committee for the SIGLE database from 1993 to 2005, she also set up the national agency for ISRN (International Standard Report Number). She is member of the AFNOR expert group who prepared the recommended metadata scheme for French electronic theses (TEF). Email: christiane.stock@inist.fr

Vaska, Marcus **19**
Marcus Vaska is a librarian at the University of Calgary's Health Sciences Library, where he provides research assistance and support to clients within the Faculty of Medicine and the Calgary Health Region. In addition to exploring innovative instruction techniques for his classes, he has developed newfound appreciation for the pursuit of literature that exists beyond traditional publishing channels. Marcus' current interests focus on the teaching methods and experiences of librarians involved in delivering grey literature searching sessions to academic audiences in a medical setting.
Email: mmvaska@ucalgary.ca

Young, Peter R. **11**
As Chief of the Asian Division at the Library of Congress since November 2008, Peter R. Young provides leadership, direction, and planning for Asian Division programs and services. The Division manages the most comprehensive library collections of materials outside of Asia representing China, Japan, Korea, South Asia, Southeast Asia, and Asian Pacific Americans. With over 3.1 million items, the Asian Division collections represent a unique resource. Young is responsible for the Library's strategy for digital information related to science and technology. Email: pyou@loc.gov

Zoppi, Franco **94**
Franco Zoppi has been working for a long time in the design and implementation of software systems in the areas of DBMS, Distributed Office Information Systems and Digital Library Systems. Initially employed at Research and Development Department of Olivetti S.p.A., then at the Network Laboratory of the Telecommunications Department of Telecom Italia, in 2001 he joined the Information System Department of the University of Pisa as Project Manager. Since 2005 he has been working as Research Associate at the CNR-ISTI, where he coordinates the CNR activities in the BELIEF-II project. He is also involved in the DRIVER-II, EFG and HOPE FP7 projects. Email: franco.zoppi@isti.cnr.it

List of Participating Organizations

American Veterinary Medical Association, AVMA	United States
Battelle	United States
British Library, BL	United Kingdom
Catholic University of America, CUA	United States
Centre National de Recherche Scientifique, CNRS	France
Centre of Information Technologies and Systems, CITIS	Russia
Chiba University	Japan
Dalhousie University, DAL	Canada
Defense Technical Information Center, DTIC	United States
Department of Energy, DOE	United States
EBSCO Publishing	United States
ERIC Digital Library	United States
euroCRIS, Current Research Information Systems	United Kingdom
Federal Library and Information Center Committee, FLICC	United States
Federal Library Information Network, FEDLINK	United States
Florida State University, FSU	United States
George Mason University	United States
George Washington University, GWU	United States
Grey Literature Network Service, GreyNet	Netherlands
Information International Associates, IIa	United States
Institut de l'Information Scientifique et Technique, INIST	France
Institute of Ecosystem Study, ISE-CNR	Italy
Institute of Information Science and Technologies, ISTI-CNR	Italy
Institute of Research on Population and Social Policies, IRPPS	Italy
International Council for Scientific and Technical Information, ICSTI	France
Istituto di Linguistica Computazionale, ILC	Italy
Japan Atomic Energy Agency, JAEA	Japan
Law Library of Congress	United States
Lexis-Nexis	United States
Library of Congress, LC	United States
Library of Zeeland, ZEBI	Netherlands
Marion Public Library	United States
National Drug Intelligence Center, NDIC	United States
National Institute of Health, NIH	United States

List of Participating Organizations

(continued)

National Library of Agriculture	United States
National Library of Medicine, NLM	United States
National Research Council, CNR	Italy
National Science Foundation, NSF	United States
National Technical Information Service, NTIS	United States
National Technical Library, NTK	Czech Republic
Naval Research Laboratory, NRL	United States
New York Academy of Medicine, NYAM	United States
Office of Scientific and Technical Information, OSTI	United States
Open Source Center	United States
Oregon Health & Science University, OHSU	United States
Portland State University, PSU	United States
Pratt Institute, School of Information and Library Science	United States
PricewaterhouseCoopers, PwC	Netherlands
ProQuest	United States
Science and Technology Facilities Council, STFC	United Kingdom
Scientific and Technical Information Center, VNTIC	Russia
Texas A&M University	United States
TextRelease	Netherlands
Tom Baker Cancer Knowledge Centre	Canada
Total Support Systems Corporation, TOSS	Japan
United States Department of Agriculture, USDA	United States
United States Department of Education	United States
Université Charles de Gaulle Lille 3	France
University of Arizona Libraries	United States
University of Bergen, UiB	Norway
University of Calgary	Canada
University of California, Irvine Libraries, UCI	United States
University of Illinois, UIUC	United States
University of Ljubljana, UNI-LJ	Slovenia
University of Victoria	Canada
Wayne State University	United States

GL 12

pantone red 032c

TWELFTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

web qr code

National Technical Library (hereafter referred to as "NTL") is central professional library open to public, which offers unique collection of 250 thousand publications freely accessible in open circulation. Its holdings form the largest collection of Czech and foreign documents from technology and applied natural sciences as well as associated social sciences. It contains total of 1,2 Mil. volumes of books, journals and newspapers, theses, reports, standards, and trade literature in both printed and electronic forms. Besides its own collection, parts of Central Library of CTU in Prague and Central Library of ICT holdings are accessible in NTL.

Held in the National Technical Library in Prague, Czech Republic, on December 6 – 7, 2010.

As corresponds to its statutes the NTL runs – among others – the project of building the **National Repository of Grey Literature**. The project aims at gathering metadata and possibly full texts of grey documents in the field of education, science and research. The NTL supports an education in the field of grey literature through annual seminars in the Czech Republic.

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Technical Library

For the latest news on GL12 or National Repository of Grey Literature, visit our Web site at **http://nrgl.techlib.cz/**

Index to Authors

A

Asserson, Anne 54

B

Bastrykin, Aleksandr M. 13
Biagioni, Stefania 92, 94,104
Blaaij, Cees de 37
Bogdanov, Yuri M. 13

C

Carroll, Bonnie C. 28
Castelli, Donatella 92, 94
Cerbara, Loredana 82
Crowe, June 28

D

Di Cesare, Rosa 82

E

Ebisawa, Naomi 111

F

Fairclough, Ian 47
Farace, Dominic 122
Frantzen, Jerry 122

G

Gheen, Tina 68
Giannini, Silvia 104
Gonda, Mayuki 111

H

Henrot, Nathalie 122

I

Ikeda, Kiyoshi 111
Itabashi, Keizo 111

J

Jeffery, Keith G. 54

L

Lin, Yongtao 19
Lines, Michael 59
Luzi, Daniela 82

M

Manghi, Paolo 92
Marzi, Claudia 118

N

Nozawa, Takashi 111

O

Olmsted, Sue 68
Oxnam, Maliaca 65

P

Pardelli, Gabriella 118
Pavlov, Leonid P. 13
Prost, Hélène 74

R

Ruggieri, Roberta 82

S

Sassi, Manuela 118
Schöpfel, Joachim 74, 122
Starovoitov, Aleksandr V. 13
Stock, Christiane 122

V

Vaska, Marcus 19

Y

Young, Peter R. 11

Z

Zoppi, Franco 94

GreyNet Members less 20%

The Grey Mosaic, Piecing It All Together

Publication Order Form

ELEVENTH INTERNATIONAL CONFERENCE ON GREY LITERATURE *Washington D.C., 14-15 December 2009*

	No. of Copies	x	*Amount in Euros*	**Subtotal**
Forthcoming Publications (February 2010)				
GL11 CONFERENCE PROCEEDINGS – Printed Edition ISBN 978-90-77484-13-5 ISSN 1386-2316 *Postage and Handling excluded*[*)]		**x**	**95.00 =**	€
GL11 CONFERENCE PROCEEDINGS - CD-Rom Edition ISBN 978-90-77484-13-5 ISSN 1386-2316 *Postage and Handling included*		**x**	**95.00 =**	€
EXCLUSIVE OFFER: **GL11 Conference Proceedings** - CD-Rom Edition With accompanying PowerPoint Presentations *Postage and Handling included*		**x**	**120.00 =**	€
POSTAGE AND HANDLING PER PRINTED COPY[*)]				
Holland		**x**	**5.00**	€
Europe		**x**	**10.00**	€
Other		**x**	**20.00**	€
			TOTAL	**€**

Customer information

Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
E-mail Address:	

Upon receipt of payment the publication(s) will be forwarded to your shipping address!

❑ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to "GL11 Publication Order"

❑ MasterCard/Eurocard ❑ Visa card ❑ American Express

Card No. _____ Expiration Date: _____

Print the name that appears on the credit card, here _____

Signature: _____ **CVC II code:** _____ *(Last 3 digits on signature side of card)*

Place: _____ Date: _____

NOTE On receipt of payment, an invoice marked paid will be sent to your postal address. If you require a *pro forma* invoice, please notify:

TextRelease
www.textrelease.com

GL11 Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
T/F +31-(0) 20-331.2420 Email: **info@textrelease.com**