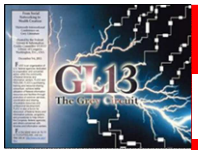
Thirteenth International Conference on Grey Literature

Library of Congress, Washington D.C. USA, 5-6 December 2011



The Grey Circuit

From Social Networking to Wealth Creation

Conference Proceedings

ISSN 1386-2316

Host and Sponsors

FLICC FEDLINK

DTIC Online Information for the Defense Community

ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO"

CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR

JST

NTK 50°6'14.083"N, 14°23'26.365"E Národní technická knihovna National Technical Library

cnrs inist

Ila Information International Associates

NEW YORK ACADEMY OF MEDICINE 1847

KiSTi www.kisti.re.kr

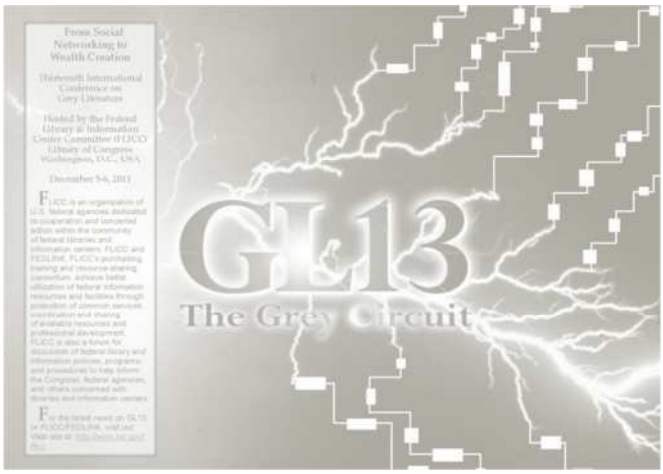
GreyNet Grey Literature Network Service

EBSCO INFORMATION SERVICES

GL13 Program and Conference Bureau

TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, The Netherlands
www.textrelease.com · **conference@textrelease.com**
Tel/Fax +31-20-331.2420



CIP

GL13 Conference Proceedings
Thirteenth International Conference on Grey Literature : The Grey Circuit, From Social Networking to Wealth Creation, 5-6 December 2011 / compiled by D.J. Farace and J. Frantzen ; GreyNet, Grey Literature Network Service. – Amsterdam : TextRelease, February 2012. – 184 p. – Author Index. – (GL conference series, ISSN 1386-2316 ; No. 13)
EBSCO Publishing, FLICC-FEDLINK, INIST-CNRS, JST, NTK, and NYAM are corporate authors and associate members of GreyNet International. These conference proceedings contain the full text of nineteen papers presented during the two days of plenary and poster sessions. The papers appear in the same order as in the conference program book. Included is an author index along with biographical notes, a list of participating organizations, as well as sponsored advertisements.

ISBN: 978-90-77484-17-3

Foreword

The Grey Circuit

From Social Networking to Wealth Creation

Social networking is the way the grey literature community remains connected in the 21st century. It encompasses a range of social media and communication tools that enable subject based communities to create, review, process, publish, and make grey literature openly accessible to public domain. Social networking is not new to grey literature, in fact it is inherent to this field of information. What's new however are the technologies available to global grey literature communities in developing, monitoring, and sustaining valued information resources and services. In this context, social networking becomes a mechanism both used and applied by grey literature communities in the processes of knowledge generation and ensuing wealth creation. The Thirteenth International Conference on Grey Literature is beholden to its title: The Grey Circuit, from Social Networking to Wealth Creation. The imagery in the conference logo rekindles the spirit of Franklin in which the uncontrolled discharge of lightening transfers power to controlled networked circuits. Today, in the spirit of science, grey literature communities are called to demonstrate their know-how and merit to wider audiences.

Dr. Dominic J. Farace
Grey Literature Network Service

Amsterdam,
February 2012

GL13 Conference Sponsors

FLICC/FEDLINK, USA
Library of Congress

DTIC, USA
Defense Technical Information Center

EBSCO, USA
Information Services

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

JST, Japan
Japan Science and Technology Agency

NTK, Czech Republic
National Technical Library

GL13 Conference Sponsors

ISTI-CNR, Italy
Institute of Information Science and Technologies

NYAM, USA
New York Academy of Medicine

IIA, USA
Information International Associates, Inc.

CVTI-SR, , Slovak Republic
Slovak Centre of Scientific and Technical Information

KISTI, Korea
Korea Institute of Science and Technology Information

GL13 Program Committee

Blane K. Dessy ***(Chair)***
FLICC/FEDLINK
USA

Janice Kaplan
NYAM, New York Academy of Medicine
USA

Bonnie C. Carroll
IIA, Information International Associates, Inc.
USA

Roberta Shaffer
LLC, Law Library of Congress
USA

Petra Pejšová
National Technical Library, NTK
Czech Republic

Joachim Schöpfel
University of Lille 3
France

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

Daniela Luzi
CNR, National Research Council
Italy

Dominic J. Farace
GreyNet, Grey Literature Network Service
Netherlands

Table of Contents

OPENING SESSION

Chair, *Blane K. Dessy, Director FLICC-FEDLINK; Library of Congress, United States*

SESSION ONE – SOCIAL NETWORKING

Chair, *Joachim Schöpfel, University of Lille, France*

SESSION TWO – SPECIAL COLLECTIONS

Chair, *Stefania Biagioni, Institute of Information Science and Technologies; National Research Council, Italy*

SESSION THREE – OPEN ACCESS AND WEALTH CREATION

Chair, *Seon-Hee Lee, Korea Institute of Science and Technology Information, South Korea*

SESSION FOUR – DATA FRONTIERS

Chair, *Janie Kaplan, New York Academy of Medicine, United States*

POSTER SESSION

Chair, *Robin Harvey, FLICC-FEDLINK; Library of Congress, United States*

Iia
Information
International
Associates

Navigators in a Sea of Data

For 23 years, Iia has provided quality information services to advance our clients' mission. By applying information science skills in combination with information technology, we help organizations handle massive amounts of information and make it useable.

Iia makes information an asset for your organization.

Corporate Office:
104 Union Valley Road
Oak Ridge, TN 37830
Phone: 865- 481-0388

D.C. Area Office:
417 West Broad Street
Suite 201
Falls Church, VA 22046
Phone: (703) 407-5770

www.iiaweb.com

Science-Forums.net
A Platform for Scientific Sharing and Collaboration

Marilyn ("Lynn") J. Davis,
USDOE Office of Scientific and Technical Information
United States

Edrick Coppock and **Lance Vowell,**
Information International Associates, Inc.
United States

Abstract
The beta website of Science-Forums.net was developed by Information International Associates, Inc. (IIa) through the Department of Energy (DOE) Small Business Innovation Research (SBIR) program. By leveraging scientist/researcher information in databases managed by DOE's Office of Scientific and Technical Information (OSTI), Science-Forums.net will allow scientists to easily collaborate on a scientific topic of interest. In addition to potentially yielding commercial applications, the functionality of Science-Forums.net is well suited to OSTI's family of scientific and technical information (STI) websites for research and development.
The objective of the SBIR—"Interactive Peer-to-Peer (P2P) Scientific Communication in the Digital Library Environment"—was to research and identify web-based tools and other concepts that will foster online interaction and collaboration among scientists and researchers. These tools can facilitate scientific discovery and innovation.
Phase I methods included (1) survey of available technologies, (2) analysis and prioritization of technologies, (3) design and planning of a foundational tool for online interaction, (4) development and implementation of approaches to obtain feedback from researchers and scientists, (5) research into privacy and communications examples and guidelines to address government privacy and survey requirements, and (6) investigation into scientists' use and interest in social media tools for research and communication.[1]
Phase II has focused on (1) development of the tool into a beta-ready site for feedback from a test group of previously identified, interested researchers, (2) refinement of the privacy policy and user terms and conditions documents, and (3) further research of bibliographic/metadata controls to create and exchange documents and other content.

Motivation/Justification for the Research
This project began with the hypothesis that web 2.0 technologies can promote information exchange, collaboration, and networking among scientists (P2P) and that they can also facilitate scientific discovery and innovation. Furthermore, there was a strong belief that scientific researchers would want to collaborate in a different manner than that provided by the typical social network (e.g., Facebook, LinkedIn, Second Life, or Twitter). The system should support a broad range of information types: technical reports, grey literature, theses, research notes and papers in progress, as well as social interaction among users, blogs and forums open to the general research community, and those open only to a select group of members.
The primary goal of the project was to research and develop a prototype tool to enable interactive P2P scientific communication in the digital library setting. This prototype tool would provide mechanisms to enable scientific and research communities, as defined by common interest, to communicate based on content provided by a digital library or another collection of information (e.g., grey literature in a specific topic). By increasing the ability for collaborators to interact, this tool will improve the diffusion of scientific knowledge.

These digital libraries, or literature collections, provide a unique opportunity to facilitate communication between scientific communities of common interest. These groups can add functionality to their electronic

dissemination products (i.e., web-based database search products) to allow users to communicate and collaborate using the technical information they seek as the catalyst for discussion and collaboration. To date, agencies within the governmental digital library environment have not deployed this type of collaboration technology. This P2P communication is a less formal scientific mechanism than a technical report. Because authors typically have some mechanism for users to contact them directly, implementation of P2P communications should be straightforward.

Discussion and collaboration invariably lead to creation of new content. Collaborative technology enables informal content creation through comments and discussion, along with the more formal creation of published and grey literature, using document and bibliographic/ metadata formatting interchange standards. Mechanisms to perform a literature search and to access, import, share, view, and cite other published and grey literature are also necessary to facilitate creation of new content.

Other Benefits

IIa anticipates that the project will result in a P2P system that is easy to use, configure, and administrate. The approaches developed and implemented for this project are important for government agencies and for corporations providing public information to various consumers on widely accessible terms, as well as through more restricted access. Any research institution or information provider that makes a body of STI available will greatly benefit from this project. This communication mechanism will enable members of scientific communities to discover others with common interests, facilitating a broader perspective on specific topics, and increasing and enhancing the amount and quality of information available for a specific community of interest.

This system will also accelerate knowledge discovery. Having the ability to communicate instantly with peers and to identify a broader field of members within communities of interest will facilitate discovery and synthesis of STI. Furthermore, enabling the development of new scientific grey literature in standard bibliographic/metadata formats will facilitate retrieval by librarians and other information professionals. Not only is this concept useful to traditional government research agencies, but it is also applicable to universities, pharmaceutical companies, and all organizations involved in scientific research.

PHASE I

Experience, Goals and Accomplishments

Based on our analysis of the requirements during Phase I, we identified appropriate DOE libraries and other agency libraries to participate in the prototype. We conducted joint application design (JAD) sessions to determine the type of P2P communications desirable in this environment. We used a structured questionnaire to poll representatives from selected agencies and national laboratories, as well as local information analysts, to determine how they currently collaborate and how they would like to collaborate in the future. IIa developed and refined requirements through an iterative effort with DOE and other agency stakeholders.

Phase I also included a comprehensive survey of available commercial technologies and standards to determine what commercial-off-the-shelf (COTS) or shareware products provide P2P communications to satisfy some or all requirements. Technologies in development at universities, within DOE, and within other government agencies were also surveyed. IIa anticipates that available technologies will provide 60 percent of the solution.

User Input Process - Interviews, Polls, etc. A two-pronged approach was used to determine baseline requirements.

1. We interviewed individuals in a small, hand-selected focus group in person to define a useful set of collaboration tools. In one of the focus group interviews, we asked a DOE nuclear researcher his opinion of new social media. A few of his quotes follow:
    - "If there are ways to have a social networking site devoted to narrow topics, that might be of value. So, making that easy for someone to establish something like that. But it would have to be bottom-up, researcher-driven."0
    - "I'm working now in the area of thermodynamic modeling of nuclear fuel. I have some colleagues in that area. I hear from them occasionally. It would be nice to foster a little more communication on that."

- "There is a potential that we're overlapping in work, and we want to avoid that. So-and-so has developed an understanding of this chemical system. I don't want to do the same thing. I want to leverage what he's already done, and vice versa."

2. We identified a larger group to answer a questionnaire and provide feedback on a regular basis. To identify this group, emails were sent to more than 11,000 DOE researchers and authors explaining the project and asking for their assistance. Based on their responses, 136 respondents were selected to participate in the feedback.

Summary

There were three objectives in Phase I: (1) determine the high-level requirements for the system, (2), conduct a survey of COTS/open source systems, and (3) produce a prototype for evaluation. These goals were met, with results as discussed below.

Requirements

Several critical themes were almost universal among the respondents. Simplicity and ease of use were the first requirements identified by interviewees. In virtually all instances, the researchers said they did not have the time or the inclination to wade through a system of complex screens and infrastructure to try and use the system. They wanted to keep the focus on their research and publication rather than become the master of a complex tool. Another common theme was the need to efficiently access, import, and share content (especially from library collections), in-process work, lab notes, and other unpublished information. While this was not a focus in the initial phases of the project, it quickly became apparent that this system could be an extremely useful tool for the GreyNet community by providing the platform to make collections of grey literature available to select communities of interest using the uploaded information and other databases as a basis for discussion and collaboration. Likewise, forum members can upload their own collections of grey literature to use as a basis for discussion in their forum.

Sensitivity to the conflict "between collaboration and competition inherent to science"[2] was another theme. Interviewees want to be confident that adequate security safeguards are in place to ensure that information is shared only by trusted individuals who are members of their forum.

Market Survey

When the initial marketplace review was conducted early in Phase I, very few scientific collaboration products were available. Almost all of products were narrowly focused on specific scientific communities. During Phase I, extensive web research was conducted, and university and national laboratory work in this area was examined. While some products were partially related, in the end, no technologies were uncovered to fully solve the problem.

Prototype

The other primary purpose of the research was to develop a limited prototype with some of the planned functions for evaluation by DOE. The prototype was to demonstrate the concept's usefulness and desirability. The prototype was constructed and shared with DOE, who concurred that the concept was viable and would be a valuable addition to existing OSTI STI web products.

Our Phase I final report states: "The project team designed, prototyped, and integrated a web-based application feature called 'Author and Subject Clustering' into the OSTI Information Bridge web site. Author and subject clustering allows researchers to view a list of prominent authors and subjects for each topical search they perform."[3]

PHASE II

Prototype Development

For the enhanced prototype in Phase II, we developed and refined requirements through an iterative process among DOE stakeholders and IIa. Therefore, the initial requirements were expanded as needed by analyzing, documenting, and reaching consensus on requirements through prototype functions in the following areas:

- User interface
- Model to identify common interest groups
- Cost-effective, real-time, interactive P2P communication
- Capability to access, import, create, search various documents/content
- Privacy
- Security
- Data protection
- Reporting

The main focus of Phase I was to develop a questionnaire for use throughout the user community to determine the system's operational requirements. Based on this feedback, a system was designed to enhance collaboration within the scientific community, and a limited, prototype was developed to highlight authors within DOE online repositories.

For Phase II, we revisited a survey of currently available commercial technologies, including COTS or shareware products with P2P capabilities that may be adapted to the digital library environment. We expected and confirmed that there was considerable development in P2P collaboration between Phase I and Phase II surveys. In fact, some commercial products were released during this time. However, they were either so complex that target scientists were unlikely to adopt them, or they did not include important functionality needed to ensure effective performance for this audience. We did find that a significant number of shareware or open source products were highly customizable out of the box to fulfill many project requirements.

Before selecting a technology for the enhanced prototype, we identified barriers to adoption for the scientific community. Along with thorough evaluation of comments from the Phase I interviews and questionnaires, we also analyzed studies, literature, and online discussions of scientists. A recurring theme is summarized in this quote: "Most scientists were reluctant to invest more than a very small amount of time to learn to use new technologies unless the benefits were substantial and related directly to their research." This observation is consistent with earlier conclusions, underlining the importance of creating an intuitive, simple prototype with valuable tools and functionality. It also reinforces the conclusion that the flexibility of today's technology—doing everything for everybody—results in complex tools that are difficult for many scientists to use.[4] This exercise resulted in a list of obstacles that the enhanced prototype would need to overcome to ensure success (see Appendix 1).

Technology

During the Phase I technology survey and review of the existing technologies, we determined that there was no COTS product to meet project needs and that a prototype was needed. Therefore, we had to determine whether to develop something from the ground up, or whether we should select an open source framework to offer some initial desired functionality and to provide a development framework so that we could quickly add and evaluate new features. After a thorough evaluation of existing technologies, we focused on an open source content management system (CMS) to provide the most functionality out of the box while also providing a development environment in which features can be added as needed. We considered Ning, Drupal, Mambo, Joomla!, Taverna, and SupportMaster. Drupal was chosen based on several factors. Drupal is open source software maintained and developed by a community of hundreds of thousands of users and developers. It is distributed under the terms of the GNU General Public License (GPL), so it is free to download and share, and users can also contribute back to the project. This open development model ensures that people are constantly working to make sure Drupal is a cutting-edge platform that supports the latest technologies that the web offers. In addition, OSTI had selected Drupal for several applications (e.g., OSTIblog), so local expertise is available on the product. The Drupal CMS was chosen to allow for rapid prototyping of basic functions. Once it was determined that the tool met expectations for rapid prototyping, modifications and add-ons were made to make Drupal accommodate unique functional requirements and address performance issues. Drupal also paid big development dividends since there are literally thousands of contributed modules available to help solve unique

requirements. However, because Drupal is so rich and flexible, performance is impacted. As established earlier, poor performance is not likely to be tolerated by busy scientists. However, the intent was to use Drupal as the prototype, rewriting the system if the concept is commercially viable.

Because a major part of the system's functionality is to serve as a repository (library) for scientific research information, we determined that it was important to use open access standards for importing documents in the library function. We analyzed document metadata using digital object identifier (DOI), ISBN, Dublin Core, (along with OSTI DOE documents), and other standards for interoperability. The Library of Congress (LOC) website was used as a model for document interchange and standards. The result of the research and development is a web-based CMS that enables self-forming groups where researchers can create useful websites around scientific topics of interest to those groups. The project team named the website and system "Science-Forums.net."

Beta System Functionality

The purpose of the "Profiles" function is to encourage and support interaction among users of Science-Forums.net and ultimately of OSTI web portals. Profile information enables users to accomplish the following:

1. ***Verify the identity of other users.*** This is especially valuable when users have similar last names. For example, when a user's affiliation is provided in a profile, it can help end-users determine which "Smith" is the author of a particular document.
2. ***Specify other names under which authors have published.*** When an author has published under more than one name, the ability to make other names known can help users identify other works by the same author.
3. ***Provide more detailed background.*** Profiles can include information on education or work history to let other users know more about a particular user's credentials or experience with a subject. This is often of interest to researchers performing topic or author searches, viewing answers to questions, and understanding the context of online discussions.
4. ***Declare research areas and fields of interest.*** Profiles with this information can facilitate connections and dialog on topics of mutual interest.
5. ***Broadcast a user's interest in collaborative opportunities***. This information can help users who are seeking or offering partners for proposals.
6. ***Identify and locate researchers with similar interests***. This occurs in social networking websites and can be emulated here.

Trust is a critical element for effective online interaction. Without trusted communication, collaboration cannot take place. In order to achieve the project vision of researchers meeting, talking, and collaborating online, it is essential to provide a reliable level of trust for those interactions. This can be achieved through identification and verification of forum participants. The use of profiles is widely accepted and is scalable for different purposes. Profiles can be used on networked systems to identify and nominally authenticate users.

During the design phase, we considered providing a trust service that authenticates new participants by verifying some of the user-provided background information. We determined that this process would be too resource-intensive for the prototype, so the trust service will be explored by stakeholders for inclusion in subsequent increments.

Other trust services that Science-Forums.net profiles can provide include introductions and messaging between previously unacquainted users while protecting user privacy.

Researchers/users that express interest in providing feedback will be the first invited to create profiles to test the forum system.[5] Authors with articles in OSTI collections may also be invited to create forums and profiles. When the Science-Forums.net framework is integrated with OSTI web portals, a profile link will be available for authors who have provided profiles.

We considered a broad range of options for the type of information to include in user profiles, ranging from a minimal amount to a very robust profile like that used for LinkedIn or Facebook. Based on feedback during requirements analysis, a baseline model was chosen at this stage, with added features to address

unique requirements of scientists and professionals. After the user creates a user profile and initially logs onto the system, then he or she is presented with a welcome screen, which provides the user with access to the Forums screen.

At this point the user can search or browse the existing forums or create his or her forum. There are two types of forums that can be created - public and private. In a public forum, any user can join, search, view and contribute. In a private forum, membership is controlled by the forum creator and can only be viewed and contributed to by invited members. Once a forum is created and selected, the user is presented a screen in which the navigation bar can now change to display the following headings.

- **Library:** the main repository for existing technical content. The user can add content directly from DOE holdings, search and add content based on ISBN or DOI, cut/paste from existing metadata (e.g., BibTex or EndNote), and manually enter metadata for user-created documents. If they are available, the user can upload the full-text documents associated with the metadata in common document formats such as pdf, xls, doc, jpeg, etc.
- **Posts/Discussions:** blog posts, announcements, Q&A, and other communications for review, feedback and discourse. Other members can comment on posted entries, and users can view threads of conversations on a topic and respond to one or more previous posts. References or links to related information can be included in the body of the comment to support any point of view.
- **Calendar**: allows forum members to add event entries to mark meeting dates, conferences, or other events of interest.
- **Members:** allows the user to view all forum members and to manage privileges. In addition, it allows the user to invite colleagues to join the forum.

The system also supports other features, such as notifications via RSS feeds, forum administration, user feedback, searches, and user support such as FAQs and contact information. In addition, a full user profile management subsystem is provided.

CURRENT STATUS

The system is nearing the end of beta testing and is under consideration by several scientific groups for adoption as their collaboration portal.

FUTURE DIRECTIONS

Depending on the level of system adoption and the availability of future funding, several enhancements have been identified and will be included in the system. In addition, if the level of performance does not meet users' needs, we will consider a customization that relies less on Drupal modules to provide better performance.

APPENDIX 1
Barriers to Adoption of Social Networking by Scientists, Businesses, Workers

Type of Barrier	Description / Excerpt from Source
Benefits of social networking are not obvious to researchers	**The major barrier to take-up of web 2.0 tools and services is lack of clarity – even among some frequent users – as to what the benefits might be.** The costs of adoption are not always trivial, and unless researchers receive active support and see clear and quick benefits, they tend to keep to the tools and services that they know and trust. Moreover, the rapid development and proliferation of web 2.0 services mean that it is hard to keep track of them, or assess their potential benefits.[6]
	Some applications promise a lot of value, but people may not see the immediate value.[7]
	Scientists are not really interested in social networking as an end in itself. They network to boost productivity.[8]
Must be easy to use, and take little time/effort to learn	While the technologies used to collect and analyze the data often must be cutting edge, specifically collaborative technologies are often best served by technologies that are simple, require very little learning, and are already easily accessible (e.g. wikis, and telephones). **Most scientists were reluctant to invest more than very small amount of time to learn to use new technologies unless the benefits were substantial and related directly to their research.**[9]
	Widespread adoption of web 2.0 services by researchers depends on their being intuitive and easy to use, and incremental in building on existing practices. Above all, they must offer clear advantages to users, and near zero adoption costs.[10]
Perceived lack of Quality and Trust (without traditional peer-review)	But a second major set of barriers revolve around perceptions of quality and trust. Both as producers and consumers of information, **researchers seek assurances of quality; and many of them are discouraged from making use of new forms of scholarly communications because they do not trust what has not been subject to formal peer review.** A significant minority of researchers believe that peer review in its current forms will become increasingly unsustainable over the next five years, and nearly half (47%) expect that it will be complemented by citation and usage statistics, and user ratings and comments. But at present they do not see such measures as an adequate substitute for peer review. **Trust is also a concern for researchers who are producing, rather than consuming, information; they are cautious about sharing results and findings in a medium which, as yet, has no standardised way to formally attribute authorship.**[11]
	But the major disincentive for many researchers may be lack of trust. Both as creators and consumers of content and services, researchers seek assurances of quality. Our study indicates that many researchers are discouraged from using new forms of scholarly communications because they do not trust what has not been subject to formal peer review. **These findings are consistent with other studies (e.g. Ware and Monkman 2008) which suggest that researchers seek assurances of quality above all through peer review, and that they do not see citation counts, usage statistics or reader ratings or other 'wisdom of the crowds' tools as providing an adequate substitute.**[12]
	David said that some of the barriers to adoption include the lack of reputation information,[13] lack of confidence in individual networks,[14] and concerns that personal data might be sold,[15]
	Users may also be suspicious of a site's commercial intentions.[16]
	Data used in networks need provenance and structure - some kind of hosting institution or scientific society verifying the credibility of data.[17]
	Sites with no checks on data run the risk of offering less valuable information.[18]
Waste of time	some researchers regard blogs, wikis and other novel forms of communication as a waste of time or even dangerous[19]

Type of Barrier	Description / Excerpt from Source
	Some non-users go further, and believe that novel forms of scholarly communications bring no benefits or are even a 'waste of time.' 'I'd rather spend the time thinking about what I'm going to do next rather than spend it telling others what I'm doing… I think it's definitely a younger person's thing.'[20]
Needs critical mass of users to be effective	These problems are exacerbated by the fragmentation of the user-base: **few services have yet achieved the critical mass needed to achieve the positive network effects that stimulate pervasive use by particular communities**. Researchers may well be right to defer a decision to take up a particular service until they are sure that large numbers of their colleagues have done so.[21]
	But there is some debate about whether many of the web 2.0 services for researchers – particularly social network services – provide sufficient added value to stimulate widespread adoption.[22]
	Moreover, the plurality of services results in fragmentation of the potential user base, which is especially problematic when benefits are closely related to number of users. Researchers may well defer a decision to adopt until they are sure that large numbers of their colleagues have done so. Thus the advantages for late movers may outweigh those for early adopters.[23]
	Social networking does not have value if you are just talking to yourself. (Ressequie, 2010)[24]
	no one site provides tools or features valuable enough to lure a majority of busy scientists[25]
	What can they offer that more established sites, such as Facebook, don't?[26]
Applications not integrated	David said that they've tried Yammer here and a wiki there, but when the applications are not combined in one product, it splits people's attention and the usage falls off. … Some don't know how to use applications together.[27]
Change may conflict with existing practices	Changes are usually incremental, for a number of reasons. **Introducing changes without input from users may conflict with existing practices and be rejected. Moreover, it takes time for new practices to develop around new features, especially if these radically challenge existing disciplinary patterns of use.** Providers therefore generally seek to align new features with existing services and patterns of usage. Thus, the fora on Nature Networks are very much in the style of conventional bulletin boards, with implicit social rules of behavior and a reasonably active moderator.[28]
	Adoption of web 2.0 tools and services, and of novel forms of scholarly communication associated with them, has reached only modest levels up to now. Use is both fragmented and uneven, and tends to support well-established practices.[29]
Institutional policies can discourage use	We also found some evidence that emerging institutional policies may act as a barrier: 'In our university we have a guideline on what may or may not be put onto the blog. I have to agree that something needs to be saved and I don't want people to say: we just discovered X.' [30]
Conflict between cooperation and competition	Some scientists may be wary of sharing too much, which probably hinders network adoption. There is tension between collaboration and competition inherent to science.[31]

References

[1] Coppock, Edrick, O'Dell, D. (2009), Interactive Peer-to-Peer Scientific Communication in the Digital Library Environment, DOE FY2008 SBIR, DE-FGO2-08ER85096, Final Report, Department of Energy, May 28, 2009.
[2] Gewin, Virginia. (2010, 2011), Collaboration: Social networking seeks critical mass. – In: Nature, and Naturejobs.com, Dec 15, 2010, http://www.nature.com/naturejobs/2010/101216/full/nj7326-993a.html.
[3] Coppock, Edrick, O'Dell, D. (2009), Interactive Peer-to-Peer Scientific Communication in the Digital Library Environment, DOE FY2008 SBIR, DE-FGO2-08ER85096, Final Report, Department of Energy, May 28, 2009.
[4] Lee, C. P. and M.J. Bietz. (2009), Barriers to the adoption of new collaboration technologies for scientists. Position paper for The Changing Face of Digital Science: New Practices in Scientific Collaborations workshop, the ACM Conference on Computer-Human Interaction (CHI) – In: matthewbietz.org, Apr 5, 2009, http://www.matthewbietz.org/blog/wp-content/uploads/chi2009-scientificcollaborationsposition.pdf.
[5] Coppock, Edrick, O'Dell, D. (2009), Interactive Peer-to-Peer Scientific Communication in the Digital Library Environment, DOE FY2008 SBIR, DE-FGO2-08ER85096, Final Report, Department of Energy, May 28, 2009.
[6] Proctor, R., R. Williams, and J. Stewart. (2010), If you build it, will they come? How researchers perceive and use web2.0. London: Research Information Network.
[7] Resseguie, David. (2010). Personal observations at EPSCoR monthly meeting on Exploiting the Use of Social Networking to Facilitate Collaboration in the Scientific Community, Nov 2, 2010.
[8] Gewin, Virginia. (2010) "Collaboration: Social networking seeks critical mass." – In: Nature, and http://www.nature.com/naturejobs/2010/101216/full/nj7326-993a.html. Naturejobs.com, Dec 15, 2010.
[9] Lee, C. P. and M.J. Bietz. (2009), Barriers to the adoption of new collaboration technologies for scientists. Position paper for The Changing Face of Digital Science: New Practices in Scientific Collaborations workshop, the ACM Conference on Computer-Human Interaction (CHI) – In: matthewbietz.org, Apr 5, 2009, http://www.matthewbietz.org/blog/wp-content/uploads/chi2009-scientificcollaborationsposition.pdf.
[10] Proctor, R., R. Williams, and J. Stewart. (2010), If you build it, will they come? How researchers perceive and use web2.0. London: Research Information Network.
[11] Ibid.
[12] Ware, Mark and Monkman, Mike. (2008), Peer review in scholarly journals: Perspective of the scholarly community – an international study. Publishing Research Consortium.
[13] Resseguie, David. (2010). Personal observations at EPSCoR monthly meeting on Exploiting the Use of Social Networking to Facilitate Collaboration in the Scientific Community, Nov 2, 2010.
[14] Gewin, Virginia. (2010, 2011), Collaboration: Social networking seeks critical mass. – In: Nature, and Naturejobs.com, Dec 15, 2010, http://www.nature.com/naturejobs/2010/101216/full/nj7326-993a.html.
[15] Ibid.
[16] Ibid.
[17] Ibid.
[18] Ibid.
[19] Proctor, R., R. Williams, and J. Stewart. (2010), If you build it, will they come? How researchers perceive and use web2.0. London: Research Information Network.
[20] Ibid.
[21] Ibid.
[22] Ibid.
[23] Ibid.
[24] Resseguie, David. (2010). Personal observations at EPSCoR monthly meeting on Exploiting the Use of Social Networking to Facilitate Collaboration in the Scientific Community, Nov 2, 2010.
[25] Gewin, Virginia. (2010, 2011), Collaboration: Social networking seeks critical mass. – In: Nature, and Naturejobs.com, Dec 15, 2010, http://www.nature.com/naturejobs/2010/101216/full/nj7326-993a.html.
[26] Ibid.
[27] Resseguie, David. (2010). Personal observations at EPSCoR monthly meeting on Exploiting the Use of Social Networking to Facilitate Collaboration in the Scientific Community, Nov 2, 2010.
[28] Proctor, R., R. Williams, and J. Stewart. (2010), If you build it, will they come? How researchers perceive and use web2.0. London: Research Information Network.
[29] Ibid.
[30] Ibid.
[31] Gewin, Virginia. (2010, 2011), Collaboration: Social networking seeks critical mass. – In: Nature, and Naturejobs.com, Dec 15, 2010, http://www.nature.com/naturejobs/2010/101216/full/nj7326-993a.html.

Social Networking: Product or Process and What Shade of Grey?

Julia Gelfand, University of California, Irvine Libraries
Anthony Lin, Irvine Valley College. United States

Abstract
Social networking debuted in 1997 and is now an established and common method of communication and is increasingly related to and supportive of academic publishing, scholarship and generating new information. Some of the most mature and popular sites are Facebook, Bebo, Twitter, Linked-In and Plaxo plus many more specialized examples. As many professional societies and individuals choose to develop a presence on social networking sites (SNSs), the utility of them has become more valuable and ubiquitous. With emerging forms of technology to provide personal space and networking opportunities, the sites gain a new presence blending information products with new means to be discovered and searched. This paper explores how academic libraries are implementing a range of social networking activities to augment their online presence through traditional websites and launch new ways for their users to find, discover, access, navigate, evaluate, create and apply content. Libraries are also extending access with greater mobile optimized sites for Smartphones (iPhone, Droid, and Blackberry) and other emerging technologies that enhance information in utilizing and managing data, images, audio and streaming media. Today, library collections and services are being transformed due to electronic publishing, data curation, preservation and archiving efforts in order to allow users dependable 24/7 access to resources with potential for interactive communication. In an increasingly global and mobile society, the new social communication mediums reduce boundaries, transcend the digital divide and invite more transparency and participation by shadowing the grey parameters in the scholarly community and redefining publishing outputs and opportunities. This evolution did not happen without cautious regard to intellectual property, privacy and confidentiality. The new by-products are not always tangible but remain instructive and promote innovation within communication as a new adopted and accepted form of authorship and creative expression. By adding the social communication features of this kind of networking that promotes introduction, networking, commentary, critique, discourse, sharing, and the building of new communities, grey content is enhanced and the sources of access multiply with an open and entrepreneurial future for SNSs.

Let's begin with a story, not a fairytale but a true fact-based memory.

A not quite long time ago in a far away land, there existed a feudal system of Publishers who sought to control the largest number of Users as their loyal subjects (archeologists would later refer to this prehistoric era as Web 1.0). Users were happy for a while until one day they discovered that they too, could become Publishers, quite easily in fact, and form their own kingdoms and communities.

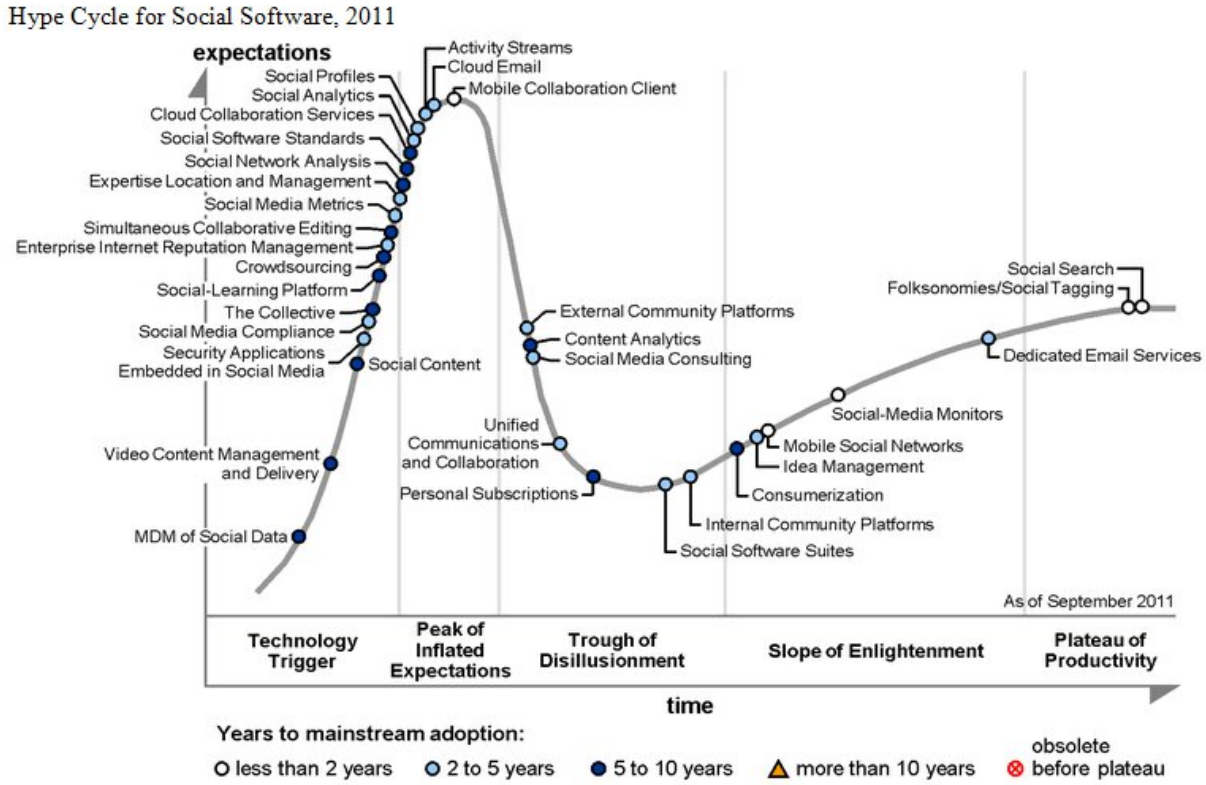
They created a new land were content is democratized and where each User could become their own king. New social websites and services soon enabled them to publish royal decrees (called Blogs and Podcasts), vote for content they Digg, discover old friends on Social Networks, Poke (and even SuperPoke) each other, become Internet Famous, gain Followers, and in a few cases become even larger and more powerful than the great Publisher kingdoms of yore. They named this new utopia Web 2.0 and immediately started to blog about how cliché the name was. Welcome to the Social Web.[1]

Background

The relationship between Grey Literature and Social Media is not at all unusual. Grey Literature has been known to include the previously challenging documentation, records, and communications in all formats including oral histories, news and broadcasts, correspondence, diaries, journals, reviews, other primary

source content and more recently websites, blogs, examples of photojournalism, videos, images, and other forms of commentary. The enhanced technologies which are now more ubiquitous in everyday life make for an intensive lifestyle change that allows individuals a way to present themselves, build and conduct independent relationships and offers personal empowerment that has never been so intensely realized prior to the development of the Internet. It often provides same or real time experiences and broadcasts across all geographical barriers as long as one has connectivity and anyone of many devices required to participate. Here we are at the end of 2011,

- Nearly 20 years after the birth of the World Wide Web
- 13 years after the launch of Google Search
- 8 years after the start of the first social networking site
- 6 years after the first YouTube video
- 4 years after the introduction of the first touch screen smartphone
- 3 years after the opening of the first "app" store
- And a little over a year after the first iPad sale[2]



Source: Gartner, September, 2011(captured October 30, 2011)

Social media has been defined by Gartner as:

> "a set of technologies targeted at forming and enabling a community of participants to productively collaborate. The ability of social-media technology to enable mass collaboration differentiates it from other collaboration technology and practices. IT tools to support collaboration have existed for decades. But social-media technologies, such as social networking, wikis and blogs, enable collaboration on a much grander scale. They enable potentially hundreds of thousands, even millions of people to simultaneously create content, share experiences and build relationships."[3]

These contemporary methods of social media are indeed different from other forms of exchange in that they can be saved, traced, shared, easily accessed and have unspecified properties challenging authorship, rights, permissions, and intellectual property. As much as they are technologies, they are tools. Suffice it to say, social media has redefined the information landscape. Print advertising revenue streams are drying up quickly due to readership preferences and patterns. The upswing of e-publishing and digital formats offers the added power of interactivity, commentary and opportunities for readers' feedback. Anyone can be an author or contributor and share their opinion with whatever bias with the entire world if they choose. The Global Web Index reported three clear trends in the consumer adoption of the Internet (not a medium):

1. Social Media has reached mass maturity – shift from massive growth and active social consumers to 'real-time' technologies – moving from creating content and publishing to sharing or from creator to distributor models for both 'traditional' media and professional content- utilizes tweets, status updates, etc
2. Open browser-based web is losing out to packaged internet platforms such as mobile apps, internet connected televisions, tablets, eReaders, gaming and video platforms
3. Continued aggressive growth of packaged platforms as traditional style content become a core part of the consumer online experience[4]

It is no surprise that as we witness the changes in how media is delivered, cable television, satellite, radio and newspapers as we currently know them are in danger of becoming obsolete. A more "packaged" Internet may transform the way we get online, the content we consume, and the ways we can create, share and communicate. [5] We also see a convergence taking place within new social media as Twitter becomes increasingly visible and important and the academic community abandons or relies less upon RSS feeds for updates than on twitter feeds, choosing to find twitterers who keep current, participate in timely conversations/debates and provide links to sources relevant to personal interests, both individuals and relevant organizations. These methods of keeping up with technology will be personally challenging as well as a goal for the workplace to handle.[6]

Social Media as Grey Literature

Grey literature has always encompassed new ideas and captured information outputs regardless of formats and as technologies have emerged, embraced them to the extent that they can be saved, preserved and archived. In the last twenty years of the digital revolution, we have been immersed in the knowledge generation era followed by one in knowledge management, tapping into the vast potential that hypermedia and the Internet allows. Higher education and research benefited from this in a very transformational way as the structure of social construction expanded to include new communication methods. Previously, a world of standard grammar peppered with occasional colloquialisms and nuances, new lingo, and expressions would be common. Today, new languages to accommodate social media have been birthed, text messaging, for one has its own dictionaries and a lingua franca of the hand-held device has determined ways to create its own ethnography. In one of the first books that chronicled this movement of Social Media the compiled conference papers allowed the editor, Edward Barrett to coin the term, 'sociomedia,' "to signify when we design computer media we are hardwiring a mechanism for the social construction of knowledge."[7] The papers in this early collection (1989) explore the role of the computer and what is information and he writes, "Information about a text becomes as important as the primary text itself. "[8] An aspect of daily life that has really assumed added visibility due to social media is the expansion of religious identity via the new communities established through these online venues. (Church of Facebook) As the *Church of Facebook* describes itself, the hyperconnected are a redefining community and encouraging the pursuit of authentic relationships with spiritual beings outside of the secular domain.[9]

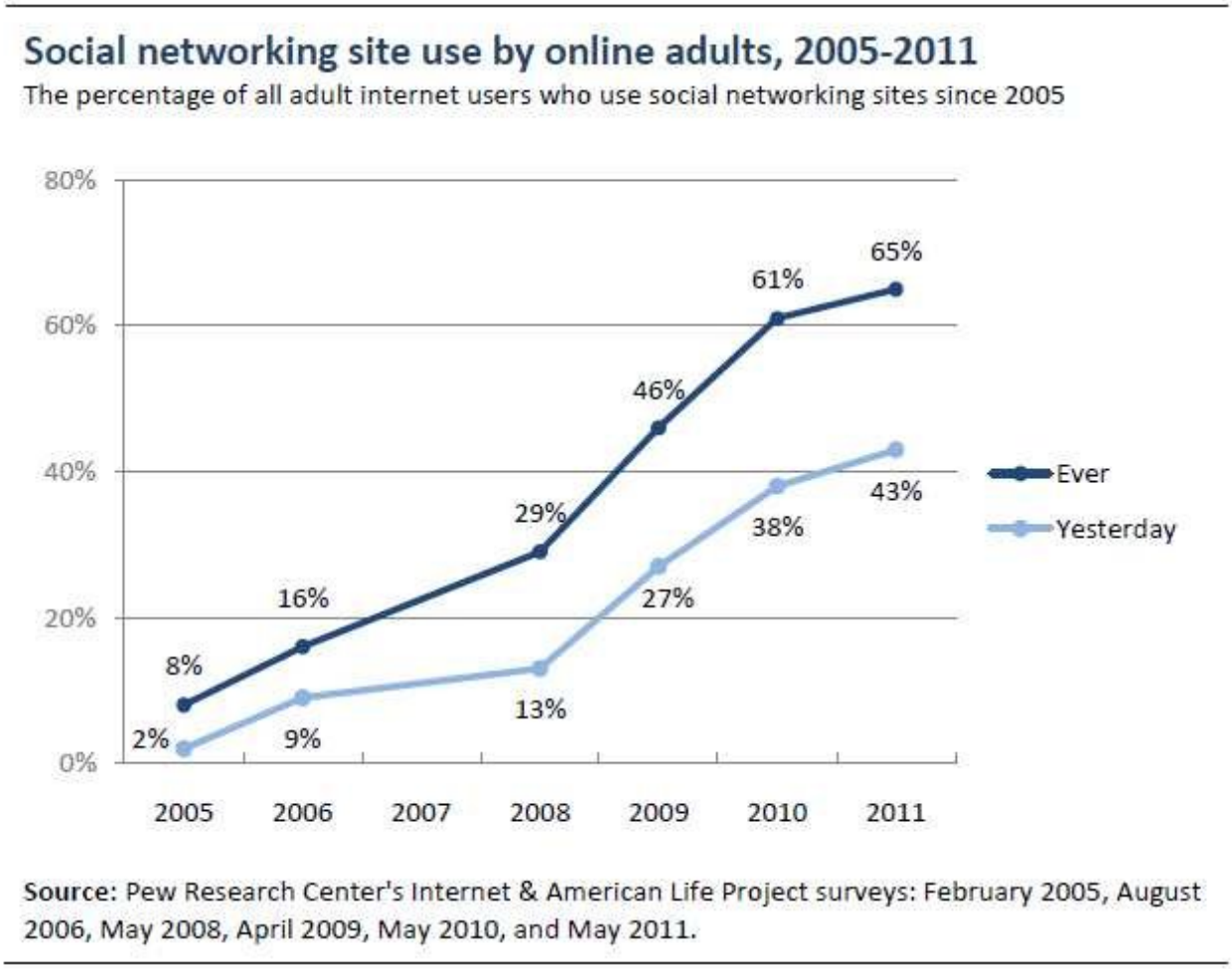
No subject, discipline, activity, workflow is left untouched by social media and its potential. The business community's thread to survival is keenly dependent on its mastery, creative and appropriate applications. It is considered by many current writers and historians of the day, that the omission of social media will delete you and your enterprise from the public radar. Still considered in its infancy, social media remains to be all about potential and users will have to be flexible and able to cope with frequent change, and a certain amount of ambiguity.

While we were finishing this paper, the news of the passing of two extraordinary and highly influential people who contributed to this topic of information sharing communities and networks cannot be overlooked. Steve Jobs, founder of Apple and John McCarthy, known as the father of artificial intelligence. Without their innovative spirits, our today would be a lot different.

As librarians, our generation is focused on knowledge creation and we commend David Lankes for including a "thread" on this in his much acclaimed, *Atlas of Librarianship*. He quotes Melvil Dewey, "The time was when a library was very much like a museum, and a librarian was a mouser in musty books…The time is when a library is a school, and the librarian is the highest sense a teacher."[10] The Atlas is intended to have a worldview of librarianship not founded on materials but outcomes and learning with concentration on the librarian rather than the institution. [11] The Mission Statement, "The Mission of Librarians is to Improve Society through Facilitation Knowledge Creation in their Communities," is a figurative illustration about how that can be done with different focus points and the one selected for this paper, emphasizes conversation and should be viewed vertically or top to bottom.[12]

How we perceive social media networking

Increasingly, people are using social media for a variety of different things. According to the Pew Internet Research Foundation, more and more users are using social media as a daily part of their lives[13]. As can be seen from the chart below, social networking site use by adults starts from a low 8% in 2005 to nearly 65% in 2011.[14]

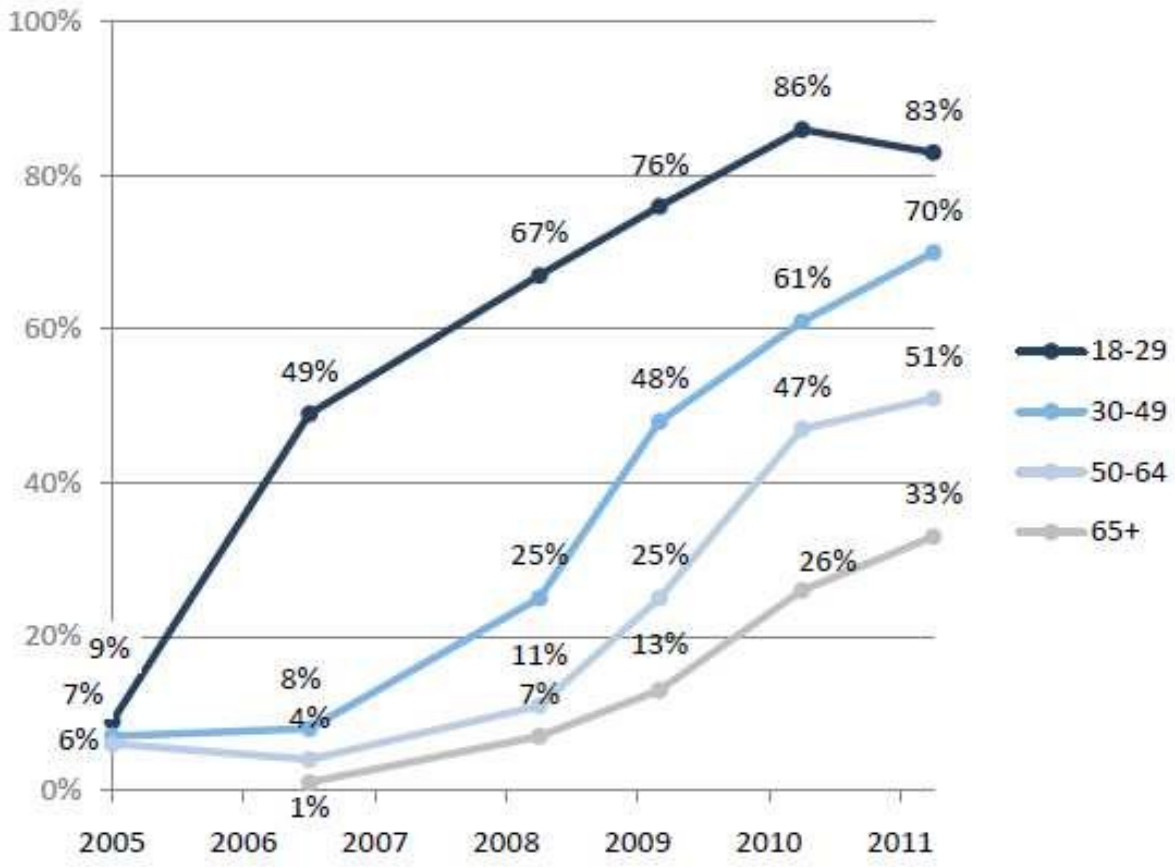


Harnessing Social Capital

Most importantly, the age of social networking users tends to be in the 18-29 age category compared with older adults, although that is changing.[15]

Social networking site use by age group, 2005-2011

The percentage of adult internet users in each age group who use social networking sites



Note: Total n for internet users age 65+ in 2005 was < 100, and so results for that group are not included.

Source: Pew Research Center's Internet & American Life Project surveys: February 2005, August 2006, May 2008, April 2009, May 2010, and May 2011.

As the number of people who use social networking and social media proliferates, libraries and librarians will find themselves in positions as curators of information to harness the power of social media for academic and scholarly research and reflect the demographic analysis. In Charnigo and Barnett-Ellis' 2007 survey of 126 academic librarians, they estimated 54 percent of the librarians surveyed thought there was little academic use for Facebook.[16] On the other hand, in a study done approximately at the same time by the Pennsylvania State University Libraries 126 of the 441 reference questions in a sampled Fall Semester study were asked through Facebook. Compared to the other methods of reference questions such as email, phone, IM, and In-person reference, the Facebook transactions were the most highly asked for reference questions.[17]

Since social media is still a relatively new concept, some libraries and librarians are resistant to the idea that social networking can have academic purpose (Dickson and Holley, 2010).[18] It is assumed that the usage of this method of communication is still evolving and will continue to evolve many years to come. This paper and our research examine how libraries use social media as a product or process to achieve the goal of improving academic scholarship in different ways. In terms of academic usage, some universities, such as how George Washington University concluded in a study in 2007 that only four percent of the students that they interviewed used Facebook for academic purposes[19] and we can speculate that this figure has probably increased over the last four years, but by how much remains unclear. As more

campuses adopt distance education to reach out to more students, course management systems may be used to facilitate information sharing and online participation.
Since Facebook has let users design programs, attach them to their own pages and publish them for other users, examples ranging from sports to entertainment, science, medicine and the arts allow users to satisfy many of their online needs without leaving their profile pages. [20]

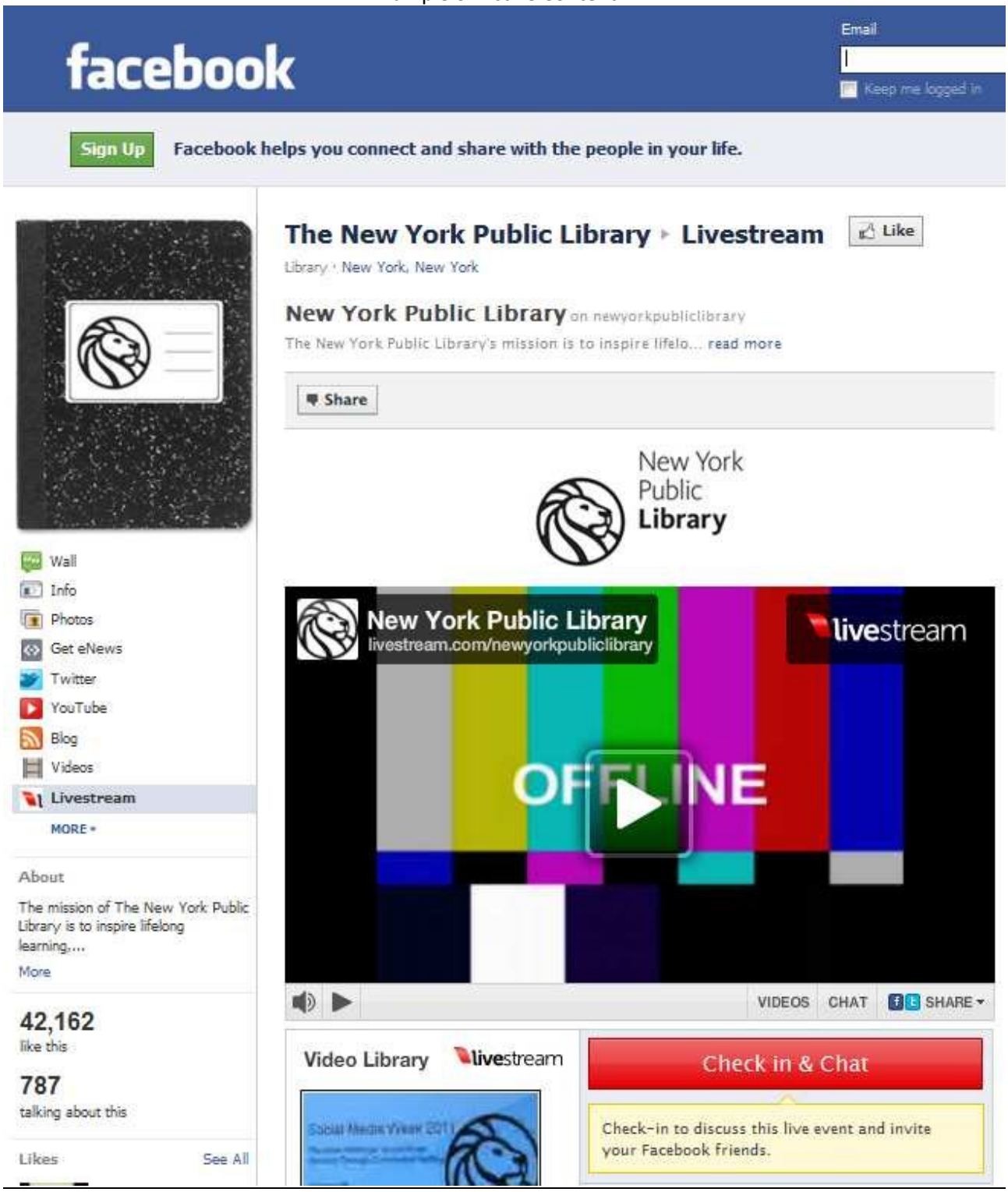
Library Applications – Passive/Active social media content

One of the goals of libraries branching into social media sites such as Facebook, Twitter, and Google+ is the ability for the library to conduct outreach to more users while they are not inside the library. The type of information that exists on the library social media site can be divided into two main categories: static information and active social media content. Static information is a one-way information exchange that tells the user a piece of information such as the library hours, location, and website information. Static information can exist on Facebook as a complement to the library's traditional website .[21] An example of static content follows:[22]

Source: Library of Congress Facebook Page (http://www.facebook.com/libraryofcongress/)

Active content is a two way dialog between the library and its users. This means anything that the library does through the library social media site such as online reference services to include chat reference through Facebook or, posting a Twitter feed of a library event. Other types of active content could also include the Facebook "wall" with dialogs, postings and comments from library patrons, users, and also other Facebook members containing reviews of content, exhibitions, streaming meetings such as our gathering today among other examples.

Example of Active Content:[23]



Source: New York Public Library Facebook Page – Livestream (http://www.facebook.com/newyorkpubliclibrary)

Today's news that The Library of Congress is also engaged in creating and maintaining a Twitter Archive of all tweets sent suggests new social and historical value of this unique communication medium.[24]

The Educause Center for Applied Research noted that the Library tops social networking in a tally of students' computer and internet activities in 2010. The list ranked 19 entries:[25]

Activity	Share of Students Using	Median use frequency
Using College Library's website	94%	Weekly
Presentation software	93%	Monthly
Text messages	90%	Daily
Social networking websites	90%	Daily
Course or learning management systems	90%	Several times per week
Spreadsheets	86%	Monthly
Instant Messaging	71%	Several times per week
Graphics Software	67%	Monthly
Using Internet from handheld device	50%	\Daily
Voice over Internet (VOIP) from computer	47%	Monthly
Following or updating microblogs	43%	Several times per week
Contributing content to video websites	42%	Monthly
Contributing content to wikis	40%	Monthly
Video creation software	40%	Monthly
Contributing content to blogs	36%	Once per quarter/semester
Audio creation software	34%	Monthly
Online multiuser computer games	27%	Once per quarter/semester
Social bookmarking/tagging	25%	Weekly
Online virtual worlds	9%	Once per quarter/semester

Somewhat related, the same source reported that text messaging is the communication form of choice for most students compared to social networking sites, instant messaging and VOIP.[26] Ease of use via mobile phones, usually readily at hand suggests the probable reasons.

Organizing information appears to be central to the construct of social media and to the agenda of most libraries. One interesting sidebar may be the contributions of another computer scientist, David Gelernter, also known as a victim of the Unibomber among other notorieties. He and his students, in this context, are the chief architects of Lifestreams, defined as:

> "a time-ordered stream of documents that functions as a diary of your electronic life; every document you create and every document other people send you is stored in your lifestream. The tail of your stream contains documents from the past (starting with your electronic birth certificate). Moving away from the tail and toward the present, your stream contains more recent documents --- papers in progress or new electronic mail; other documents (pictures, correspondence, bills, movies, voice mail, software) are stored in between. Moving beyond the present and into the future, the stream contains documents you *will* need: reminders, calendar items, to-do lists."[27]

This is one of the first visible chronicles of everyday communication, exchanges and activities. Analogous to the capstone of autobiographical journal experiences, this record of sharing put document management or organization to a test for indexing, labeling or tagging.

With the subtheme of this conference, "From Social Networking to Wealth Creation," one has to consider what the different intelligence communities are. Today, the corporate sector has been described as "Data Rich, Insight Poor." This oneliner from Casper Craven, founder of the British research firm, Trovis, states,

> "the research warns that companies ignoring social media channels do so at their peril."[28]

This is clearly a demonstration of the new competitive intelligence the corporate sector is relying upon – who can utilize channeling to follow leads, connect to new customers and establish new markets. As higher education continues to emulate the business world, instruction follows business trends.

In a different domain, but one that is equally conducive to fast changes refers to the youngest population and one that the public, school and academic library community is monitoring. Children growing up today are unaware that telephones ever had cords or had to be used in a static environment. Additional studies by the Pew Foundation, Outsell, the Horizon Reports, and countless other releases from global marketing and research firms are plentiful but the report, "Zero to Eight: Children's Media Use in America" released last month by Common Sense Media is the first examination of how "even very young children are frequent digital media users, experiencing mobile media, using computers and playing video games"[29] in early childhood defined as ages 2-4 and 5-8. In addition, "there continues to be a substantial digital divide, including both computers and mobile devices"[30] referring to family income and the ability to invest in home computers, smart phones and promise connectivity. Other findings note:

- Children under 2 spend twice as much time watching TV and videos as they do reading books
- TV continues to dominate children's media use
- Broadcast television is the most accessible and widely used platform for educational content among lower-income children
- Media use varies significantly by race and socio-economic status, but not much by gender
- Even some young children are media multi-taskers[31]

Evidence of media multitasking is observed as young as among 5-8 year olds with such examples as doing homework while watching television, speaking on the telephone and monitoring Facebook, but increasingly common among 12-18 year olds. Media as babysitter is also common, while out running errands, or by parents who use media to occupy themselves when they are considered to be playing with or engaging with their children. Think of the last time you visited a park or supermarket and observed parents holding a phone, wearing Bluetooth, while disciplining a child! The major conclusions confirmed:

- that increasing dosages of baby videos on TV and DVDs on a regular basis with children under 2 are concerning
- bedroom television contributes to obesity, declining school grades and other negative effects
- educational media via television remains the most popular platform
- there is an enduring and substantial digital divide among lower income families, even when targets are for health and social promotion
- quality entertainment is challenged with negative impact of content and advertising of violence or inappropriate gender roles
- perceptions of a decline in reading and literacy may not be as great as earlier studies
- there is still much to learn about neuroscience effects of multitasking by very young children[32]

Perceptions of social media characterize information products that one used to consider perpetual ways of life that were deemed safe and were part of legacies as well as future generations. As reading habits changed, information became far more abundant, accessible, yet fragmented traditions of scholarship and lifestyle have been challenged. Some academic examples include the book review. When one published a scholarly monograph or even a trade book, novel, or released a film, had an exhibition, the anticipation of the reviews clouded the authors or artists' horizons for months as that could dictate the professional opportunities of one's future, the commercial viability of the project, compensation packages, new merchandising potential and any host of other related by-products.

Today, reviews are immediate, distributed worldwide and can commend the work hailing its success at the cash register or box office, or pan it with a few tepid or carefully selected words. In a recent op-ed piece, a historian shared his observations of why one should read book reviews by concluding that they are "enduring value of the form," and going on to state," Reviews do influence my book buying –just in a roundabout way. I'm sometimes inspired to buy books by authors whose review impress me." He concludes by saying that as a reviewer and reader of reviews, he can now see how a "pretty nifty Twitter-friendly version could have been built around the best lines from the [full review]."[33] New social reading

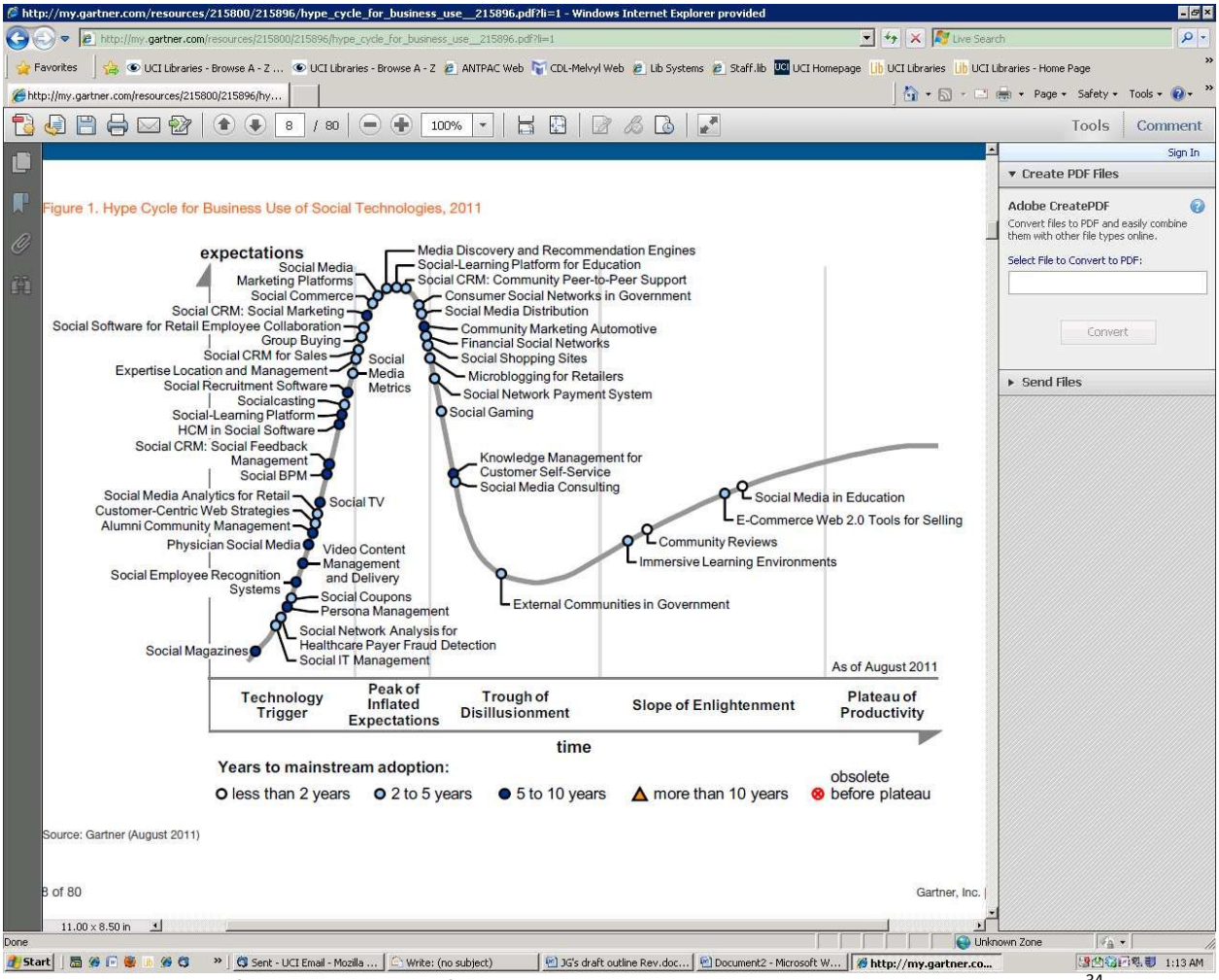
experiences are abundant and content ideas can be shared with people one may never encounter outside this virtual world such as in BookGlutton.com

Signs of addiction to social networking are apparent as one's identity is defined by where one is found. Earlier explorations into virtual worlds termed such visitors as geeks and techies, today, one is expected to be a joiner, self-promoter, au current with microblogging options that allow followers the 140 character tweets. Relationship building today often has its beginnings in the throes of social networking. Immediacy and constancy are what defines it as we think of the quality of the friends we acquire on these journeys.

New Communication Strategies

Forrester Researcher and Gartner are pioneers in tracking consumer behavior and indicating trends in media usage. Public opinion polling, voting patterns and behavior, campaigning, newscasts, GIS activity have each been a hallmark of grey literature. The political landscape offers a wealth of examples of grey literature that comes out of social media and could include the spectrum of that and other related matter. President Obama's 2008 campaign ignited the populace via methods of social media like nothing in the past and redefined the potential for this medium. Today's "Occupy Wall Street" again demonstrates how social media can mobilize populations with a common cause and agenda and predictably, have some staying power.

The following is an example of a Gartner Hype Cycle for Social Media in Business Technologies done in August 2011 demonstrating the adoption rate of business use of different social media:



Source: Hype Cycle for Business Use of Social Technologies, 2011. Gartner, September, 2011.[34]

A SWOT for this analysis of social media may include the quadrants: Challengers and Leaders on the top with Niche Players and Visionaries below.

And yet another very academic example is the most interesting corpus of work by the political scientist, George R.Boynton at the University of Iowa who studies and teaches about new media in political discourse. He tracks the spread of political sentiments and social issues by different forms of new media, tracing how fast news travels. His research "focuses on dissemination of information of "real time" events, such as Iran's nuclear production activity and President Obama's Nobel Peace Prize. By undertaking a systematic analysis of Twitter feeds and other social media, he sheds light on the shifting information landscape and explores emerging areas of research on the intersection between these new media and politics. Boynton publishes reports about his new media research with supporting data on his Web site, and, of course, posts updates on Twitter (@bobboynton)."[35] For this, he was recognized as the recipient of the the 2010 Primary Source Award for Research for a project that deploys eight computers to continually harvest from the Web data on new media trends.[36] For an example of the Twitter Streamgraphs used to track such activity visit http://www.crl.edu/sites/default/files/images/focus/Boynton_stream_graph.jpg [37]

Social Network Data

Returning to a theme of Grey Literature and data sets, the Discovey Research Group has culled some usage statistics from a variety of well-known social media platforms and shared in recent months and if anything this data has increased:

- Average Facebook user has 130 Friends
- Over 25Billion pieces of content are shared on Facebook each month
- Over 200 Million active users of Facebook via mobile phones; overall there are 500 Million users
- In late 2010, one year ago, Twitter had over 100 Million registered users with 300,000 signing up each day
- Twitter received 180 Million unique visitors to its site monthly and over 600 Million searches daily
- Twitter users are averaging 55 Million tweets daily; 637 tweets a second
- YouTube receives 2 Billion views a day and the 3rd most visited website
- On average, visitors spend 15 minutes daily viewing and adding content to site
- 24 hours of video being uploaded to YouTube each minute[38]

As a new form of grey literature, these data sets will add to the timely corpus of information about these trends and continue to inform us about how we function in this increasingly web, semantic and multimedia age. Detecting new influences, seeking behavioral indicators, and tracking the patterns and timing of these functions and how they contribute to new innovations, stimulate the economy, generate new knowledge is all for the greater good. Higher education is testing and trying many of the features of this medium by adopting social bookmarking practices in the classroom (Brainify.com) and utilizing social media to submit college admission essays.[39]

The Global Divergence of Social Networking depicts the growth in social media worldwide but especially in China, the US, India and Brazil, across all demographic segments. In the United States what we know is that:

- Social networking and photo sharing are the leading social channels
- Social networking and video sharing are the fastest growing platforms
- Product reviews, Q&A services and forums have the smallest differences between young and old consumers
- Despite the rise of Twitter still 50% more bloggers in the US
- Fragmentation by demographic components suggest that the target audiences and customers that use social media are not fully developed and requires special insights[40]
- Meanwhile, the number of U.S. unique users was 50.8% December 2010, a sharp drop from 62.1% the previous June. This suggests that the use of Twitter outside the U.S. has experienced significant growth over the past six months.[41]

Conclusions

Libraries, like business are trying to stay relevant and increase market shares of their constituencies. The competitive intelligence indicators suggest that engagement, branding, overcoming the digital divide and utilizing the available technologies in smart efficient ways leads to greater participation and collaboration,

key goals in the promotion of learning and exchange. Some disconnects or challenges exist for the academic community, as we are warned about the risks of creating poor content.

> "If your library has been making the common mistake of primarily using social media as a marketing platform for its collections and programs, then its content probably lacks interest or relevance to the majority of its followers. One of the things that makes content successful is making sure there is a "hook" to capture readers' attention. Interesting content will generally be humorous, useful or newsworthy. Take care to ensure that whatever is being posted is not only useful but also entertaining."[42]

Entertaining is often perceived as inappropriate for the academic marketplace, but being relevant may require that to appeal to the target age cohort and compete with other more entertaining content. It is clear that the latest World map of social networks popularity suggests use of social media by all ages.[43] We think it is possible, and that social networking will continue to be both product and process and dim in its hues of grey. Social Networks we speculate will continue to evolve both in the personal and external professional settings. Communication security will experience challenges but the typical features of having groups, fan pages, widgets, applications, games, profiles, communication abilities with degrees of privacy to allow for collaboration, recruitment, branding, engagement, retention, mentoring, leadership development, relationship building, promoting of diversity. With more than 90% of people in the workplace using eMail compared with fewer than 40% of workers who use social media tools, as found by a recent study commissioned by Skype, we predict a steady decline in eMail and a surge in use among a growing range of social media tools.[44] Users will become increasingly savvy and the unforgiving economy will lead by changes in business communication and thus education will follow. That is our scenario and grey literature will not be referred to that way because its ubiquity and increasing ease of use will continue to transform how life is conducted. When asked back in 2007 in *Friends: the Social Networking Sites for Engaged Library Services* whether social networking applications are a long term phenomenon, Stuart Weibel, Research Director at OCLC, emphatically stated, "they already ARE."[45]
Let's end by asking all those in this room who over the last week (due to travel) have contributed to, consulted or read a blog, to please rise; who collaborated on or contributed to a wiki to rise; who watched a podcast to rise; who updated personal information in fb to rise; who sent a tweet to rise; who sent an eMail message to please rise. Need we say more?

References

[1] Ford, Rob and Wiedemann, Julius, eds., (2010). *The Internet Case Study Book*. Cologne: Taschen, 232.
[2] Zero to Eight: Children's Media Use in America (2011). *Common Sense Media Research Study*, released October 25, 2011. Http://www.commonsensemedia.org/research/zero-eidht-childrens-media-use-america
[3] Bradley, Anthony J., (December 10, 2009). The six core principles of social media: Based collaborations. Gartner ID G00172930. Viewed October 29, 2011.
[4] Global Web Index (2011). *Annual Report*. http://globalwebindex.net/explore-the-data/data/annual-report-2011/
[5] *Ibid.*
[6] Threads on LITA-L listserv, December 1, 2011. Keeping up with technology.
[7] Barrett, Edward, ed. (1992). *Sociomedia: Multimedia, Hypermedia, and the Social Construction of Knowledge.* Cambridge, MA: MIT Press, 1.
[8] Ibid, 8.
[9] Rice, J. (2009). Church of facebook: How the hybperconnected are redefining community. Colorado Springs, CO: David C. Cook Publishers.
[10] Lankes, R. David (2011). *Atlas of Librarianship*. Cambridge, MA: MIT Press, 31.
[11] *Ibid*, 15
[12] *Ibid*, 30
[13] Madden, M. and Zickuhr, K. (2011, August 26). 65% of online adults use social networking sites. In *Pew Internet & American Life Project*. Retrieved October 25, 2011. http://pewinternet.org/Reports/2011/Social-NetworkingSites.aspx
[14] *Ibid.*
[15] *Ibid.*
[16] Charnigo, L.and Barnett-Ellis, P. (2007). Checking out Facebook.com: The impact of a digital trend on academic libraries. *Information Technology and Libraries*, 26 (1), 23-34.
[17] Mack, D., Behler, A., Roberts, B., and Rimland, E. (2007). Reaching students with Facebook: Data and best practices. *E-JASL: The Electronic Journal of Academic and Special Librarianship, 8(2).* Retrieved November 7, 2009.

[18] Cickson, A. and Holley, R.P. (2010). Social networking in academic libraries: The possibilities and the concerns. *New Library World*, 11 (11/12), 468-479. doi:10.1108/0307480101094840
[19] Bietila, D., Bloechi, C. and Edwards, E. (2009). *Proceedings from the ACRL Fourteenth National Conference*, 2009, Seattle.
[20] Baumann, M. (May 2008). Facebook Apps: Deeper Social Networking. *Information Today*, 25 (5), 1.
[21] *Dickson and Holley*.
[22] Library of Congress Facebook Page. http://www.facebook.com/libraryofcongress/
[23] New York Public Library Facebook Page via Livestream. http://www.facebook.com/newyorkpubliclibrary
[24] Library of Congress to receive entire Twitter Archive. http://www.federalnewsradio.com/?nid=239&sid=2658996 (December, 7, 2011)
[25] Educause Center for Applied Research (2010). *Chronicle of Higher Education*, August 26, 2011, 51.
[26] *Ibid*.
[27] Lifestreams: Organizing Your Life. http://cs-www.cs.yale.edu/homes/freeman/lifestreams.html
[28] Williams, Peter (October 29, 2011). Data rich, insight poor. *Information World Review*. http://www.lwr.co.uk/social-media/3011134/Data-righ-insight-poor
[29] *From Zero to Eight: Children's Media Use in America*, 9. http://www.commonsensemedia.org/sites/default/files/research/zerotoeightfinal2011.pdf
[30] *Ibid*, 10.
[31] *Ibid*, 11-13.
[32] *Ibid*, 29-30.
[33] Wasserstrom, J. (September 9, 2011). Why read book reviews? *The Chronicle Review*, B20.
[34] Gartner (August 2011). Hype Cycle for Business Use of Social Technologies, 2011: 8. Gartner, Inc.: G00215896. Viewed October 30, 2011.
[35] G.R. Boynton Website. http://www.boyntons.us/website/grb/grb-start-page.html Viewed October 28, 2011.
[36] Center for Research Libraries (CRL) (2010). Primary Source Awards, 2010: Robert G. Boynton. http://www.crl.edu/focus/article/6642 Viewed October 30, 2011.
[37] Twitter StreamGraph. http://www.crl.edu/sites/default/files/images/focus/Boynton_stream_graph.jpg Viewed October 30, 2011.
[38] Discovery Research Group (2011). Social Media Research and Listening. http://www.discoveryresearchgroup.com/social-media-research/ Viewed October 30, 2011.
[39] The New College-Admission Essay: Short as a Tweet. *Chicago Tribune*, December 20, 2011.
[40] Digital Stats, April 2010. http://digital-stats.blogspot.com Viewed October 30, 2011.
[41] Sysomos (2011). Exploring the Use of Twitter Around the World. http://www.sysomos.com/insidetwitter/geography/ Viewed October 30, 2011.
[42] Solomon, L. (2011). *Doing Social Media So It Matters: A Librarian's Guide*. Chicago, IL: ALA Editions, Special Reports, 53.
[43] Sambhi, H., (December 2009). World map of social networks, in *Social HR*. Toronto, ON: Carswell: 119.
[44] Sturdevant, C. (January 3, 2011). Socializing the enterprise. *eWeek*, 28 (1): 34.
45 http://onlinesocialnetworks.blogspot.com/2007/12/blog-post.html

Additional Selective References

Baird, T. (2009). *How to drive traffic to your website with Facebook: Growing your online business with Facebook, and much more*. La Vergne, TN: US Emereo.

Blossom, J. (2009). *Content nation: Surviving and thriving as social media changes our work, our lives and our future.* New York: Wiley.

Cassidy, E.D., Britsch, J., Griffin, G., Manolovitz, T., Shen, L, Turney, L. (2011). Higher education and Emerging Technologies: Student usage, preferences and lessons for library services. *Reference & User Services Quarterly* 50 (4): 380-391.

Gillette, F., June 27-July 3, 2011. The rise and inglorious fall of Myspace. *Bloomberg Businessweek*: 52-59.

Katz, J.E., LaBar, W. and Lynch, E. (2010). *Creativity and Technology: Social Media, Mobiles and Museums*. Edinburgh: MuseumsEtc.

Kear, K. (2011). *Online and social networking communities: A best practice guide for educators*. London: Routledge.

Micek, D. and Whitlock, W. (2008). *Twitter Revolution: How social media and mobile marketing is changing the way we do business and market online*. Las Vegas, NV: Xeno Press.

ShelfLife: Significant resources on social media. *Choice Reviews Online,* November, 2011. http://www.cro2.org/default.aspx?page=reviewdisplay&pids=3688423

Solis, B. (2011). *Engaged!: The complete guide for brands and businesses to build, cultivate and measure success in the new web.* Rev. and updated. New York: Wiley.

Sterne, J. (2010). *Social media metrics: How to measure and optimize your marketing investment*. New York: Wiley.

IEEE Computer Society, November 2010. Technology-mediated social participation. *Computer*, 43(11).

Weber, L. (2009). *Marketing to the social web: How digital customer communities build your business,* 2d ed.. New York: Wiley.

Knowledge Communities In Grey

Claudia Marzi
Institute of Computational Linguistics, National Research Council, Italy

Abstract
The dynamic nature of modern human social interactions, and the increasing capability of wireless and mobile devices for creating and sharing contents, open up the opportunity for a wide dissemination of information through complex knowledge sharing systems.
As the shared knowledge components build cognitive ties, there is no real sharing of knowledge without a common understanding of it.
In this article, particular emphasis is laid on technologies in Natural Language understanding and knowledge management for providing structured, intelligent access to the continuously evolving content, generated on-line in a pervasive collaborative environment.
In detail, robust automated techniques for term extraction and knowledge acquisition are used to tap the information density and the global coherence of text excerpts sampled from both general-purpose and subject-specific social networks. We show empirically that the two sources may exhibit considerable differences in terms of content accessibility and informativeness.
Topics: Subject based Communities; Social Networking
Keywords: Grey Literature, Web Communities, Knowledge sharing, Concept Maps

1. Introduction

The development of digital technologies and the continuous evolution of telecommunication networks are rapidly heading our society towards a culture of participation and to a more and more interactive communication. The dynamic nature of modern human social interactions, the adaptive networking protocols and data management systems are fostering pervasive information and communication environments.
The web represents an unlimited universe of information and data, and offers the steadily increasing availability of ubiquitous accessible information.
As accessibility improves, however, the huge amount of data and information available on the web need to be identified, classified, analyzed, filtered, so as to enhance the generation and assimilation of new knowledge.
Large volumes of information, even structured information, have to be managed, and generation and assimilation of knowledge have to be facilitated. Knowledge needs to be represented, standardized and distilled from multiple sources.
In this context, Social Networks can enhance fore-front ideas and highly innovative contents; they offer the enormous potential to transform research, and research results, into a knowledge co-creation process.

2. Methodology

2.1 Research questions

Given this scenario, the following questions arise naturally.
To what extent can Social Networks provide a real opportunity for sharing and disseminating novel information and generating knowledge?
Can they really be supportive of a steady flow of technical and scholar writing, or do they only provide a general communication channel for ephemeral communication exchanges?
Is there a specific added value in the way Social Networking can foster people's interest in sharing and building information?
Is interactive, informal and ubiquitous information exchange developing a new social framework for the creation of public-domain knowledge?
We suggest that all these questions can be addressed by applying advanced Natural Language Processing tools for automated content extraction to the analysis of web-based text collections, sampled from both general-purpose and specialized examples of social networks.

2.2 Research rationale

The Information Extraction literature provides different modes and tools for knowledge acquisition and representation: from highly structured, standardized and objective knowledge information systems based on ontological hierarchies and relations to more dynamic, subjective tools for volatile knowledge representation such as word clouds and concept maps.

Technologies in Natural Language understanding offer an objective measure of the information density of a text document or document collection and ways to map out the distribution/development of information. This makes it possible to compare the information structure across texts and get a sense of their level of content sharing and knowledge coherence.

This approach will highlight current automated tools for concept acquisition and ontology learning that are conducive to an incremental access and management of content, to establish a fruitful bridge between modes of knowledge sharing/creation and dynamic, incremental approaches to automated knowledge acquisition and representation.

2.3 Methodological approach

Natural Language Processing (NLP) tools can augment text documents with layers of mark-up data, making the hidden linguistic structure of the document overtly represented and accessible. The input text is segmented down into words and multi-word structures, mutually linked through syntactic relations. Moreover, salient terms are identified in context, to provide access keys to the basic contents of the document. In classical NLP architectures, this is carried out in a step-wise fashion, with layers of annotation being cascaded in a feeding relation, from *tokenized* texts to trees of dependency relations. Typical parsing steps are: i) *tokenization,* ii) *morphological parsing* and iii) *dependency trees*.

Tokenization amounts to assigning a string of characters the status of single token, where a token is the most basic parsing unit, approximately corresponding to a linguistic word, but also including non-lexical units such as dates, addresses, proper names, acronyms, measuring expressions, etc. For tokens to be identified as independent words and assigned their corresponding part-of-speech tag (or grammatical category), their set of morpho-syntactic features (e.g. number, gender, tense, etc.) and their lemma (or lexical exponent), they have to undergo a level of *morphological parsing*. In its simplest instantiation, morphological parsing requires the existence of large repositories of word forms, where each form is glossed with a set of morpho-lexical features. However, in languages with rich morphologies, a closed-list approach to morphological parsing is subject to serious risks of failure, as shown by the German example in (1) (borrowed from Anderson and Lightfoot, 2002):

(1) *Lebensversicherungsgesellschaftsangestellter*
(life insurance company employee)

In fact, no German lexical repository can be expected to be large enough to contain all possible compounds of this kind. A principled solution is to split the compound into its simpler constituent words (Leben + Versicherung + Gesellshaft + Angestellter), for the latter to be looked up in a lexical database as individual entries.

Once word tokens are identified and categorised for contextually-appropriate part-of-speech and lexical exponence, they are grouped into larger constituents defining their *syntactic dependency relations*. A dependency relation is a binary relation linking two tokens in context, usually represented as a pointed arc going from the *dependant token* (a complement or a modifier) to its syntactic head (usually a complemented or modified verb or noun). Dependency relations can also be defined between the constituents of complex NN compounds as shown by the following examples:

(2) a. *life insurance company employee*

b. *china tea cup*

The pointed arcs above tell us that *life insurance* is a type of *insurance*, *life-insurance company* is a type of *company* and *life-insurance company employee* is in fact an *employee*. Incidentally, it should be appreciated that not all dependency chains in NN compounds must look like those in *life insurance company employee*, as shown by the dependency structure of *china tea cup* above*,* where both *china* and *tea* entertain a modifying relationship with *cup*.
Linguistically annotated documents provide a jumping-off point for the acquisition of a more and more abstract representation of the document content, in line with the so-called "layer cake" approach to ontology learning (Buitelaar, Cimiano and Magnini 2005), whereby:

- ✓ words are structured into terms (e.g. *life insurance company* is a complex term),
- ✓ terms are grouped into conceptual classes (e.g. *insurance company* is a co-hyponym of *telecommunication company*),
- ✓ concepts are linked together through vertical (taxonomical, e.g. *life insurance company* is a hyponym of *company)* and horizontal (ontological, e.g. people are typically employed in a *company*) relations.

Such a wealth of information provides the basis to a computational platform for automated document content sharing, access and dissemination, that allows document contents to be queried by concepts and concept relations rather than by fixed text patterns or key-words (Lenci et al., 2008). For example we can search a text for information about the number of employees of a given insurance company or its overall yearly income and the like. With no linguistic information such as in (2.a-b) above, a text can be navigated only through fixed word patterns. Linguistic annotation offers a more abstract level of information which can selectively be searched for intelligent information access.

Another, orthogonal level of linguistic information that can usefully be represented through NLP technologies is the *content accessibility* of a document, defined as the level of readability of a text document calculated on the basis of its processing difficulty. Processing difficulty is a multi-factorial concept, which can be decomposed into several, fairly independent factors, such as lexical richness, lexical density and (morpho-)syntactic complexity. Each factor can be assessed independently through measurable parameters. This type of analysis allows us to get an objective measure of the information density of a text document or document collection and to map out its distribution/development through the document(s). This makes it possible to compare the information structure of different text collections and get a precise sense of their level of informativeness, content sharing and knowledge coherence.

In particular, *lexical richness* has to do with the lexicon of a text document, defined as the set of word types attested in the document. Trivially, a richer lexicon has a higher set cardinality than a poorer one. More subtly, lexical richness also involves word frequency distributions. Rare words are in fact taken to be more difficult to process than more common words and often denote the most salient pieces of content information of a document together with its level of subject-specificity. Accordingly, by inspecting the tails in the Zipfian distribution of different document lexicons we can get a flavour of the different degrees of lexical richness of the corresponding documents (or document collections). *Lexical density*, on the other hand, gives a measure of the rate at which the content of a collection is updated through the introduction of novel concepts and is defined as the number of new words that a text excerpt introduces in a document collection, divided by the length of the excerpt.

Salient domain-specific concepts and relations are most often conveyed in text through statistically significant terms. Relevant terminological units can be tracked down automatically by projecting abstract morpho-syntactic patterns such as "NP PP" (*i.e.* "find a syntactic structure made up out of a Noun Phrase immediately followed by a Prepositional Phrase) onto linguistically annotated texts. Text strings fitting into the targeted morpho-syntactic pattern are then filtered out through a further step of statistical post-processing, to assess their potential for termhood. Since Smadja's (1993) seminal work, statistical methods offer reliable means of acquiring domain specific expectations concerning the joint distribution of words in sufficiently large training corpora. Association measures such as Pointwise Mutual Information (Church and Hanks, 1989) have become standard utilities to measure the degree of collocational association of word pairs in context, by exploiting the intuition that words belonging to the same bracketed pair will co-occur in

corpora significantly more often than what would be expected under a model of chance co-occurrence (based on the frequency of the individual words). Based on this intuition, we can further identify the most salient terms attested in a document collection and their degree of subject-specificity, by comparing their frequency distribution in the target collection with the distribution of the same terms in a balanced, general-purpose corpus.

Finally, a score of (morpho-)syntactic complexity can be calculated on the basis of i) the average length of text clauses (the longer a clause, the more difficult to parse), ii) the way words are arranged in context (an unusual/marked order of words is more difficult than a standard word order), iii) the per-sentence length of the attested dependency chains (shorter chains are easier to be parsed and understood), and iv) the per-word distance between a head and its dependant (the longer the distance the more difficult the relationship) (Dell'Orletta et al., 2011).

This multi-factorial information allowed us to compare the information density of two different text collections: samples of text excerpts from general-purpose social networks, based on friendship relations and on social proximity, and samples of texts produced within the frame of specialized subject-based communities, based on content sharing and supporting relationships.

3. Experimental evidence

3.1 Results

Two distinct experiments were carried out on English and Italian texts. For all experiments, we used the web-based battery of NLP and knowledge-management tools made available on-line by the *Dylan Lab* [1] (Dynamics of Language Laboratory, Institute for Computational Linguistics - Italian Research Council).
In the English experiment, we conducted a cross-evaluation assessment of both grammatical and content-word parameters in three different text collections: i) a sample of messages posted in general-purpose social networks (e.g. Facebook), ii) a sample of message exchanges within subject-based web communities (e.g. LinkedIn), iii) as a base-line, a sample of Grey Literature writings (e.g. *GL 12 Conference Proceedings*). The distribution of terms in the three samples was comparatively evaluated on the basis of the degree of domain-specificity of terms. This is shown in table 1, which gives the average number of terms that were automatically found to be domain-specific by comparison with a general purpose corpus of English language newspapers (*The Wall Street Journal* section of the Penn Treebank). Characteristically, texts exchanged by subject-centred communities contain an average number of domain-specific terms comparable to the number of terms conveyed on average by GL papers, and considerably higher than the average number of terms occurring in general social network comments.

Table 1

	domain specific terms (single and multiple)
Social networks[2]	1.75
Subject-based communities[3]	14.75
GL Papers	15.75

On the other hand, table 2 gives a measure of the average syntactic complexity of our samples, resulting from the weighted integration of several quantitative parameters: lexical rarity of content-words, distribution of part-of-speech tags (an objective measure of morpho-syntactic complexity), average length of chains of dependency links stemming from a single syntactic head, and average head-complement distance measured by the number of intervening words (Dell'Orletta et al. 2011):

Table 2

	difficulty level
Social networks	35.30
Subject-based communities	43.85
GL Papers	55.20

Once more, the three text samples, ranked by increasing values of syntactic difficulty, reflect a gradient of content accessibility which appears to mirror the degree of communicative formality (from less formal to more formal) scored in our text types (Figure 1).

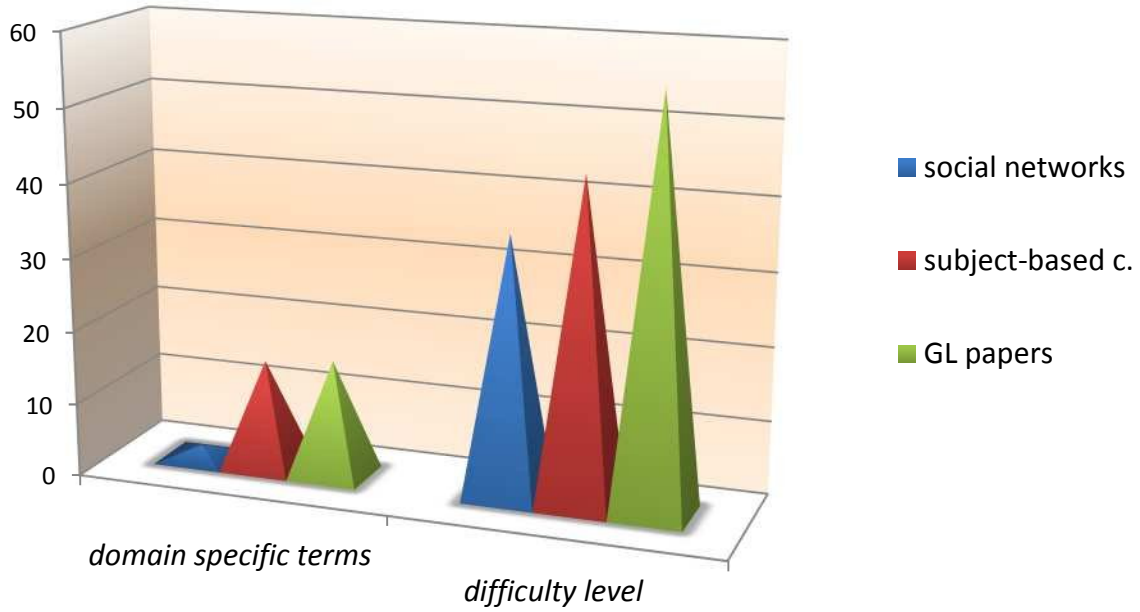


Figure 1

In the Italian experiment, we compared the overall levels of lexical coherence in two samples of Italian post exchanges: i) through a general-purpouse social network (various Facebook accounts' contacts all based on friendship relations); ii) through subject-based technical blogs of a research institution (CNR intranet). Lexical coherence is automatically estimated by measuring the flow of new lexical items that are incrementally added in a post exchange referring to the same issue. This is calculated as the number of novel words introduced by each newly posted comment divided by the length of the comment. Results are summarized in the graph below (Figure 2), providing the overall trend in the average word frequency and lexical density for the two text samples (over an exchange of maximum 20 posted comments).

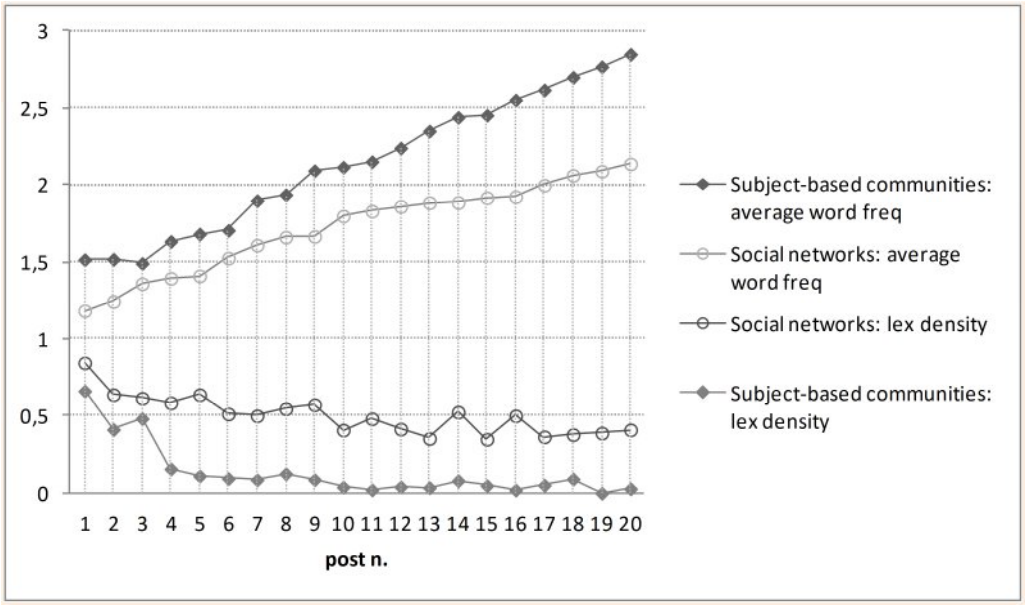


Figure 2

The two trends are remarkably different. Social networks show a slower rate of average word frequency, witnessing a higher variety of lexical choices (the same words are repeated less often). On the other hand, subject based technical writing tends to be lexically more coherent, with a systematic trend towards repetition of the same words. That these words are mostly technical is shown by the average number of domain-specific terms and their high level of difficulty/rarity as shown in table 3:

Table 3

	Single terms	*Multiple terms*	*Average domain specific terms*	*lexicon difficult level*	*Readability level*
Subject based CNR intra-blog	20	5	12.5	50.2	84.7
Social networks	1	3	2	12.5	73.7

3.2 Discussion

The two experiments were intended to test the empirical hypothesis that only subject-based collections offer a coherent flow of shared and structured knowledge, general purpose social networks being more erratic and ephemeral in the choice of discussion topics and domains. This hypothesis is basically confirmed by our results. The medium of Social Networks tends to make communication simpler, with shorter sentences than in traditional texts, medium/highly frequency distributed words, simpler syntax (one verb per sentence), and a high readability score. This is true of all texts that were sampled from writings exchanged through social networks, irrespective of their topic.

However, a simpler communication does not necessarily guarantee coherence of information flow and steady knowledge sharing. Only in those cases where there is a strong interaction between medium and content (as in subject-based community exchanges), we can also observe domain-specific terms, high lexical coherence, more levels of syntactic embedding, more complex readability levels. These are in fact, in our view, the hallmarks of knowledge building and informative flow.

4. Concluding remarks

NLP tools for content analysis and Information Extraction are instrumental in establishing a direct relation between modes of knowledge creation/sharing and dynamic, incremental approaches to automated knowledge acquisition and representation. They allow us to assess the content of a text in terms of its level of readability, domain-specificity, lexical coherence and density of its conceptual maps. They can be used to measure not only the effectiveness of a text in conveying information but also the extent to which this information is structured in terms of shared knowledge.

As cognitive proximity consists of sharing capabilities and knowledge in a broad context, subject based communities, primarily focused on supporting relationships and content sharing, act at the same time as providers and users of all kind of Grey Literature materials in a highly distributed and collaborative scenario, and represent a conducive environment for knowledge creation and transfer. They can represent, as collaborative networks, a key element in the advancement and dissemination of knowledge in scientific domains as well as in diverse aspects of everyday human life.
Conversely, general-purpose social networks, reflecting either friendship or superficial relationships, are virtual meeting place, but tend to generate ephemeral information and to create superficial and mosaic knowledge.
As a general concluding consideration, Social Networking is a medium with a strong potential, a house of cards powerful and delicate at the same time.

References

ANDERSON S. R., LIGHTFOOT D. W. (2002). The language organ: linguistics as cognitive physiology. Cambridge University Press.

BONIN F., DELL'ORLETTA F., VENTURI G., MONTEMAGNI S. (2010). A Contrastive Approach to Multi-word Term Extraction from Domain Corpora. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, Malta, 19-21 May, 3222-3229.

BAEZA-YATES R., RIBEIRO-NETO B. (1999). Modern Information Retrieval. Addison Wesley, ACM Press, New York.

BUITELAAR P., CIMIANO P., MAGNINI B. (2005). Ontology learning form text. IOS Press, Amsterdam.

CHURCH, K., HANKS P. (1989). Word Association Norms, Mutual Information and Lexicography. *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada.

DELL'ORLETTA F., MONTEMAGNI S., VENTURI G. (2011). Read-it: Assessing Readability of Italian texts with a View to Text Simplification. *Proceedings of the 2nd Workshop on Speech and Language processing for Assistive Technologies*, Edinburgh, UK, 73-83.

DELL'ORLETTA F. (2009). Ensemble system for Part-of-Speech tagging. *Proceedings of Evalita'09*, Reggio Emilia, December 2009.

FRANTZI K. T., ANANIADOU S., MIMA H. (2000). Automatic Recognition of Multi-Word Terms: the C-value/NC-value method. *International Journal on Digital Libraries*, 3(2), 115-130.

LENCI A., MONTEMAGNI S., PIRRELLI V. (2006). Acquiring and Representing Meaning: Computational Perspectives. In A.. Lenci , S. Montemagni ,V. Pirrelli . (eds.) Acquisition and Representation of Word Meaning. Theoretical and computational perspectives. *Linguistica Computazionale*, XXII-XXIII, IEPI, Pisa- Roma, 19-66.

LENCI A., MONTEMAGNI S., PIRRELLI V., VENTURI G. (2008). Ontology learning from Italian legal texts, in J. Breuker, P. Casanovas, M. C.A. Klein, E. Francesconi (eds.), *Law, Ontologies and the Semantic Web - Channelling the Legal Information Flood, Frontiers in Artificial Intelligence and Applications*, Springer, Volume 188, 75-94.

MANNING C. D., SCHÜTZE H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.

MARZI C., PARDELLI G., SASSI M. (2011). A Terminology based re-definition of Grey Literature. *GL12 Conference Proceedings,* TextRelease, Amsterdam, 27-31.

MARZI C., PARDELLI G., SASSI M. (2010). Grey Literature and Computational Linguistics: form Paper to Net. *GL11 Conference Proceedings*, TextRelease, Amsterdam,118-121.

NOOTEBOOM B. (2000). Learning and innovation in organizations and economics. Oxford University Press, Oxford.

SMADJA, F. (1993). Retrieving Collocations from Text: Xtract. *Computational Linguistics*, MIT, Cambridge MA, USA, 19(1), 143–177.

TENENBAUM J. M. (2006). AI meets Web 2.0 Building the Web of Tomorrow, Today. *AI Magazine*, American Association for Artificial Intelligence, 47-68.

TEXT ANALYSIS TOOLS http://www.ilc.cnr.it/dylanlab/ developed at *Dylan Lab* (ILC-CNR).

[1] Tools available at http://www.ilc.cnr.it/dylanlab/index.php?page=software&hl=it_IT
http://www.ilc.cnr.it/dylanlab/index.php?page=software&hl=en_US

[2] Facebook excerpts are messages posted during a time slot of 20 days on various accounts, varying in age, social class and professional level. The figure in the table is an average over excerpt samples whose standard deviation is 1.24

[3] LinkedIn excerpts are messages taken from the GreyNet Group, during a tile slot of 20 working days.

Using Social Media to Create Virtual Interest Groups in Hospital Libraries

Yongtao Lin and Kathryn M.E. Ranjit,
Health Information Network Calgary, University of Calgary, Canada

Abstract

Social networking has positively impacted the realm of human interactivity. Although libraries have traditionally been viewed as a place for research and information seeking, Web 2.0 technologies, such as blogs and Instant Messaging (IM), are starting to change users' experiences of current library services.

In this study we aim to examine whether the opportunity for enhanced researcher-librarian interaction that Web 2.0 tools provide creates a synergistic experience for health research interest groups. In addition, we will explore whether these tools increase efficiency in obtaining information and/or improve quality and quantity of research evidence.

This study was conducted by two hospital librarians working in tertiary hospitals. An email was distributed to the health care professionals from these two sites, inviting participation in two online research interest groups: Clinical Practice Guidelines and Patient Safety. A pre-survey of participants was intended to assess the comfort level of this group with Web 2.0 tools and to gauge their level of use both professionally and personally. We created a Virtual Interest Group (VIG) environment to incorporate the following Web 2.0 tools into the existing library website: user blogs, enabling commenting to facilitate a knowledge-sharing atmosphere; chat software to assist with easy access to acquired information; and Delicious tagging for a more systematic documentation of grey literature. A post-survey was conducted three months later in order to re-evaluate the participants' experiences with social media, in particular with the online interest group environment.

Findings from this study can be used to highlight future trends around the discoverability of grey literature with social media tools and to establish a basis for integrating Web 2.0 tools in library websites and services.

Introduction

Online social media, since its introduction, has transformed the means of human interaction and communication in the twenty-first century. Social media, "a group of internet-based applications that … allows the creation and exchange of user generated content"[(1)], has advanced users' capacities in publishing information, communicating with others, and connecting and sharing with like-minded people. Social tools such as Facebook and Twitter have become dominant tools for creating knowledge, accessing information, collaborating and networking within and beyond a physical community.

Teaching and learning has often been a field where innovations in computing technology are adopted earlier than in other domains.[(2,3)] Educators have, therefore, begun to explore how new paradigms of learning, such as Education 2.0, can be implemented using social media tools, including Wikis, blogs, online groups, virtual worlds, social tagging and bookmarking. These applications enable innovative behaviours in acquisition, access, processing, retrieval, presentation and distribution of information within the learning space.[(4-7)]

The impact of social media technologies is not pervasive in education; everything is affected, especially in the scholarly and research activities. Our society is increasingly interconnected through social media. Social networking can assist in developing online resources, providing a collaborative research environment. Knowledge creation and scholarly communication are moving away from a world in which a few producers generate content to transmit to a set of users. Instead, research dissemination now has various routes via a wide range of collaborative tools.[(8)] The influence of social media on the rankings of articles has begun to be acknowledged. Binfield (2009) reviewed the metrics used by Public Library of Science (PLoS) to analyze usage of articles appearing in their journals. They include not only page views and downloads, but also citation counts, comments, ratings, social bookmarks, and blog coverage.[(9)] More research data is being made available online, thus greatly impacting the perception of research results. Several well-designed studies[(10-14)] have also been published on scientists' use of Web 2.0 for information seeking of scientists, the impact of new technologies on health research, and dissemination of information through social networks.

Collective learning and social media

The advent of social media technologies imparts a new formative experience in collaborative group learning, which is often referred to as *Collective Learning*. Related to an instruction model known as active learning[3], collective learning focuses on "the process of learning and ensures learners are cognitively engaged."[15, p38] Collectively, the individuals can discuss, explore and develop educational content, and answer each other's questions. This enhances the engagement of learners with content, as well as their capability to actively learn and collaborate.[16]

While control is shifted from instructors to learners, undoubtedly, there are challenges to overcome. The changing roles require learners to have a strong motivation to work in groups to drive the self-directed learning process. When learning takes place through online social interaction, it challenges collaborators with sets of uncertainties and assumptions and the need to develop mutual trust between peers in a virtual research environment.[15] Furthermore, the success of learners and the unstructured process are extremely difficult to measure.

The potential that social technologies, such as blogging, social bookmarking and social tagging, can offer foster immediate dialogue and stimulation. These applications feature the capability to support virtual social interactions and build social communities, all promising an easy-to-use, more open and dynamic platform.

The collective learning paradigm is extremely helpful in organizations[17], where employees are expected to explore, experience, and teach themselves new technologies without formal training. Primary characteristics of social media, including accessibility of most social media sites; flexibility in refining and altering; currency and usability of these technologies help in supporting collective learning based environments. Even the social media categories (namely, *wikis, blogs, microblogging, media sharing, social bookmarking, social friendship networks and social news sites*)[15, p40] align well with the collective learning model's goals of interactivity and proactivity. The presence of collective learning is evident in organizational learning groups such as communities of practice and virtual research environments (VRE) via social media platforms. In the context of a VRE, social networking sites such as Facebook provide an easy means of finding research partners and keeping them in touch with changing interests. Building a virtual community platform[18-20], where subject information guides, information gateways, literature alerts, literacy training and chat references are one-stop accessed, gives subject users a more convenient and more attractive environment to preserve their research results and to share their experience.

Social media and libraries

Libraries are major information providers and facilitators. As new communication technologies support collaboration among scientists, researchers, students, faculty members and all other users, it offers libraries and librarians the opportunity to be more proactive in facilitating the use of social networking for resource finding and sharing. Joe Fernandez (2009)[21] presented a SWOT analysis of social media in libraries, allowing libraries to see the strengths and weakness that social media offer, recognize the opportunity of using social media to develop a more dynamic relationship between themselves and their users, and develop awareness of the threats that they might face when getting into the Web 2.0 realm and help them respond to the challenges accordingly.

There have been numerous studies highlighting the ongoing research activities on the integration of social media in library information and knowledge management services. The Web 2.0 tools are becoming mainstream and are increasingly more and more expected by users. For example, blogs are a popular tool used by librarians and information scientists to collaborate, connect and instruct, or to provide news and current awareness services to users[22], both as a form of publication and as tools for marketing their resources and events.[23] [4] Moreover, Bar-Ilan (2007)[24] proposed that library blogs are ideal for disseminating, commenting, and expressing opinions. Social bookmarking services such as Delicious allow users to tag (describe), save, manage and share web pages. Various library subject/resource guides are created using this social bookmarking service to facilitate the retrieval of information on the Internet.[6] Information seeking behaviour within the virtual world of Second Life[5] was found to be rich, complex interactive with multiple facets. Health libraries develop integrated personal information portals[25] to promote evidence-based information to health information consumers. Information literacy programs are a

core service component in nearly all types of libraries, but a challenging one to meet. Learners have found themselves more engaged through interactive online learning modules created by Web 2.0 technologies.[26]

Social networking and virtual interest groups (VIG)

Social networking, incorporating Web 2.0 technologies, has transformed the way information professionals are creating, storing, retrieving and disseminating information. Frameworks have been developed to facilitate the development of virtual communities.[28,29] Virtual interest groups (VIGs), are online communities that use social networking sites to allow collaboration between users logically grouped with common shared interests.[30, 31] Research has shown that the VIGs created and facilitated via social media platforms provide an integrated environment that supports the work of a community or group of collaborating researchers.[19]

Background

The study was conducted in two tertiary hospitals in Calgary, the fourth largest municipality in Canada and largest in Alberta.[27] The Health Information Network Calgary, a network of hospital libraries located at six major Calgary sites, provides information and library services to patients, the public and health care professionals working in the Calgary Zone and in Cancer Care. Collaboration across disciplines, geographical distances and political boundaries is increasingly accomplished in the hospitals where we provide information support.

Two key subject areas of interest that we, as librarians, have developed through our work in reference, instruction and knowledge management support are c*linical practice guidelines* and p*atient safety*. Tailored information services are already in place contributing to the knowledge creation of the groups, whether through subject-based e-resource guides, or through alert services to inform researchers and health care practitioners the current literature on the related topics. Webinars and discussion forums are hosted to facilitate the sharing of information among researchers and practitioners.

Objective and Research Methodology

Our study aimed to examine whether the opportunity for enhanced researcher-librarian interaction that Web 2.0 tools provide creates a synergistic experience for health research interest groups, and to explore how the content is co-developed, assessed and further improved by users. We also hoped to gain a better understanding whether social media tools improve the discoverability of grey literature.

Emails were distributed to library users in two hospital libraries to recruit study participants for two VIGs: clinical practice guidelines and patient safety. A preliminary online survey was administered in order to understand the participants' experience using social medial tools, how they used the technologies and possible barriers to using these tools for online research purposes. The survey was also intended to help the librarian facilitators identify training needs for the social media tools selected.

Blogger (www.blogspot.com) was adopted to create two blogs as a gateway platform for the other two tools integrated in our research group activities. Group members were invited to share resources and questions with their peers via this platform. Subject-related information literacy training and information consultation between researchers and librarians, and among researchers themselves were conducted via a browser-based chat service. Two platforms were explored: *Meebo* (https://www.meebo.com/) and *Chatroll* (http://chatroll.com/). Chat was used to connect with researchers virtually in real time and to enhance collaborative decision-making. Tags created for websites about clinical practice guidelines, patient safety and social media on Delicious (www.delicious.com) were linked from the two blogs. A number of networks from these tags were formed among researchers when these specified tags were subscribed to and the participants were followed.

After three months, a post-survey was distributed via email to gather feedback about the pilot project, assess comfort level with social media tools for research purposes, determine barriers to uptake during the pilot, and seek recommendations for the future services in VIGs.

Results

Pre-Survey

Twenty-five people indicated interest in participating in the VIG pilot and responded to the pre-survey. Of these 25, all elected to join the Clinical Practice Guidelines VIG and eight indicated interest in Patient Safety. Most participants had experience with at least one social media tool prior to participation in the VIGs (Table 1) and only two participants self-reported discomfort from lack of experience with social media.

Table 1. Responses of 25 participants who were asked, "What social media tools do you use?"

Tool	Number of participants	Percentage
Flickr	2	8
Twitter	6	24
Delicious	0	0
Facebook	18	72
MySpace	0	0
YouTube	13	52
Ning	0	0
Blogger	2	8
LinkedIn	2	8
Other	1	4

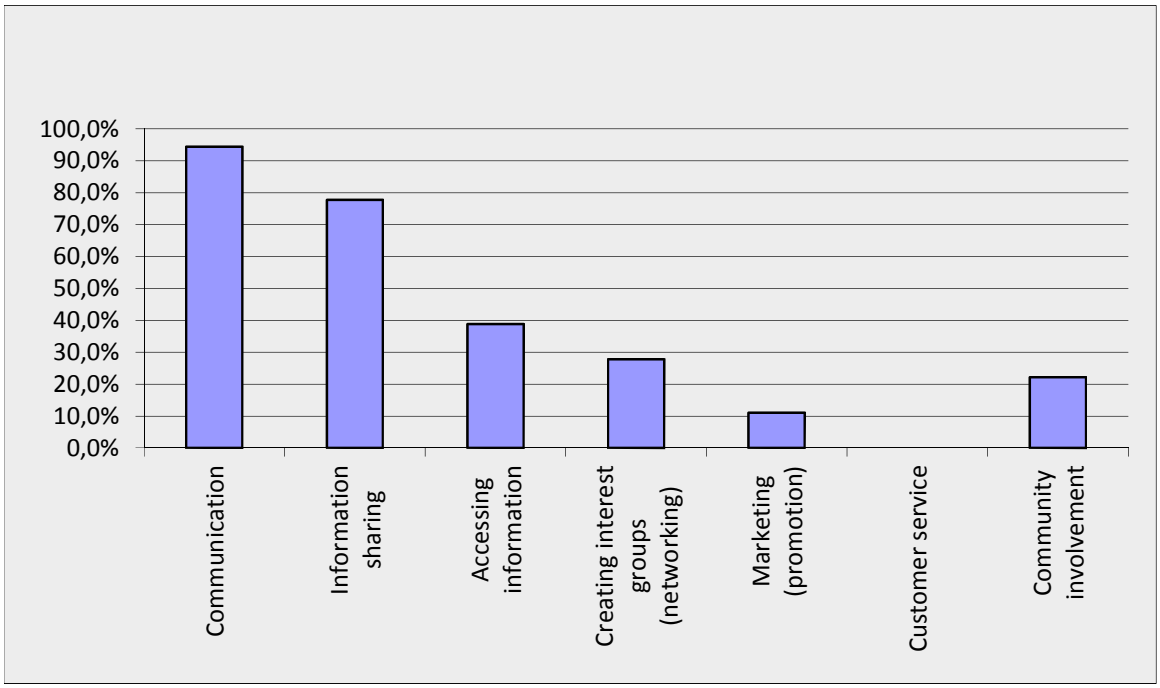


Figure 1. Participants' previous uses of social media tools.

Most participants had used social media tools for communication and information sharing purposes prior to the pilot (Figure 1).

Self-reported barriers to using social media tools included limited workplace access, time constraints, discomfort with the medium, and privacy issues. Training was requested in creating good online content and directions to use the tools proposed for the VIGs.

Blog

A blog was created for each VIG. Twenty-one participants registered for an account with one or both blogs. Twenty posts were published during the three-month period. Types of information shared in the blog posts included an introduction to each piloted social media tool and "how-to" directions, two reviews of articles pertinent to the subject areas, information about guideline review tools including AGREE-II, favourite free resources including websites and a participant's departmental newsletter.

Chat

Four live chat sessions were held. Eight people participated in one or more chat sessions. Chat sessions were scheduled approximately every three weeks. Sessions were an open forum but each session was prefaced with questions provided to the group to foster discussions.

- Session 1: Troubleshooting issues with signing up for the blogs, why were participants interested in joining the VIGs, social media in health care and previous experience with the VIG topic
- Session 2: Social media in health care, social media use in their workplace, troubleshooting issues with Delicious
- Session 3: Reflections on recent posts, using guideline appraisal tools, exploring the use of other social media tools for research and learning purposes
- Session 4: Favourite resources in the VIG topic: websites, medical journals/databases, consultation with colleagues, etc. and why these are good resources.

Types of information shared in chat sessions included recent developments related to the interest group topic, reflections and discussions about recent posts, interest group resource sharing, discussions about the use of social media for specific information sharing purposes and new Social Media tools to explore.

Delicious

The librarian facilitators created three initial bookmark lists (tags) to share with the virtual interest groups: Clinical Practice Guidelines, Patient Safety and Social Media in Health Care. Twelve participants signed up for Delicious accounts. Participants bookmarked favourite websites during in-person training sessions. Users were encouraged to follow each others networks; however, no one was able to complete this task after the training sessions.

Types of information collected in tag lists included government websites, alert websites, online reports, statistical reports, online appraisal tools, popular reviews and videos pertaining to the interest group topics.

Post-Survey

Eight participants (32%) responded to the post-survey. Participants were asked about each tool used during the pilot, indicating how much they used each tool and assessing the value of each for research and education purposes. If the participant did not use a particular tool, they were asked to indicate why. Finally, participants were asked to characterize their whole experience in the VIG pilot.

Overall, participants reported that participation in each tool led to increased comfort level using these social media tools. Five out of eight respondents were unable to interact with one or more of the social media tools during the pilot. The greatest reported barrier to participation was time constraints (Table 2).

Table 2. Self-reported barriers to use of individual social media tools used in the Virtual Interest Groups.

	Blogs (n=2)	Chat (n=5)	Delicious (n=5)	In-person (n=3)	Training
Too busy	2	5	3	3	
Insufficient training			1		
Not interested					
Not useful			1		

The blog content was viewed as useful to most (five out of six) participants who read them. Chat sessions were reported as most useful for sharing and discussing interest group resources, as well as for connecting with VIG collaborators and troubleshooting issues. Of the tools introduced in this pilot, Delicious bookmarking was considered to be the most difficult tool to use. Participants reported difficulty in creating accounts for this tool and in using it to find useful resources even after in-person and online training.

Participants joined the virtual interest groups for a variety of reasons (Table 3). Most indicated the motivation to learn more about the virtual interest group topics and to connect with their colleagues over these topics.

Table 3. Participants' motivations for joining the Virtual Interest Group pilot study.

Motivation	Percentage of participants[a]
I didn't have any experience in social media so I wanted to learn some basics.	37.5%
I had some previous experience with social media and wanted to learn how social media is used for work-related purposes.	25.0%
I wanted to be involved in the discussions of the topic(s): patient safety and/or clinical practice guidelines.	62.5%
I wanted to be part of the groups to make more connections to people with interest in patient safety and clinical practice guidelines.	50.0%
I signed up because the librarian recommended the interest groups and potential benefits to me.	25.0%

[a] Percentage calculated based on number of Post-Survey respondents (n=8)

Overall, participants characterized the experience as positive and reported an increased awareness of using social media tools for information sharing. Negative experiences noted were primarily due to lack of time to participate fully.

Discussion

Evolution of third interest group in social media and health care

Social Media is gaining an ever increasing role in health care.[32] While this pilot was designed to introduce the use of social media for learning and collaborative purposes in health care topics, it was not originally intended to address the prevalence and best practice of social media use in the health care environment. This topic, however, became a recurring theme throughout the pilot and therefore through this collective learning model the group adapted the VIGs to address questions in this topic as needed.

Social media training

The intent of this study was to facilitate as much as possible in the online environment, including training in the use of social media tools. Online learning is increasingly becoming the best practice model for continuing education in the researchers' health care environment.[33] However, some participants were unable to participate as desired, perhaps due to the limited time invested in online learning as each tool was introduced (Table 2). It was observed that facilitated in-person training increased the comfort level and motivation to participate in the VIGs. The introduction of an in-person training session at the outset for learners who prefer this learning style should be considered for conducting similar projects. This introduction would likely increase participation.

Enhanced information creation and grey literature sharing in virtual interest Groups

While this pilot was small, there is strong evidence of increased information sharing and creation. Some resources that were deemed valuable by this interest group were shared in each of the social media tools used in this pilot, such as government websites, online reports and freely available research tools on the Internet. Participants from different departments and roles at three different sites came together to participate in the VIGs. The collaborative online environment allowed these professionals to share and learn from others in their organization with whom they would not usually have the chance to collaborate in-person. Information gathering took place in the use of each of the different tools. This may have been most evident in the chat sessions where questions were able to be addressed in real time. A more stable form of the shared information could be followed later in the form of a blog post or by tagging into Delicious. Examples of this can be in the sharing of departmental Twitter, which were not previously known or an organizational newsletter that was later shared in a blog post.

Future Directions

Future research will look into facilitated VIGs among health professionals and with librarians as participants. With librarians able to focus as participants and content creators instead of facilitators, the opportunity to create more awareness of grey literature in health VIGs may be more easily achievable. As more data is collected the researchers would like to explore a semantic approach of visualizing these conversations using tag-based data mapping, which will greatly improve the discoverability of grey literature in health topics.

References

(1) Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Bus Horiz 2010 2;53(1):59-68.

(2) Johnson L, Smith R, Willis H, Levine A, Haywood K, New MC, et al. The 2011 Horizon report. New Media Consortium 2011 01/01.

(3) Bonwell CC, Eison JA, Association for the Study of Higher Education, ERIC Clearinghouse on HE, George Washington Univ, Washington, DC School of Education and Human Development. Active learning: creating excitement in the classroom. 1991 ASHE-ERIC Higher Education Reports. ; 1991.

(4) Goss H, Jr. Extending library instruction: using blogger to collaborate, connect, and instruct. Journal of Library & Information Services in Distance Learning 2010 Oct-Dec;4(4):166-184.

(5) Ostrander M. Talking, looking, flying, searching: information seeking behaviour in Second Life. Library Hi Tech 2008;26(4):512-524.

(6) Darby A, Gilmour R. Adding Delicious data to your library website. Information Technology and Libraries 2009 JUN;28(2):100-103.

(7) Dlouha J, Dlouhy J. Use of Wiki tools for raising the communicative aspect of learning. 2009.

(8) van Zyl AS. The impact of social networking 2.0 on organisations. Electronic Library 2009;27(6):906-918.

(9) Binfield P. PLoS One: background, future development, and article-level metrics Proceedings of the 13th International Conference on Electronic Publishing (Milan, 10–12 June 2009) 2009:69-86.

(10) Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of Web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 10/01;78(10):645-655.

(11) Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. Am J Infect Control 2010 04/01;38(3):182-188.

(12) Saver C. Tweeting, posting, and yammering: the role of social media in the OR. OR Manager 2010 Feb;26(2):1, 12-4.

(13) Segbers R. Go where the customers are. Marketing (and managing) your patient experience with social media. Mark Health Serv 2010 Winter;30(1):22-25.

(14) Moore A. A very sociable guy, or girl...: could social media 'branding' be a useful measure of a scientist's impact? Bioessays 2010 Feb;32(2):95.

(15) Agarwal N. Collective learning: an integrated use of social media in learning environment. In: White B, King I, Tsang P, editors. Social Medial Tools and Platforms in Learning Environments. 1st ed.: Springer; 2011. p. 37-51.

(16) Bolhuis S. Towards active and self-directed learning. Preparing for lifelong learning, with Reference to Dutch Secondary Education. 1996 04/01.

(17) Backstrom T. Collective learning: a way over the ridge to a new organizational attractor. Learning Organization 2004 01/01;11(6):466-477.

(18) Pan W, Lan X. Building a virtual community platform for subject information services at Shanghai Jiao Tong University Library. The Electronic Library 2009;27(2):271-282.

(19) Voss A, Procter R. Virtual research environments in scholarly work and communications. Library Hi Tech 2009;27(2):174-190.

(20) Sonnenwald DH, Lassi M, Olson N, Ponti M, Axelsson A. Exploring new ways of working using virtual research environments in library and information science. Library Hi Tech 2009;27(2):191-204.

(21) Fernandez J. A SWOT analysis for social media in libraries. Online 2009 SEP-OCT;33(5):35-37.

(22) Chen C, Wu Z, Ran C, Tang Q, Chen S, Zhang X. A dynamic RSS information push service mechanism based on ontology of user information needs. Electronic Library 2009;27(2):222-236.

(23) Aharony N. Librarians and information scientists in the Blogosphere: an exploratory analysis. Library & Information Science Research 2009 Sept;31(3):174-181.

(24) Bar-Ilan J. The use of Weblogs (blogs) by librarians and libraries to disseminate information RID B-3452-2009. Information Research-an International Electronic Journal 2007 OCT;12(4):323.

(25) Koonce TY, Giuse DA, Beauregard JM, Giuse NB. Toward a more informed patient: bridging health care information through an interactive communication portal. Journal of the Medical Library Association 2007 Jan;95(1):77-81.

(26) King S, Kaplan R, MacDonald K. Teaching PubMed in cyberspace: the development of a self-learning package. Med Ref Serv Q 2008;27(3):272-283.

(27) Statistics Canada. Population of census metropolitan areas 2006; Available at: http://www40.statcan.gc.ca/l01/cst01/demo05a-eng.htm. Accessed 1/9/2012.

(28) Tilley CM, Bruce CS, Hallam G, Hills AP. A model for the development of virtual communities for people with long-term, severe physical disabilities. Inf Res 2006;11(3):55-69.

(29) Perspectives on knowledge sharing system for special interest group (SIG): Culture, power issues and user interface. 2009 International Conference on Computer Technology and Development, ICCTD 2009; 13 November 2009 through 15 November 2009; 2009.

(30) Kallurkar S. Targeted information dissemination 2008; Available at: http://www.dtic.mil/docs/citations/ADA480150. Accessed 1/9/2012.

(31) Myhill M, Shoebridge M, Snook L. Virtual research environments - a Web 2.0 cookbook? Library Hi Tech 2009;27(2):228-238.

(32) Squazzo JD. Best practices for applying social media in healthcare. Healthc Exec 2010;25(3):38.

(33) Alberta Health Services. E-learning: What and how. 2010; Available at: http://insite.albertahealthservices.ca/3645/asp. Accessed 1/8/2012.

what is grey?

The Central Registry of Publication Activity - CREPC

Information source of grey literature produced at slovak universities

thesis DISSERTATION and HABILITATION

proceedings SCIENTIFIC and CONFERENCE

studies SCIENTIFIC and RESEARCH

research reviews FINAL and CONTINOUS

works PUBLISHED at the INTERNET

preprints

ABOUT CREPC:

- established in 2007 as a development project of Ministry of Education of the Slovak Republic
- the main objective: comprehensively register publication activities of the Slovak universities in electronic form
- the grey objective: evidence of grey literature produced at slovak universities
- material and technical support: The Slovak Centre of Scientific and Technical Information
- softver solution: product of SVOP Ltd.
- content: 165 933 bibliographic records
- public access: www.crepc.sk

CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR

CONTACT:
The Slovak Centre of Scientific and Technical Information
Lamačská cesta 8/A, 811 04, Bratislava, Slovak Republic

Supported by national project:

Európska únia
Európsky fond regionálneho rozvoja

Operačný program VÝSKUM a VÝVOJ

National Infrastructure for Technology Transfer Support in Slovakia – NITT SK **http://nitt.cvtisr.sk**
Supporting research in Slovakia / The project is co-funded by the EU

Information-Seeking Behaviour of Slovenian Researchers: Implications for information services

Polona Vilar, Primož Južnič, and Tomaz Bartol
University of Ljubljana, Slovenia

Abstract
The paper present one part of a survey of information behaviour of Slovenian researchers. Results show that Slovenian researchers in most areas show usual traits of scholarly information behaviour. Exceptions are the non-use of Web 2.0 tools for research purposes and low use of open-access materials. Survey confirmed that ICT is influential in perferences regarding resource formats, access, means for information exchange, organization of resources, writing, reading, etc. The use of grey literature is quite intensive, but depending on the academic area and research field: researchers from humanities, natural sciences and other technical fields are more inclined then social science researchers to use grey literature as the source of information for their research, and business sector is of all sectors the most concerned with patents and standards.

1. Introduction

In recent years a lot of research has been dedicated to information behaviour of various user groups, among these also to scientists. With more and more resources becoming available online, there is an evident and rapid trend towards the development of different information behaviour by scientists: what information resources they are using, how and when. Recent studies of scholarly information behaviour all show significant changes in the ways researchers communicate (Maron and Smith, 2008), publish their works (Dallmeier-Tiessen et al., 2010; Muench, 2011), collaborate (Borgman, 2009), look for information and use it (Rowlands and Fieldhouse, 2007; Information Behaviour, 2008; Palmer, Teffeau and Pirmann, 2008). The origin of all these changes is undoubtedly the impact of digital technologies.

To go with this trend we have in 2011 conducted an extensive survey of information behaviour of researchers in the Republic of Slovenia. We here present part of its results. The study as a whole aims to shed more light on the patterns of their information behaviour thus facilitating the work of research organizations, information providers such as libraries, providers of publicly funded information sources such as public research agencies. The results help to better understand the research work, evaluate current offer of information sources, as well as plan the future provision.

The segment of the results presented in this paper focusses on issues of resources and media used for research purposes. We looked at use and preferences of information sources, types of information sources used (with emphasis on grey literature), and impact of information and communication technilogies on information-related activities.

2. Review of Literature

2.1. Information behaviour of researchers

Among other information related activities these studies present findings regarding scholars' use of social media, use of various information sources and of various formats. It is undoubtedly becoming clear that immense impact of the digital technologies has led to certain new patterns in scientific information and communication behaviour. Researchers have noticed that although scholars still perform some 'older' activities, such as browsing or berrypicking (Bates, 1989, 2007), they also express some previously unseen behaviours, e.g. skimming (looking at one to two pages at a time), navigating (looking around at what is available, i.e. 'the electronic sweet shop'), power browsing (reading abstracts and titles, even indexing terms, rather than full text), squirrelling (downloading material to `read' later), cross-checking (collecting information from different sites) (Rowlands and Fieldhouse, 2007; Maron and Smith, 2008; Information behaviour, 2008; Palmer, Teffeau and Pirmann, 2008).

In terms of resources the fact is that researchers increasingly use different resources. For some scientific communities like High Energy Physics needs immediate access to information resources and even open

access to to full text is not enough anymore. So the grey literature became the most important communication channnel (Gentil-Becot A. 2010).

For several years in the context of e-science the initiatives of data curation, reuse of data and information, have been gaining attention. By data curation we mean management of existing research data, adding value to this data, immediate sharing of it for re-use, and long-term preservation of this data for later re-use (Choudury, 2010, Lord et al., 2004; Lord and Macdonald, 2003). This issue has been tackled by OCLC, One of the questions which arises due to the changing role of libraries serving science is not, should libraries (especially academic) undertake data curation or not, but, in what way could libraries be included in these already ongoing activities? (Gold, 2010).

2.2. Slovenian research situation and infrastructure

Development and growth of research activity in Slovenia is best shown by increase of number of research groups. In 1998 in Slovenia operated 753 research groups, in the fields of natural sciences 121 research groups, in the fields of engineering 346 research groups, in the fields of medical sciences 75 research groups, in the fields of agricultural sciences and biotechnology 60 research groups, in the fields of social sciences 98 research groups and in the fields of humanities 53 research groups. Ten years later, in 2008 in Slovenia operated 1128 research groups, which means 50% increase. In the fields of natural sciences 181 research groups, in the fields of engineering 558 research groups, in the fields of medical sciences 95 research groups, in the fields of agricultural sciences and biotechnology 80 research groups, in the fields of social sciences 142 research groups and in the fields of humanities 72 research groups.

In the year 1998 the number of registered researchers was 6971. Ten years later, this number almost doubled, with 12.182 registered researchers. The age structure of researchers was as follows: 39% of researchers were aged up to 35 years, 31% from 35 to 44 years, 19% between 45-54 years, and 11% above 55 years of age. There were 38.9% female researchers. Figure 1 shows thenumber of researchers, is number of FTE researchers (Figure 1). In the year 2009 on different programmes and projects (basic, applicative and postdoctoral), financed by Slovenian Research Agency (SRA) 4560 researchers have taken part (Peclin, Juznic, 2012).

Republic of Slovenia allocates around 1.5% of its BDP to support science and research In Slovenia investment in public sector is around 100 € per capita, which puts Slovenia, relatively low, in the middle among European union countries. The results, measured by number of ppublications look more promising. In the year 2002 Slovenia had 809 papers per million inhabitants average of EU members was 629. In the year 2005 Slovenia had 1.104 papers per million inhabitants. EU members average was 887 and in the year 2008 the ratio 1.637 vs. 1.037. In the year 2002 Slovenia was on 9 th position, in the year 2005 on 8 th and in 2008 on 5th.

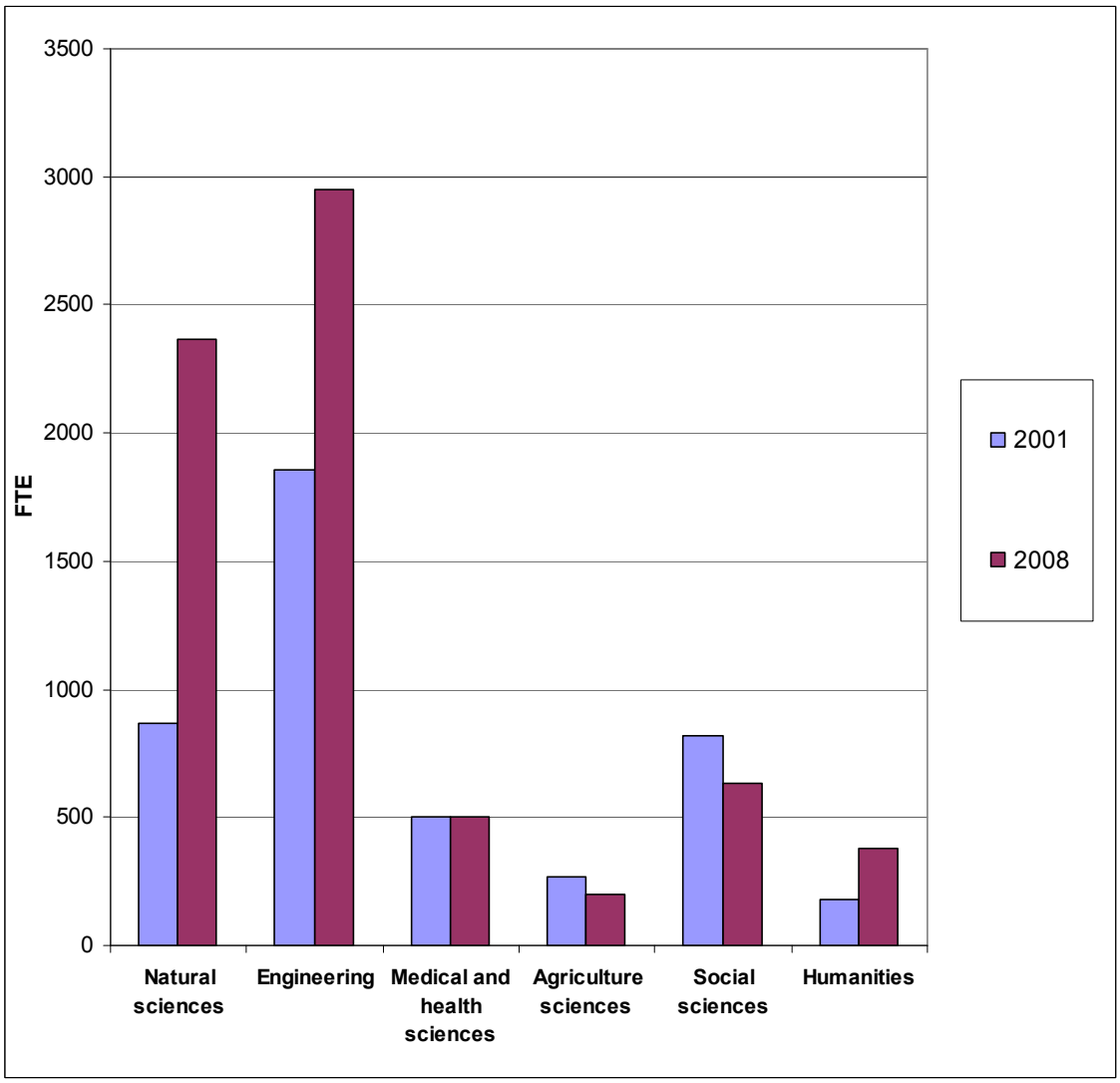


Figure 1: Number of researchers (FTE) by scientific disciplines

3. Research

3.1. Methodology

3.1.1. Sample

We used a random sample of all currently active and officially registered researchers in Slovenia. Contact details were obtained from Slovenian Research Agency (ARRS), the agency in charge of all publicly funded research in Slovenia. The agency has set up a publicly available system SICRIS (Slovenian Current Research Information System; website: http://sicris.izum.si) to monitor the activity of every publicly funded researcher in the country. Within this system every researcher is registered with a unique research ID number. In the year 2011 there have been around 4.800 (?) active[1] researchers. The ARRS agency provided us with researchers' contact details. Sample consisted of every eighth researcher according to allocated research ID number, which meant that there were 592 researchers who received a personal email invitation to participate in the online survey (opened from September 14th to November 14th 2011). Response rate until October 24th when we finished preparation of this text (although the poll was at that time still open) was 33.1% (196 people) of which 20.1% (119) responded adequately. Although not all questions were answered by all participants, the 119 responses still allowed a fairly detailed analysis.

3.1.2. Methods of gathering data

We used an online survey consisting of 25 questions: 18 content questions (Likert scale type) and 7 demographic questions. The latter seven asked participants for their gender, age, type of their current work (research, teaching), years of experience with either research or teaching, employment status (independent researcher, working at a research organization, university, commercial organization), and research area[2].

Content of the survey (initial 18 questions) was dedicated to various aspects of information behaviour. As mentioned, they were mostly Likert scale type questions. In this study relevant were:

- preferred format of information sources (printed, electronic, either)
- opinion on the impact of information and communication technology on various aspects of research work (searching and gathering of information sources, relevance judgement, organization of acquired sources, citation checking, reading, communication, independent writing, collaborative writing)
- number of printed and electronic sources in the personal archive
- frequency of use of information sources in the personal archive
- frequency of instances when teh researcher uses electronic tools for searching for sources, but prints the sources for reading
- ratio of certain types of scientific sources which are normally cited in researcher's publications: scientific publications, electronic publications, open-source publications
- ways of acquiring scientific publications (personal subscription to printed or electronic journals, organizational subscriptions of printed or electronic journals, from e-archives or repositories, interlibrary loan or document delivery services, directly from colleagues) and the frequency of each
- attitudes and opinions concerning various aspects of handling data curation
- types of sources used in research work (eg. formal sources like books, journals, reference material, or informal (grey) sources like project reports, dissertations, social networks, blogs, forums, websites) and frequency of each
- tools used to start search for information sources for research purposes (e-journal providers, specialized bibliographic databases, specialized information portals, web search engines, library catalogue)

As explained above, every participant received an email invitation and we used the data assembled till October 24th.

3.1.3. Data analysis

The data was analyzed using SPSS software. Besides descriptive analysis we also preformed bivariate statistics to see if any connections between demographic and content variables could be identified.

3.2. Results

3.2.1. Characteristics of respondents

There were 46,5% female researchers. The age structure of respondents was as follows: 27.6% were aged from 20-30 years, 36.7% 31-40 years, 17.3% 41-50 years, 12.2% 51-60 years and 6.1% above 60. The structure of respondents according to research area is shown in Table 2. Of the 119 respondents who are all, as explained, active researchers, 91 are involved in research, 60 in teaching and 16 in other activities (more than one current activity is possible for an individual; for example, besides being researchers some are also university teachers, medical doctors, etc).

Table 2: Structure according to research area[3]

	No.	%
Natural Sciences	30	25.2
Technical Sciences	20	16.8
Medicine	12	10.1
Biotechnology	8	6.7
Social Sciences	21	17.6
Humanistic Sciences	15	12.5
Interdisciplinary Research	13	10.9
All	119	100

We can say that they are rather experienced both in research and in teaching. The data can be seen in Table 3. In terms of their employment status, 3.4% work as independent researchers, majority are employed either at university (52.9%) or some type of public research institute (32.7% in research

organizations which are part of some institution and 5.9% in independent research organizations), 2.5% work in economy and 4.2% in other organizations, such as hospitals (again, one individual can be employed in one or more institutions). When asked about their current work tasks (multiple answers were possible), 91.9% answered that they do research, and 60.6% that they also teach – from this we can see that two thirds are actually involved in both activities.

Table 3: Experience in research and teaching.

	Research experience		Teaching experience	
	No.	%	No.	%
Less than 1 year	1	1.0	16	19.3
1-5 years	29	29.6	27	32.5
6-10 years	24	24.5	13	15.7
11-15 years	15	15.3	9	10.8
Over 15 years	29	29.6	18	21.7

3.2.2. Information behaviour

3.2.2.1. General traits of information behaviour

Generally speaking, we found that Slovenian researchers do not differ to a great extent from what has been discovered for today's scientists. In the first section we present some general traits of their information behaviour. If we look at the use of types of information sources – with this we mean the ways researchers come to information they need for research purposes – (Table 4), we see that researchers intensively use formal sources (which is, of course, understandable). However, also the use of informal sources (grey literature) is quite strong. Over one third of researchers (35.3%) use them always or often, and another fourth (24.2%) ocasionally. Research reports and dissertations are used often by one fourth (25.3) and ocasionally by half (51.5%). Over one fourth (27.5%) often or always acquire information from e-archives, if we add those who use them ocasionally (38.5%), this is more than two thirds.

Personal contacts are another strong source of information. From Table 4 it is evident that communication produces a lot of research-related information. It is evident that researchers communicate in all fields: with colleagues within their own organization, with Slovenian colleagues, and with colleagues abroad. Intense contacts with foreign colleagues indicates how aware are Slovenian scientists of importance of international contacts and of international character of science in general. Table 4 also shows that they very often use colleagues to acquire resources (articles). 23,6% do this often or always and another 47.3% ocasionally.

Table 4: Types of information resources used and their frequency

Resource type	Frequency of use (%)				
	Never	Almost never	Occasionally	Often	Always
Printed books	2.0	11.1	36.4	30.3	20.2
E-books	3.0	12.1	48.5	24.2	12.1
Printed journals	0	21.2	34.3	28.3	16.2
E-journals	0	3.1	12.2	43.9	40.8
Reference sources	5.1	20.4	46.9	19.4	8.2
Patents, standards, reports	22.2	41.4	24.2	11.1	1.0
COBISS/OPAC (Slovenian union cat.)	0	13.1	33.3	34.3	19.2
Bibliographic databases	10.1	14.1	28.3	33.3	14.1
Raw data sources	39.8	26.5	15.3	15.3	3.1
Proceedings	4.0	23.2	40.4	27.3	5.1
Prepints	9.1	41.4	29.3	20.2	0
Reviews	14.1	44.4	30.3	11.1	0
Research reports, dissertations	2.0	21.2	51.5	25.3	0

Communication with colleagues in own org.	2.0	12.1	37.4	40.4	8.1
Communication with colleagues in Slovenia	4.0	22.2	48.5	19.2	6.1
Communication with colleagues abroad	4.1	20.4	46.9	20.4	8.2
Social networks	65.7	19.2	12.1	3.0	0
Forums, disc. Groups	28.3	36.4	23.2	11.1	1.0
Library	16.5	18.6	41.2	17.5	6.2
Email alerts	19.2	18.2	34.3	21.2	7.1
Blogs	51.5	31.3	10.1	6.1	1.0
Invisible college (conferences, meetings, etc.)	3.0	19.2	44.4	27.3	6.1
Web portals	30.6	29.6	24.5	12.2	3.1
Websites	5.1	14.1	41.4	32.3	7.1
E-archives	17.3	30.6	33.7	14.3	4.1

Table 5: Ways of acquiring resources and their frequency

How resources are acquired	**Frequency (%)**				
	Never	**Almost never**	**Occasionally**	**Often**	**Always**
Personal subscription of printed journal	63.3	17.4	11.9	6.4	0.9
Personal subscription of e-journal	68.8	18.3	10.1	2.8	0
Organizational subscription of printed journal	7.3	21.8	33.6	24.5	12.7
Organizational subscription of e-journal	5.5	1.8	16.5	45.0	31.2
E-archive, repository	13.8	20.2	38.5	23.9	3.7
Interlibrary loan	13.6	36.4	39.1	10.0	0.9
Colleagues	5.5	23.6	47.3	20.9	2.7

As for library services, we can say that researchers are not very enthusiastic about them. Table 5 shows that the use of library among researchers is poor (41.2% use it ocasionnaly, 35.1% never/almost never), as is the use of ILL for acquisition of resources (Table 5 tells us that 13.6 never use it, and 36.4 almost never use it). However, we were rather pleased to notice that some library services are rather well accepted: 53.5% use OPAC often/always; 44.8% often/always start search with OPAC (Tables 4 and 6).

Table 6: Information resources used to start research-related searches

Resource	**Frequency of use (%)**				
	Never	**Almost never**	**Occasionally**	**Often**	**Always**
E-journal sites (eg. Science Direct, Sage,…)	3.2	8.4	27.4	33.7	27.4
Specialized bibliographic databases (eg. Medline, Inspec,…)	9.4	12.5	25.0	33.3	19.8
Information portals, cross-search engines (eg. DiKUL – search portal of UL)	28.1	31.3	24.0	11.5	5.2
Web search engines (eg. Google)	2.1	9.3	11.3	45.4	32.0
COBISS/OPAC (Slovenian union cat.)	4.2	20.8	30.2	31.3	13.5

3.2.2.2. About data curation and use

Researchers were also asked to express their opinions regarding some aspects of data curation (Table 7). They are very much in favour of having the data available (71.7% agree or strongly agree) and are also prepared to provide their own research data for curation and re-use (72.8% agree or strongly agree). But, they have doubts concerning ethical dilemmas: 58.1% agree or strongly agree that it could be questionable to provide data for re-use, since it could lead to theft of results, misuse of data, etc. This is also evident when they express opinions specifically about their own research area; although they mostly think their area would benefit from this, a significant portion also thinks it would be questionable. Interesting is their opinion about libraries playing a role in data cuartion. The first question shows quite strong oposition to this idea, but the second, control question proved that they would also support it.

Table 7: About data curation

Issues of data curation	Frequency (%)				
	I don't agree at all	I don't agree	I agree nor I don't agree	I agree	I agree comple-tely
Data should be available to other researchers	3.3	9.8	15.2	40.2	31.5
I am willing to provide my research data	5.4	9.8	12.0	47.8	25.0
There are ethical dilemmas regarding data curation& use (eg. theft of results, misuse of data)	7.5	14.0	20.4	44.1	14.0
In my area data re-use is questionable	8.0	43.2	21.6	20.5	6.8
My area would benefit from data curation & re-use	0	9.7	21.5	40.9	28.0
Libraries should not be involved in data curation	4.3	10.1	18.8	31.9	34.8
Libraries should be involved in data curation	6.9	12.6	27.6	27.6	25.3

3.2.2.3. Impact of information and communication technologies

Already from the data presented earlier it is evident that ICT has had a profound influence on Slovenian researchers and on their information behaviour (which, of course, does not differentiate them from researchers elsewhere): researchers like electronic materials and use them widely. They are strong users of web search engines (77.4% use them often or always), and websites (39.4% of frequent or regular users). This on the one hand again proves that researchers in some areas behave much like general public. On the other hand, the use of e-journals clearly characterizes them as scholars (61.1% use e-journals often or always). E-preferences are also evident from other data:

- 49.6 prefer to have resources in electronic format (compared to 5% who prefer printed resources),
- 51.3% have over 200 electronic papers in their personal archive,
- 38.1% of researchers cite 81-100% e-resources in their publications.

We also asked researcers how often they print electronic materials when they want to read them: the answer was that 50.0% often and 14.0 always do that.

When asked to estimate which areas of their work were made easier or harder by ICT, results (presented in Table 8) showed that a great majority more easily searches and acquires resources (99%), organizes them (83.5%), chains citations (91.3%), writes (in collaboration (84.8%) seems to be even a little easier than alone (71.9%)) and, of course, communicates (93.9%). Areas which are more difficult for many are relevance judgement (23.7%) and reading (25%). Very few think that ICT doesn't have any influence.

Table 8: ICT makes easier/harder

Activity	Frequency (%)				
	Much easier	Easier	No change	Harder	Much harder
Search & acquisition	8.0	19.1	0	0.9	0
Relevance judgement	22.8	40.4	13.2	19.3	4.4
Organization	48.7	34.8	13.0	3.5	0
Citation chanining	54.8	36.5	6.1	2.6	0
Reading	6.9	28.4	39.7	22.4	2.6
Communicating	61.4	32.5	5.3	0.9	0
Independent writing	26.3	45.6	22.8	2.6	2.6
Collaborative writing	45.5	39.3	12.5	1.8	0.9

3.2.2.4. Some surprising findings

In some areas of information behaviour Slovenian researchers managed to surprise us. The use of Web 2.0 tools for research purposes is almost non-existing: social networks are never or almost never used by 84.8%, weblogs are obviously not considered an information source – as 82.8% don't use them, and, similarly, web forums are never or almost never used by 64.6% of respondents (Table 4). Also, it seems that quite a significant proportion of researchers (20%) don't use (and cite) electronic resources very often when they write (Table 9).

When we look at other information resources (Table 4), we see that half never or almost never use preprints and a good third similarly never or almost never uses email alerts. Cross-search services and specialized portals (Table 5) seem to be perceived as equally unimportant (60.2% never or almost never start their search using those). Interestingly, also open-access materials don't seem to be very popular, as data shows that they are not used to a great extent – open-access materials comprise less than 20% of citations for 58.3% of researchers (Table 9). All these findings differ from the usual findings for contemporary scientists and would be worth looking into.

Table 9: Shares in citations[4]

Shares	Frequency (%)				
	Below 20%	21-40%	41-60%	61-80%	81-100%
Share of scientific publications	3.7	2.8	10.1	20.2	53.8
Share of open-access publications	58.3	15.6	13.5	6.3	6.3
Share of electronic sources	20.0	14.3	7.6	20.0	38.1

3.2.2.5. A closer look: Information behaviour in relation to demographic variables

We performed bi-variate statistics (Chi-Square tests) to investigate some connections between variables. A link can be identified between age of respondents and their perceptions of various aspects of digital tools, formats, ways of communicating and acquiring information, younger reserachers tend to prefer digital environment.

Gender of respondents is not an issue which would influence preference or behaviour. This is, of course, not surprising, since gender has not been identified as a factor influencing scholarly information behaviour. However, we did identify research discipline as being influential in some respects:

- Natural Sciences:
  - are keen to use research papers, dissertations,
  - don't acquire their resources by print journals or interlibrary loan,
  - use raw data,
  - cite higher proportion of scientific literature.
- Social Sciences:
  - cite higher proportion of scientific literature,
  - don't think libraries should undergo data curation.
- Technical Sciences:
  - don't use raw data,
  - use standards, patents,
  - are in favour of libraries doing data curation.
- Humanistic Sciences:
  - use research papers, dissertations,
  - prefer printed sources,
  - acquire their sources by using print journals,
  - cite higher proportion of scientific literature, lower share of e-sources and lower share of open-source materials,
  - are not keen on using ICT to organize resources.

- Interdisciplinary Research:
  - use e-archives to acquire their resources
  - are in favour of data curation from various aspects: that data should be available to others, that tey would provide their own data and that libraries should deal with data curation
- Medicine:
  - use websites,
  - use colleagues to acquire their resources and information (invisible college),
  - use raw data, but are sceptical in terms of ethical dilemmas of its use,
  - think that ICT helps independent writing.
- Biotechnology:
  - use research papers, dissertations,
  - prefer e-sources,
  - support availability of raw data,
  - are keen on using ICT to communicate and organize resources.

Other influential demographic variables seem to be employment status, experience in teaching and researching, and current job tasks.

4. Discussion

Our results in this preliminary analysis clearly show that Slovenian researchers in most areas don't differ significantly from what has been discovered for today's scientists. There are some areas which show some discrepancies – primarily the use of Web 2.0 tools for research purposes. While it was clearly shown that researchers strongly lean on personal contacts (i.e. social networks), this has for some reason not transferred to digital environment. Digital scholarship has in recent years been gaining importance in the professional lives of scientists, both for research and for teaching (see for example Veltsianos and Kimmons (2012), Researches of tomorrow (2011), Moran, Seaman and Tinti-Kane (2011)). Non-use of weblogs and digital social networking tools can to a large extent be contributed[5] to work overload (most researchers also teach), lack of motivation and maybe also as a reaction to popular use of social netwook tools in today Slovenia, not so much to technological issues. Last political elections that taken place and used social network tools as a political parties promotion tools. Some stated that currently they are 800.000 facebooks in Slovenia, which is a lot for two milion of Slovenians, as 60% of Slovenians have their profiles (63% are females) (RIS, 2011). Slovenia is, in fact, one of the most typical European Union (EU) countries with respect to information technology usage (Petrič et al. 2011). Some current research imply that it appears that many users are mainly using Facebook to partake in passive activities, instead of providing active social contributions (Ryan, Xenos, 2011). Such findings suggest that not all Facebook users are using the site to improve their social capital, which can also some impact on nonuse.

Another issue which has obviously not gained a lot of researchers' attention is open-access. We plan to look into these issues more deeply, to be able to shed more light on our results. One of the reasons might also be the traditionally well-organised access to academic journals in Slovenia (Juznic, 2009).

Otherwise there is no doubt that information and communication technologies have a big impact on the life and work of scientists, regarding many issues: resource formats, access, means for information exchange, organization of resources, writing, reading, etc. We also saw that, regardlessof the quality of infrastructure, researchers are, as always, independent, innovative and creative to find ways to acquire information which they need and to use it appropriately.

In some respects Slovenian researchers are (much like researchers in general) quite similar to general public – perhaps more than they would like (or more than providers of formal information sources would like). This is proven by e.g.:

- Intensive use of web search engines, which is much more common than the use of 'formal resources', such as e-journal providers or specialized bibliographic databases.
- Preference of e-materials and tools, immediate access to full-text,
- Dissatisfaction with existing library services and non-use of these services (with exception of OPAC).

There are, of course, areas which clearly distinguish scientists from general public. One of those is undoubtedly reference judgement. As proven also by our results, scientists are (for obvious reasons) much more than average users concerned with judging the content and quality of their information sources.

As for grey literature, we saw that it is quite intensive, but depending on academic area and sector of employment:

- Technical, natural, humanistic and biotechnological sciences are more inclined to use grey literature as the source of information for their research.
- Business sector is of all sectors the most concerned with patents and standards.

When we asked researchers what they think about data curation, the answers were positive, but careful: in principle they are in favour of it, but many have dilemmas regarding ethics. It is also quite obvious that they (with exception of medicine) don't use these resources.

5. Conclusion

Our research has shed more light on information behaviour, and to a certain extent, also information needs of Slovenian scientists. When the study, which is partly presented here, is finished, we hope that the picture will be even more complete. Some interesting issues arose which we intend to investigate further.

What implications could our findings have for information services to scientists? Libraries will need to further rethink the services which they offer to scientists. There is no doubt that formal resources will still retain their value and importance, however, they will be shaped by information and communication technologies and new patterns of scientific publishing (e.g. open access, e-archives, repositories). Some library services are already less important to scientists – hence the decline of library visits – while others, like OPACs (with access to full-text) and provision of access to e-journals, retain their value. There are other interesting issues which libraries might consider taking up: setting up e-archives and repositories, approach data curation.

Undoubtedly formal information tools need further development in terms of more intuitive interface design, change of record structure - and that is potentially another big task for libraries. For this, of course, wider co-operation is needed, e.g. with search engine developers, publishers.

We must mention that our results were obtained with web survey and are thus a reflection of self-reported activities and decisions of scholars. Due to the limitations of this method the results should be crosschecked using some other methodology. For example, observation of scholars at work would probably give us more detailed insight into their activities, and follow-up interviews would certainly shed more light into the context of their work, on their decisions and on reasons for their actions. However, it would be very difficult, if not impossible, to assemble so many cases as we did using our current methodology.

6. References

Bates, M.J. (1989), “The design of browsing and berrypicking techniques for the online search interface”, Online Review, Vol. 13 No. 5, pp. 407-24.

Bates, M.J. (2007), “What is browsing—really? A model drawing from behavioural science research”, Information Research, Vol. 12 No. 4, available at: http://InformationR.net/ir/12-4/paper330.html (accessed October 10, 2011).

Borgman, C.L. (2009) "Scholarship in the Digital Age: Blurring the Boundaries between the Sciences and the Humanities (Keynote Talk)" Digital Humanities Conference. College Park, MD. Jun. 2009. Available at: http://works.bepress.com/borgman/216 (accessed October 10, 2011)

Chowdhurry, S. (2010). "Data curation: an ecological perspective", College & Research Libraries News, vol. 71 no. 4 (April 2010), pp. 194-196.

Dallmeier-Tiessen, S., Goerner, B., Darby, R., Hyppoelae, J., Igo-Kemenes, P., Kahn, D., Lambert, S., Lengenfelder, A., Leonard, C., Mele, S., Polydoratou, P., Ross, D., Ruiz-Perez, S., Schimmer, R., Swaisland, M., van der Stelt, W. (2010). "Open Access Publishing - Models and Attributes", The SOAP Consortium, 2010. Available at http://edoc.mpg.de/get.epl?fid=71514&did=478647&ver=0 (accessed October 10, 2011)

Gentil-Beccot, A. (2010). »The driving and evolving Role of Grey Literature in High-Energy Physics.« In D. J. Farace & J. Schöpfel (eds.). Grey Literature in Library and Information Studies. De Gruyter Saur, Berlin/New York

Gold, A. (2010). "Data Curation and Libraries: Short-Term Developments, Long-Term Prospects", available at http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1027&context=lib_dean (accessed November 11, 2011)

Information Behaviour (2008), “Information behaviour of the researcher of the future – Executive summary”, available at: www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf (accessed October 10, 2011).

Južnič, P. (2009). “Using a decision grid process to evaluate library collections and e-journals. *New libr. world*, 110 (7/8), 341-356.

Lord, P., Macdonald, A. (2003). "E-Science Curation Report: Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision", available at http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf (accessed 22 October 2011)

Lord, P., Macdonald, A., Lyon, L., Giaretta, D. (2004). "From Data Deluge to Data Curation", available at http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf (accessed October 22, 2011)

Maron, N. and Smith, K.K. (2008). "Current Models of Digital Scholarly Communication: Results of an Investigation Conducted by Ithaka Strategic Services for the Association of Research Libraries", available at: http://www.arl.org/bm~doc/current-models-report.pdf (accessed October 10, 2011).

Muench, V. (2011). "Open Access shaking the basics of academic publishing", Online, Jul/Aug 2011, pp. 18-21.

Moran, M., Seaman, J., & Tinti-Kane, H. (2011). "Teaching, learning, and sharing: How today's higher education faculty use social media for work and for play", available at http://www.pearsonlearningsolutions.com/blog/2011/05/09/teaching-learning-and-sharing-how-todays-higher-education-facutlyuse-social-media/ (accessed December 22 2011).

Palmer, C.L., Teffeau, L.C. and Pirmann, C.M. (2009). "Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development", available at: www.oclc.org/research/publications/library/2009/2009-02.pdf (accessed October 10, 2011).

Pečlin, S. Južnič, P. (2012). »Research in the fields of medicine in Slovenia – research potentials, funding and publications." in press.

Petrič, G., Petrovčič, A., Vehovar, V., (2011) Social uses of interpersonal communication technologies in a complex media environment. European Journal of Communication. 26(2) 116–132

Researchers of Tomorrow: A three year (BL/JISC) study tracking the research behaviour of 'Generation Y' doctoral students: Second Annual Report 2010-2011. May 2011. Available at :http://www.efc.co.uk/projects/rot/Researchers%20of%20Tomorrow%20Year%202%20report%20final%20110510.pdf (accessed December 22, 2011).

Ryan, T., Xenos, S. (2011), Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. Computers in Human Behavior 27 (2011) 1658–1664

Rowlands, I. and Fieldhouse, M. (2007), "Information behaviour of the researcher of the future", Work Package I: Trends in Scholarly Information Behaviour, available at: www.jisc.ac.uk/media/documents/programmes/reppres/ggworkpackagei.pdf (accessed October 10, 2011).

Schöpfel, J., Stock, C. (2009). Grey literature in French digital repositories: a survey. In GL10 Conference Proceedings. Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society. Amsterdam. Available at: http://archivesic.ccsd.cnrs.fr/docs/00/37/92/32/PDF/GL10_Schopfel_Stock_final.pdf (accessed October 10, 2011).

Veletsianos, G. and Kimmons, R. (2012). "Networked participatory Scholarship: emergent techno-cultural pressures toward open and digital scholarship in online networks", Computers & Education, vol. 58, pp. 766-774.

Williams, P. and Rowlands, I. (2007), "Information behaviour of the researcher of the future". Work Package II: The Literature on Young People and their Information Behaviour, available at: www.jisc.ac.uk/media/documents/programmes/reppres/ggworkpackageii.pdf (accessed October 10, 2011).

Acknowledgment and funding

This research was conducted as a part of the project V5-1016 funded by the Slovenian Research Agency.

[1] Active means that in the current year 2011 the researcher has been allocated at least 100 publicly funded research hours.
[2] According to the classification of Slovenian research agency (ARRS): 1 – Natural Sciences,, 2 – Technical Sciences, 3 – Medicine, 4 – Biotechnical Science, 5 – Social Science, 6 – Humanistic Sciences, 7 – Interdisciplinnary Research.
[3] As explained in the introduction ,we used the classification of the Slovenian Research Agency (ARRS).
[4] With citations we mean resources which are cited by researchers in their own publications.
[5] At this point we can only provide anecdotical evidence, personal experience and results of ad-hoc interviews with individual scientists from various areas – but we intend to look into these issues more deeply in the future.

GeoStoryteller:
Taking grey literature to the streets of New York

Debbie Rabina and Anthony Cocciolo,
Pratt Institute, School of Information and Library Science (SILS), United States

Abstract
In this paper, we briefly describe the design and implementation of the GeoStoryteller project, with particular emphasis on the resources used to create the digital narratives that are the core of the application. GeoStoryteller is a project where learners engage with archival photos and multimedia narratives in historically relevant places using a combination of augmented reality technology and web-based delivery via mobile devices. The initial application of GeoStoryteller is in partnership with the Goethe-Institut New York, the worldwide cultural organization of Germany. In this partnership, learners—particularly German language students in the United States—engage with content that details the historical events and makes use of real places related to German immigration to New York City (1840-1945). Second, based on the design and implementation of this project, we offer a framework for creating location-based mobile learning projects that could be used by others interested in implementing similar projects. Lastly, we profile the sources used to create the digital narratives, the majority of which (76%) were grey literature materials.

Designing GeoStoryteller

GeoStoryteller builds upon recent research and development from the emerging subfield of the Digital Humanities known as GeoHumanities, which is a term that highlights the growing interconnections between geography and the humanities. GeoStoryteller strengthens this connection specifically by layering historic narratives on a specific location and by making it available to learners' Internet-enabled mobile device. In the case of the application of GeoStoryteller with the Goethe-Institut—named German Traces NYC—learners can view location-sensitive maps and lists of historic sites related to German immigration to New York. Once a learner arrives at a physical site, he or she can see historic photos layered against the imagery visible through the camera's phone (also known as augmented reality), as well as watch videos about that particular site that include historic narratives and archival photos. Users can also play short trivia games that answers can only be found from being on physical location and post their accomplishments to Facebook or Twitter.

For developing the videos—which we also refer to as GeoStories—we turned to published, non-published, digital and non-digital sources available from cultural and memory institutions (libraries, archives, museums and historical societies). Using these sources we developed historical narratives as text, and set those to audio recordings that were augmented with archival photos.

We then used geotechnologies to deliver the narratives to users' mobile devices (smart phones, such as iPhone, Android, Blackberry, and tablets such as iPad) at the places where these events occurred. Geotechnologies used include global positioning systems (GPS), digital mapping services (available through Google) and Layar, an augmented-reality browser for mobile devices.

For example, users can find such sites as the Ottendorfer Library, the oldest public library in Manhattan, opened originally to support the German immigrant community in Kleindeutschland—or Little Germany (today known as the East Village; see Figure 1). While on site, learners can learn about the people who crated the physical site, why they created the site (e.g., philosophies that motivated their action) and how that site has changed over time to reflect cultural and world events (e.g., anti-German sentiment during both world wars).

Figure 1. Augmented-reality interface for retrieving multimedia stories.

Framework for creating location-based mobile learning projects

From our experience developing GeoStoryteller, we offer a framework that can be used in the planning and implementation of similar projects.

At the core (the inner section) of this framework is traditional humanities research and development (see Figure 2). In the case of the GeoStoryteller project, this included print, non-print, digital, digitized, non-digital, archival, and published sources, all used to construct historical narratives of German immigration in New York. Placing humanities research and development at the core of the framework highlights what distinguishes a digital humanities project from other digital projects. Additionally, this type of research and development could be done with other fields of the humanities, including philosophy, literature, and the arts.

Next, we proceed to the "Theory and Interface Development" ring. At this stage, researchers must consider the theory that will inform their project, and how this theory will be reflected in the user interface available to the learner. For example, some of the theoretical questions we asked were, how does situating historical content in physically relevant locations affect learner engagement? And does making augmented-reality content available to learners further enhance engagement? Such questions necessarily lead to decisions about how to present the content within a digital interface.

In the third ring, learners engage with the socio-technical environment created by the researchers. In the case of GeoStoryteller, this includes not simply a user interface, but also the physically relevant locations that the interface prompts the user to explore. Additionally, social interactions may occur during this stage among multiple others in the environment (e.g. a librarian in the Ottendorfer library). It cannot be assumed that the best learning experience comes from the digital device, but could result from the serendipitous interaction in the real environment.

From this stage, we proceed to the fourth ring, which is the formal user research. In this stage, we use traditional (e.g., surveys and interviews) and digital (e.g., tags, hits, time stamps) social science research methods to uncover the working of the interface, address the theoretical questions, and evaluate the learning outcomes and user engagement. These studies can then influence any earlier stages (e.g., specific areas of content that did not engage the interest of learners could be revised, confirming or refuting the learning theory, or requiring changes to the user interface).

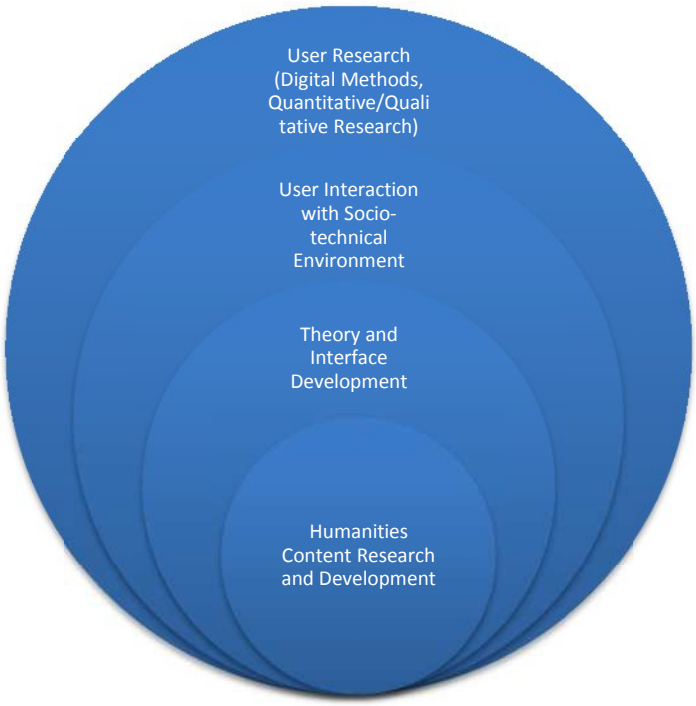


Figure 2: A Process for Digital Humanities Research and Development

Source materials

Sources of evidence used in humanities research are mostly records that are found in newspapers, photographs, articles, oral histories or other "records of human experience" (Borgman, 2009, paragraph 33). These sources come from a multitude of institutions, in multiple formats, and with many—and often unclear—usage rights and policies. Records used in humanities research are often reinterpreted as new contexts emerge. The multiplicity of tools used to create the GeoStories was quite wide. We were informed by our knowledge of local history and bibliographic sources and searched collection we anticipated would have materials of relevance, such as public libraries, historical societies, and local press dating back to the mid 19th century. In addition, we searched for other period sources such as travel guides, restaurant menus and store catalogs. More targeted searches included hobbyists websites and business and personal archives.

GermanTracesNYC includes a total of forty GeoStories. One of the guiding principles for the researchers was to include only visual materials that are in the public domain, licensed under Creative Commons licenses, or were given to the researchers with permission to use by the holders of the original copyrighted materials. This is keeping in the spirit of GeoStoryteller, which is a Creative Commons open-source platform, and within the contractual agreement between Pratt Institute and the Goethe Institut, which specified that GermanTracesNYC will be maintained in the public domain. On average, twenty hours of research was required to assemble sources materials for each GeoStory. Source materials for creating both the narratives and the accompanying slideshows included monographs, newspapers articles, photographs and other visual materials (illustrations, drawings, etc.). Most of the materials were digitized by public institutions such as the New York Public library and are in the public domain. Other materials, digitized by commercial interests such as Google Books, were used for visual materials only if they were in the public domain, and otherwise were used as bibliographic sources. In all, approximately 400 unique items were used in the project, an average of 10 items per GeoStory, although some sources were used more then once. Figure 3 provides a snapshot of the number and types of sources used to create the first five GeoStories. The total number of sources used for each story was between 12 to 22, with an average of 15.6 per story.

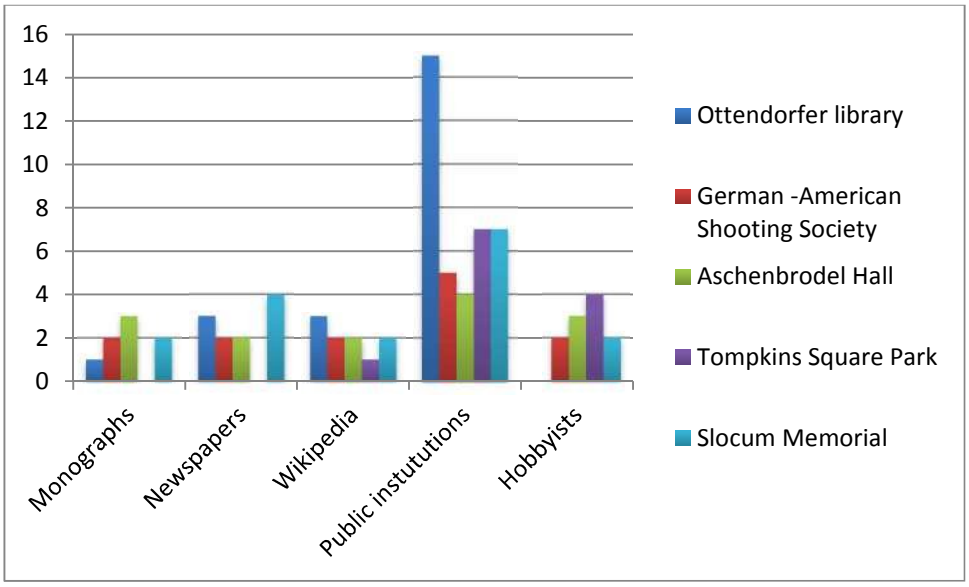


Figure 3: Sources used in five stories

Public institutions provided many of the sources, particularly the visual sources, used in creating the GeoStories. Institutions such as the New York Public Libraries or the Library of Congress have made enormous efforts in recent years to digitize their historic collection and make the digitized items available to the public. The public institutions used most frequently in the project were The New York Public Library, the Museum of the City of New York, the Library of Congress and the New York City Landmark Preservation Commission. Figure 4 provides a breakdown of the public institutions used in five GeoStories.

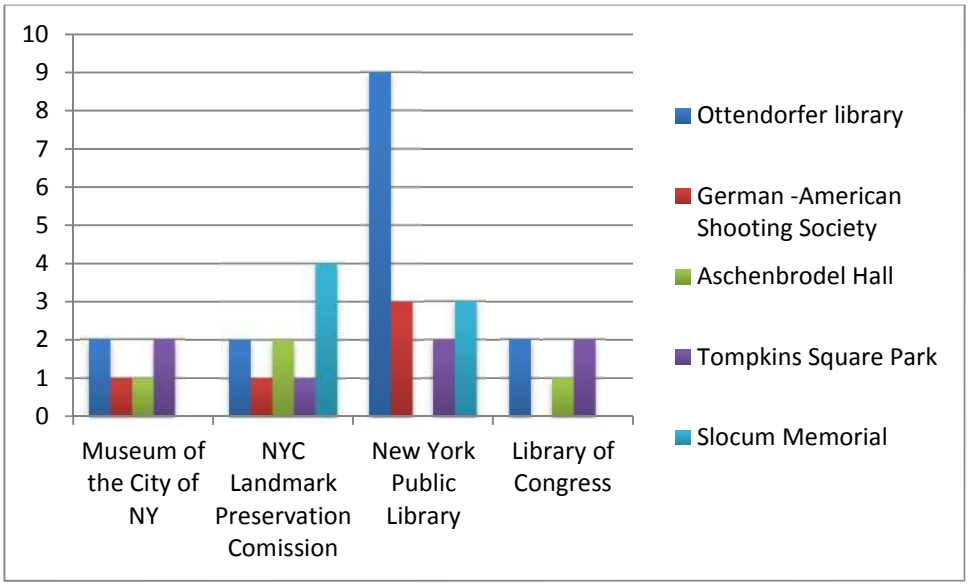


Figure 4: Public institutions used in five stories

It order to determine the how significantly grey sources featured in the materials used to create the GeoStories, one must first return to the definition of grey literature. Schöpfel (2011) provides an overview of the changes that occurred over time to the definition of grey literature but acknowledges that the definition that prevails in the one formalized during the 3rd International Conference on Grey Literature in 1997, and included in the *Encyclopedia of Library and Information Sciences* (Schöpfel and Farace 2010). According to this definition, grey literature is "that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers". Working with this definition we established a series of questions that served as a checklist in determining whether a source should be considered grey or white literature.

1. Who is the original publisher of the source?
2. What was the original source of publication?
3. Who digitized the source?

4. Where is the source currently held?
5. Is the source discoverable in traditional findings aids (catalogs or indices)?

Using this checklist, we established that the majority of sources used as source materials for the GeoStories, meet the definition or grey literature. Figure 5 depicts the use of sources for five stories. All stories used significantly more grey materials then other types and one story used only grey literature.

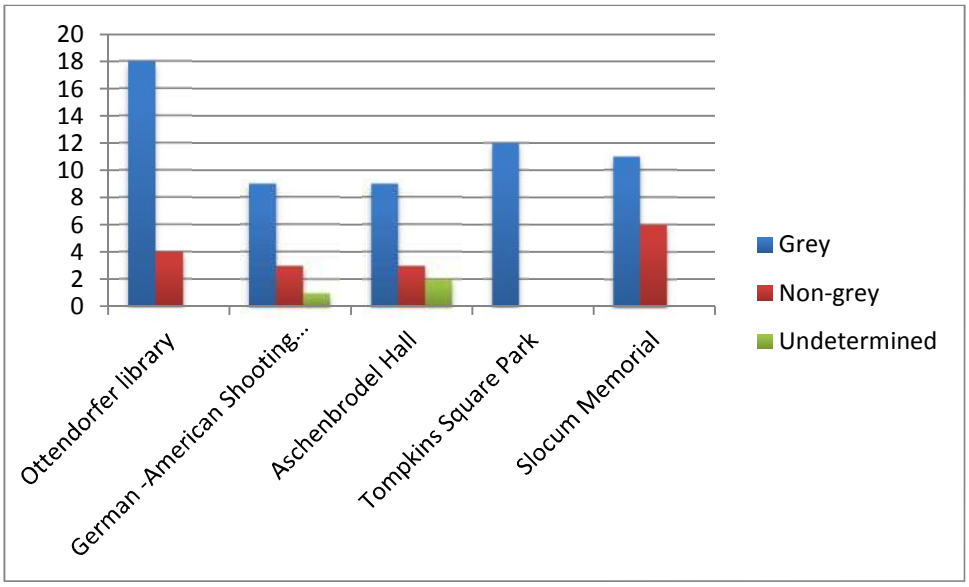


Figure 5: Grey literature used in five stories

When looking at the combined sources used to create the five stories, we see the 76% of the sources used were grey literature, 21% were white literature and 4% were undetermined. While the list of questions we developed was helpful is determining whether a source was grey or not, arguments can be made to the contrary, depending on how broadly or narrowly one defines discovery tools. Articles from the New York Times are considered to be white literature since the New York Times in available through several indexes, including Gale U.S. History In Context, InfoTrac Custom Newspapers, ProQuest Historical newspapers. Monographic materials discoverable through library catalogs such as WorldCat or the New York Public Library catalog, such as *King's Handbook of New York City* (King 1892) were considered white, but other monographic materials, such as the 1919 edition of *Childhood's Favorites and Fairy Stories,* available from Project Gutenberg, presented more of a challenge and ultimately ended up in the 'undetermined' column. The overwhelming majority, 76%, were identified as grey literature. These included items such as the 1874 illustration digitized by the Library of Congress (Library of Congress 1874).

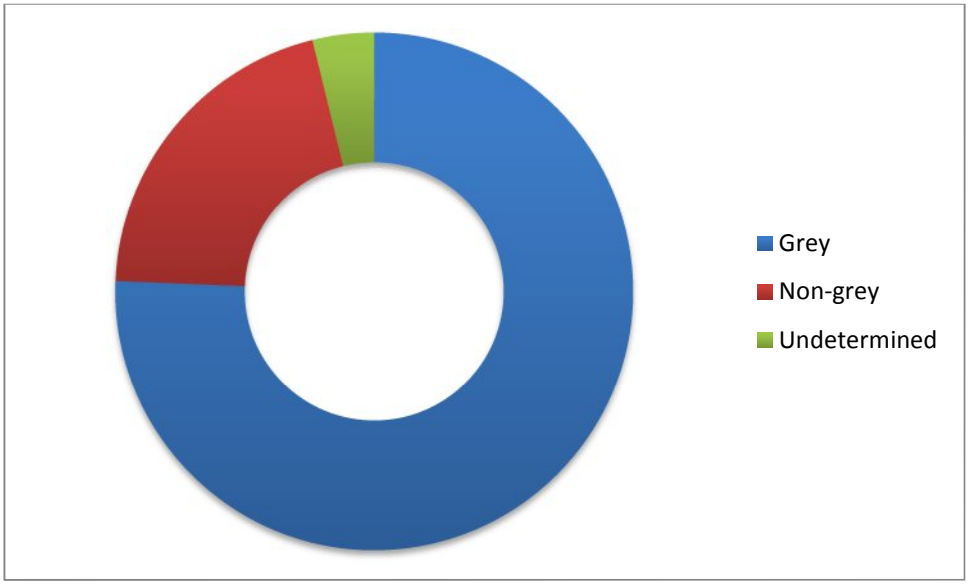


Figure 6: Overall use of grey literature in five stories

Conclusion

A number of trends made possible by the Internet converged to allow the creation of the GeoStories of GermanTracesNYC. First, digitization of legacy collections in libraries and cultural institutions provided a substantial corpus of sources that the researchers could use. Second, the ability to bypass traditional finding tools such as catalogs and indices to search directly in the collections of choice and through the aid of search engines, make the digitized collections discoverable. Lastly, the available of materials that can be used without permission, either through Creative Commons licenses or from the public domain. Grey literature constitutes the majority of digitized materials that were used in the project and supported the core of the model (figure 1) used for the development of this project.

Bibliography

Borgman, C.L. (2009), "The digital future is now: A call to action for the humanities," *DHQ: Digital Humanities Quarterly,* Vol. 3 No. 4, available at: http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html# (accessed 25 October 2011).

Childhood's Favorites and Fairy Stories (1919), p. 35. http://www.gutenberg.org/files/19993/19993-h/19993-h.htm#page35

King, Moses. (1892). *King's handbook of New York city: An outline history and description of the American metropolis*. Moses King. Reprint: Barnes and Noble 2001.

Library of Congress. The red flag in New York - riotous Communist workingmen driven from Tompkins Square by the mounted police, Tuesday, January 13th / Matt Morgen(?). (1874). http://hdl.loc.gov/loc.pnp/cph.3c11180

Schöpfel, Joachim (2011). Towards a Prague definition of grey literature. Twelfth International Conference on Grey Literature ; (2010-12-06 / 2010-12-07 ; ; Praha (CZ) ;)

Schöpfel, Joachim, Farace, Dominic J. (2010). 'Grey Literature'. In </J/ Bates & M.N. Maack (eds.), *Encyclopedia of Library and Information Sciences,* Third Edition, pp. 2029-2039. CRC Press.

Acquisition and distribution of technical reports and conference proceedings on science and technology in Korea

Seon-Hee Lee, Soon-Young Kim, Beom-Jong You, and Heeyoon Choi
Korea Institute of Science and Technology Information

Abstract
According to a survey in 2006 conducted by Korea Institute of Science and Technology Information (KISTI), technical reports and conference proceedings are the most important types of grey literature for researchers in Korea. Important, recent, and detailed research results are stored in technical reports and are presented at conferences. Despite their importance, these types of grey literature are associated with difficulty when seeking to access. As a national science and technology information center, KISTI has made an effort regarding the acquisition and distribution of technical reports and conference proceedings on science and technology to promote easy access. KISTI has transformed grey literature into white literature to support R&D researchers.
This paper examines the acquisition and distribution status and analyzes the usage statistics pertaining to technical reports and conference proceedings in Korea. KISTI has acquired and distributed domestic and foreign technical reports and conference proceedings for several decades. KISTI provides metadata and the full-text of R&D technical reports produced by government funded projects and conference proceedings through National Discovery for Science Leaders (NDSL). For an analysis of the usage statistics for technical reports and conference proceedings, the frequency of full-text downloads and the use of Document Delivery Service (DDS) are used. The KISTI statistics system, termed the Online Analytical Processing (OLAP) system, shows that conference proceedings and technical reports were used 1,921,150 times and 1,140,411 times, respectively from September of 2009 to September of 2011. These statistics indicate that open access for conference proceedings and technical reports experiencing heavy use by researchers in Korea. KISTI is using new tools and communication skills to distribute more grey literature in effort to assist R&D researchers. New ways of distributing grey literature through Social Networking Services such as Twitter, Facebook are even available on smartphones for users in Korea in the smart age.

Keywords *Grey Literature, Conference Proceedings, Technical Report, Social Network, Usage of Grey Literature*

Introduction

Grey literature such as technical reports and conference proceedings contains detailed research results as a resource for scientific and technical information. Moreover, this type of literature is published frequently than journals. Therefore, grey literature is considered to be an important information resource, but locating and accessing grey literature is not easy. In fact, it is often overlooked by information centers and libraries. It is important to create access points and provide grey literature to researchers for the improvement of research and development (R&D) in Science and Technology. As a national science and technology information center, Korea Institute of Science and Technology Information (KISTI) took acquisition and distribution steps related to technical reports and conference proceedings on science and technology. KISTI has thus transformed grey literature into white literature to support R&D researchers. This paper examines the acquisition and distribution status and analyzes the usage statistics for technical reports from government agencies and scientific research groups and conference proceedings from academic societies in Korea. In addition, the new methods distributing grey literature through the internet or via smartphones will be mentioned. This offers directions regarding the creation of access point for grey literature according to the new information environment. Although much effort has been made and while these sources of grey literature are produced in great numbers in an environment containing more advanced means of communication technology, getting access to this type of literature remains challenging.

Objective and Scope of the Study

An objective of this study is to understand users need of grey literature and the status of acquisition and distribution of the grey literature in Korea. Scope of this study is focused on acquisition and distribution of technical reports and conference proceeding through National Discovery Science Leaders (NDSL) in Korea.

Method of the Study

Literature review for grey literature is conducted for understanding domestic and overseas trends. This paper examines the acquisition and distribution status and analyzes the usage statistics pertaining to technical reports and conference proceedings in Korea. For an analysis of the usage statistics for technical reports and conference proceedings, the frequency of full-text downloads and the use of Document Delivery Service (DDS) are used. This study use data from KISTI statistics system, the Online Analytical Processing (OLAP) system. For new access points for grey literature in Korea, actual Social Networking Service in NDSL is invested.

Literature Review

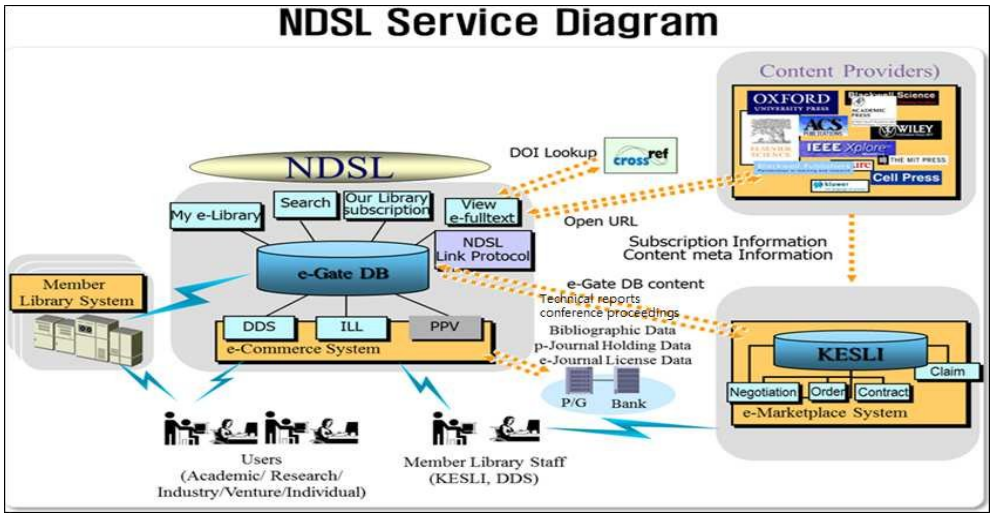
Choi and Cho (1997) categorized grey literature into light, medium, and dark grey literature and studied grey literature production, related databases, and online access through the internet. Debachere (1995) reported that grey literature is frequently original and usually recent. Yoo (2009) noted that conference proceedings are simpler to obtain in libraries than other types of grey literature, such as reprints, because this type has a greater possibility for publication. Most libraries and information centers in Korea subscribe to foreign digital conference proceedings through a purchasing consortium called KESLI (Korea Electronic Site License Initiative) because this initiative allows scholars to obtaining information quickly and reduces the necessary effort by librarians' regarding the collecting and cataloging of this information resources. Nam and Cho (2006) studied the importance grey literature for scholars by analyzing references in the Korean journals listed in Science Citation Index (SCI). According to their study, grey literature represents about 24% in all references types used. KISTI studied the development of an acquisition and distribution model for the foreign grey literature in 2002 and 2006. Jeffery and Asserson (2005) described a systematic process for grey literature and suggested a R&D process for its management. McDermott (1995) mentioned the importance of national level bibliographic control for grey literature. Gelfand (2000) emphasized an international cooperation network for this type of literature.

Survey for Grey Literature (GL)

Librarian and information specialists in 38 science and technology institutes participated in a survey conducted by Korea Institute of Science and Technology Information (KISTI) in 2006. According to this survey results, technical reports and conference proceedings are the most useful types of grey literature for researchers in Korea because most research results are stored in technical reports and presented at conferences. An analysis of result of this survey shows the difficulties in each field for grey literature are as follows: acquisition 36.5%, identifying & accessing 27.3%, bibliographic control 27.3%, and managing 89.5%. The most desired format for grey literature is an electronic format (the internet) at 74.5%. Electronic version is more popular than printed version. The most desired format for records was full-text at 57.7% and metadata & abstracts at 32.7%.

Overview of Grey Literature in KISTI in Korea

KISTI's has collected grey literature in print or micro-film format and distributed it through Document Delivery Service (DDS) since 1962. However, users have wanted electronic versions of grey literature since the internet age began. This tendency was detected in the survey result. Based on the survey result, KISTI tried to create more access points for technical reports and conference proceedings on a portal system, termed the National Discovery for Science Leaders (NDSL) (Fig. 1). NDSL contains articles, patents, conference proceedings, technical reports, trend information and analysis reports, standards and other types of grey literature. Electronic and printed journals and conference proceedings can be searched by R&D researchers to construct comprehensive metadata. E-conference proceedings such as the IEEE/IEL electronic library proceedings are subscribed by Korean Electronic Site License Initiative (KESLI) consortium members for joint purchasing



<Fig. 1> NDSL (National Discovery for Science Leaders)

Acquisition and Distribution of ST GL on the NDSL System

■Conference Proceedings

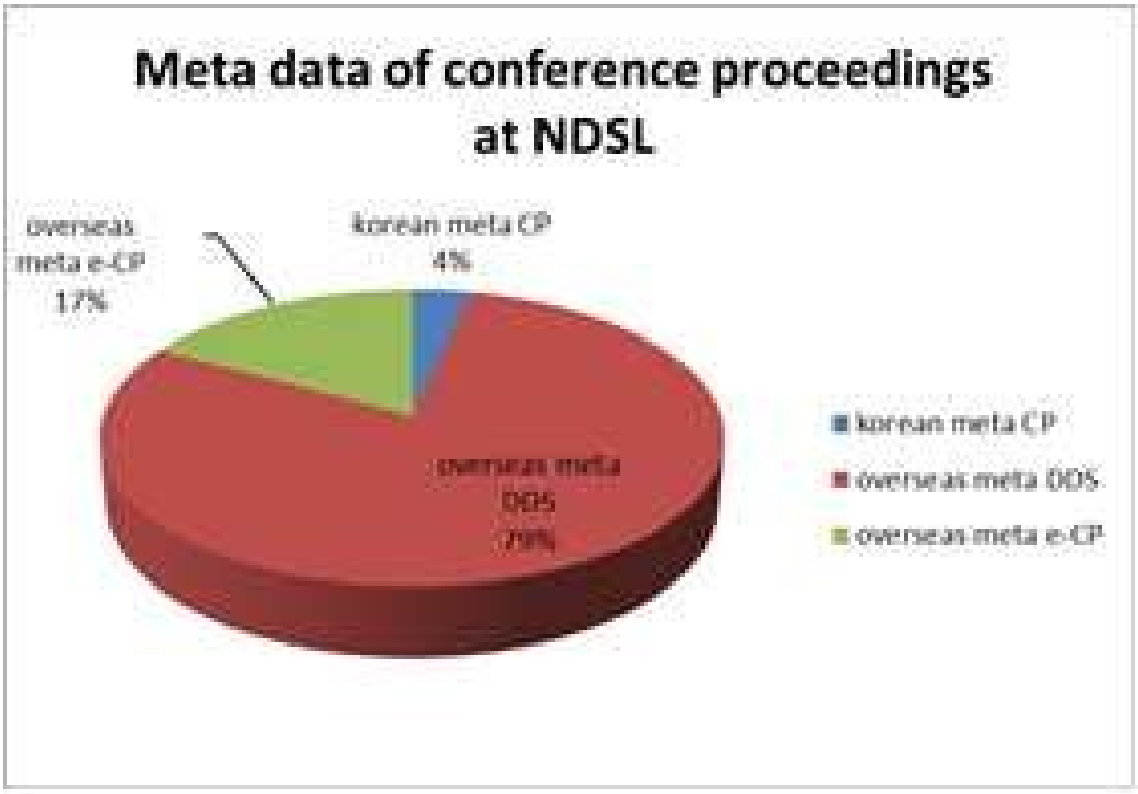
KISTI holds 16,051 conferences proceedings in paper format. Distribution of science and technology conference proceedings on NDSL is focused on making access points than providing full-text, so metadata for domestic and overseas conference proceedings was built. Table 1 shows that NDSL has 6,913,399 records of metadata for overseas print conference proceedings. Users can identify and locate the print conference proceedings and ask document delivery for them. NDSL hold 1,479,345 metadata of overseas e-conference proceedings and 304,935 metadata of Korean e-conference proceedings metadata. If users' organization subscribes the e-conference proceedings and one of Korea Electronic Site License Initiative (KESLI) members, users can access full-text of conference proceedings through metadata on NDSL. Even though KISTI does not provide full-text for conference proceedings, users can identify, locate and access them.

<Table 1> Metadata of ST Conference Proceedings

format	E-Conference Proceedings		Print Conference Proceedings*
matadata	Metadata of Korean CP	Metadata of Overseas CP	Metadata of Overseas CP for DDS
No. of metadata	304,935	1,479,345	6,913,399

*KISTI print CP holdings: 16,051 titles

The metadata of overseas conference proceedings for document delivery accounted for 79% of all metadata from Fig. 2. Korean conference proceedings can also be accessed through the ST Academic Society Village (http://society.kisti.re.kr) constructed by KISTI.

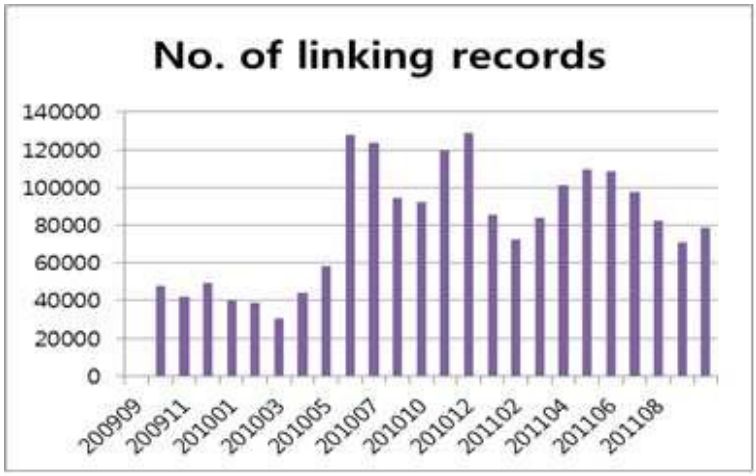


<Fig. 2> Metadata of Conference Proceedings at NDSL

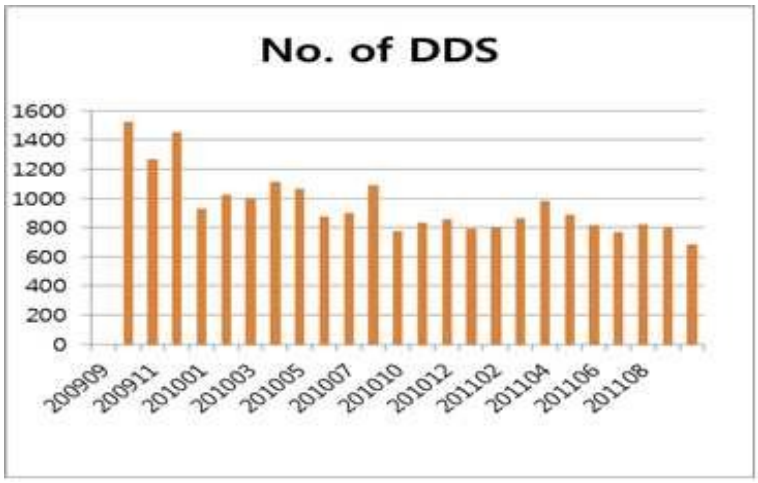
KISTI analyses the usage statistics from the Online Analytical Processing (OLAP), KISTI's statistics system. In Table 2, the numbers of uses for conference proceedings linking to full-text shows 1,921,150 records. If users' institutes did not subscribe to the proceedings, the users cannot access full-text versions of e-conference proceedings. In such case, users can use Document Delivery Service for a document. The numbers of Document Delivery Service requests for conference proceedings through the NDSL were 22,520 records from September of 2009 to August of 2011. There were more usage requests for e-conference proceedings that involved linking to a full-text version than that of the number of requests for the paper Document Delivery Service. The usage numbers for the paper the document delivery is slowly declining.

<Table 2> Usages Types of Full-Text Linking vs. DDS at NDSL (2009/9-2011/8)

type of use	full-text linking	DDS	Total
No. of records	1,921,150	22,520	1,943,670



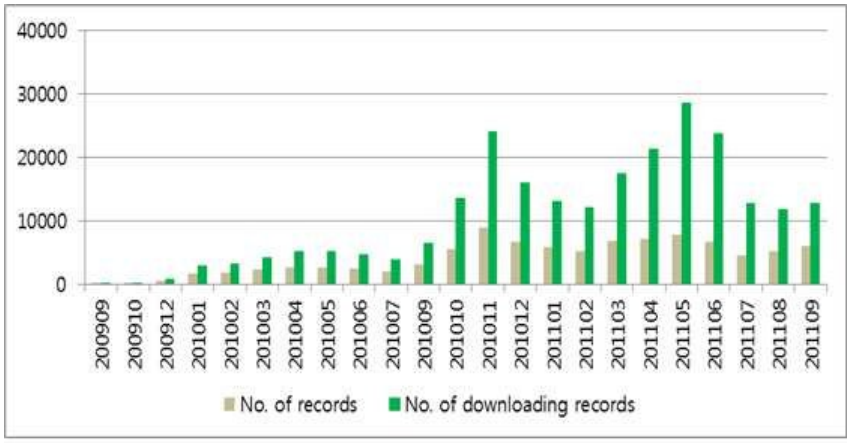
<Fig. 3> Usages Requests for Linking to Full-Text Versions for CP through NDSL Metadata



<Fig. 4> Usages Request for DDS for CP through NDSL Metadata

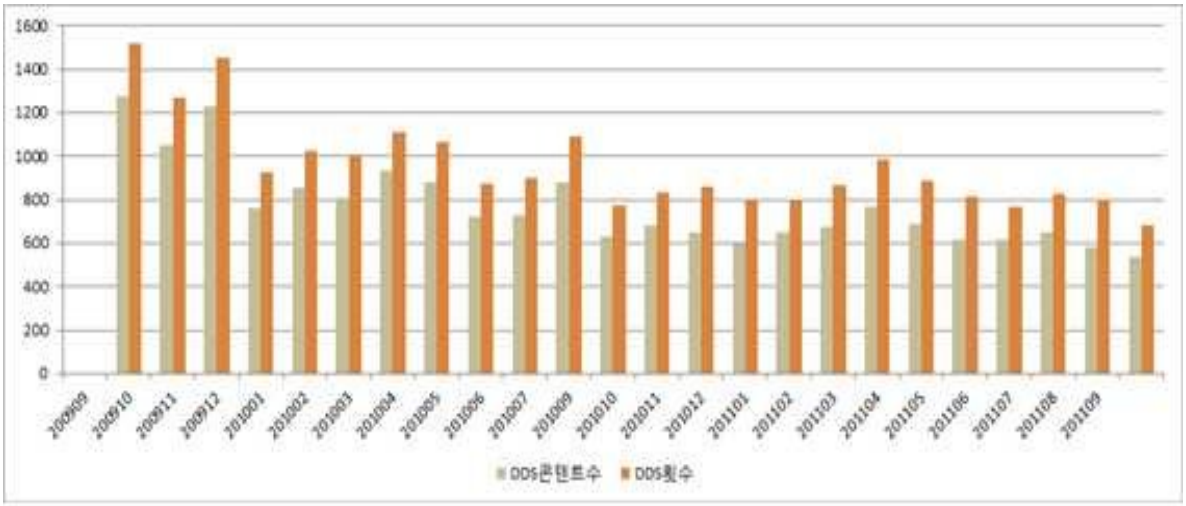
■ Technical Reports

KISTI provides science and technology technical reports that contain national R&D information obtained through the National Science and Technology Information Service (NTIS) system. NDSL provides full-text of 132,732 domestic e-technical reports. The usage numbers related to domestic e-technical reports are increasing gradually. Before this service, it was difficult to access to full-text of national R&D technical reports over the internet. The numbers of download records are 789,554 records for 2009-2010 (12 months) and 1,140,411 records for 2010-2011 (12 months). The numbers of downloads full-text of technical reports increased by 144%. Providing easy access to full-text of Korean government funded R&D research results in technical reports is important for other researchers in Korea.



<Fig. 5> Usage of E-Technical Reports (2009-2011)

KISTI holds 53,064 titles of domestic and overseas print technical reports and 149,955 titles of overseas micro-SRIM technical reports. NDSL provides metadata for technical reports for them but some of them do not have metadata. Fig. 6 shows the usage data for the Document Delivery Service for technical reports, showing a gradual reduction. However, users are still use Document Delivery Service for technical reports because some of them are still in print format only or because e-technical reports are not available to the public.



<Fig. 6> Usage of Document Delivery Service for Technical Reports (2009-2011)

New Ways for distributing GL with Social Networking Services

Satisfying users' needs and providing more convenient services, KISTI is expanding the number of access points for grey literature to exceed that of a digital library. KISTI is attempting to construct new ways of distributing grey literature system using Social Networking Services (SNSs). KISTI joined Twitter for this (@NDSL_kr), developing a service for users. The Twibort search system (ndsl_kr twibort) provides 1,000 NDSL records when users short search terms in the Twitter interface. The result can transferred to Twitter users and other social media services such as Facebook. Users also can access GL on NDSL through a smartphone while moving, meaning that they can access GL anywhere, and anytime.

Conclusion

Issues pertaining to the change from grey to white

KISTI has managed non-commercial distribution channels for grey literature for several decades and will continue in an effort to play a central role as a supporter for R&D researchers in Korea.

There are new ways of producing grey literature such as institutional repositories, new ways of communication using social media and etc. In spite of the development of information technology and tools, it is still difficult to access and identify grey literature. The effort of offering new tools and channels with which to access and disseminate scientific and technical grey literature should be continued to support R&D researchers. Additional ideas are given below.

- Identifying, making access points and collecting electronic and print GL
- Increasing metadata for GL on the web as an access point
- Providing full-text GL on the web
- Using new technology and tools for dissemination, e.g., smartphone, the iPad
- Participating social networking services: new ways of communication
- Providing other types of GL, e.g., fact data, statistics
- Discovering new types or formats for GL
- Continuing Document Delivery for paper-only GL
- Cooperating with other domestic or overseas GL networks
- Finding and using institutional repository for GL
- Adapting new research environments and channels for ST communication
- Protecting intellectual property
- Archiving GL

References

Cho and Choi. 1997. A study on output and distribution for grey literature in the internet. *Korea information management conference proceedings* 1997(4): 55-58.

Hyun-Yang Cho. 2008. A comparatives study on usability of grey literature among science. *Journal of the Korean library and information science society* 42(1): 273-294.

J. C. McDermott. 1995. Defining Grey Literature for the 21st Century. *Proceeding of the Second International Conference on Grey Literature.* 9-17.

Julia Gelfand. 2000. Intervew with Dominic Farace, Founder of GreyNet. *The International Journal on Grey Literature* 1(2): 73-76.

Keith G Jeffery, Anne Asserson. 2005. Grey in the R&D Process. *Proceeding of the International Conference on Grey Literature.*

Korea Institute of Science and Technology information. 2006. *Development of the acquisition and distribution model for the foreign grey literature*. Seoul: KISTI.

K. Nagesware Rao and V.G. Talwar. 2011. Content-based document recommender system for aerospace grey literature 7(3): 127-135.

M. C. Dabachere. Problems in obtaining grey literature. 1995. *IFLA Journal* 21(2): 94-98.

Suyeon Yoo, Kijeong Shin, and Heeyoon Choi. 2009. Collecting foreign digital grey literature through a consortium in Korea . *The Grey Journal* 7(3): 142-146.

Young-Joon Nam and Hyun-Yang Cho. 2006. A study on the active flow of foreign grey literature in Korea. *Journal of the Korean library and information science society* 40(3): 273-294.

I search

I find

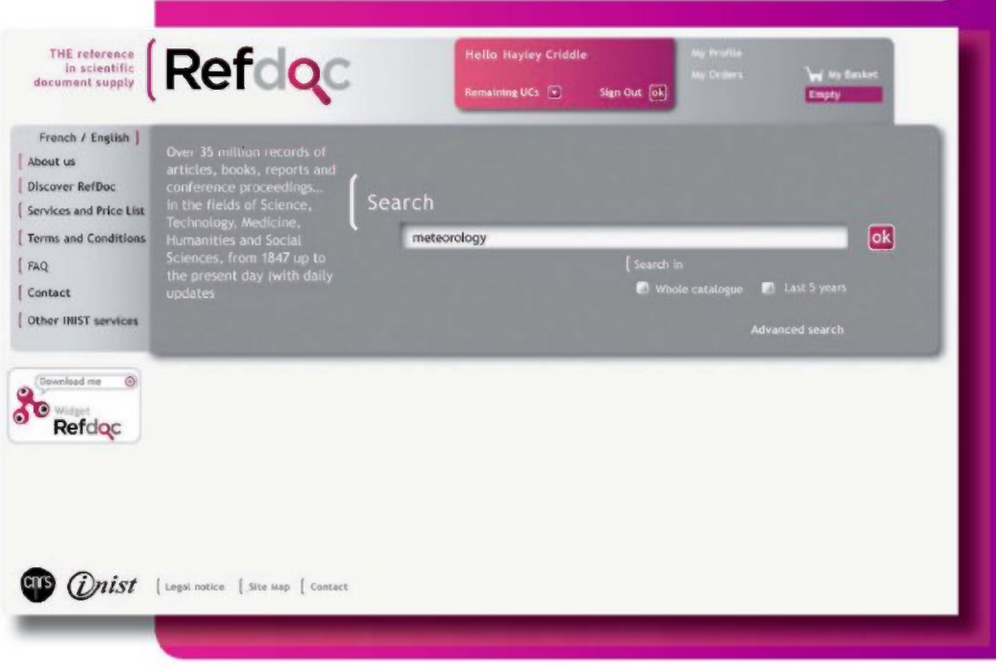
I order

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



bbcom 5677

cnrs inist diffusion

INIST DIFFUSION

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

Open Is Not Enough
Grey Literature in Institutional Repositories

Joachim Schöpfel, Charles de Gaulle University Lille 3, France
Isabelle Le Bescond, University Lille 1, France
Hélène Prost, INIST-CNRS, France

Abstract
The paper contributes to the discussion on the place of grey literature in institutional repositories and, vice versa, on the relevance of open archives for grey literature. Even in an open environment, grey literature needs specific attention and curation. Institutional repositories don't automatically provide a solution to all problems of grey literature. Our paper shows some scenarios of what could or should be done. The focus is on academic libraries.
The paper is based on a review of international studies on grey literature in open archives. Empirical evidence is drawn from an audit of the French repository IRIS from the University of Lille 1 and from ongoing work on the development of this site.
The study includes a strategic analysis in a SWOT format with four scenarios. Based on this analysis, the paper provides a set of minimum requirements for grey items in institutional repositories concerning metadata, selection procedure, quality, collection management and deposit policy. The communication is meant to be helpful for the further development of institutional repositories and for special acquisition and deposit policies of academic libraries.

1. Introduction

Along with other documents and items, grey literature contributes to the success of institutional repositories. Its non-commercial and alternative nature puts grey literature in close proximity to the community-driven culture of open archives.
But does this mean that "grey literature is at home in open archives" (Luzi, 2010) and that it should be re-defined through this new vector of scientific communication?
After years of debate on open access and grey literature, the international conference GL12 at Prague offered two different perspectives. Marzi et al. (2010) stated that "open access is the key to knowledge" and that "web-base sharing facilities and distributed access to openly available information" are key features of grey literature. For Marzi and her colleagues, institutional repositories became the future of grey literature, and grey literature hardly exists without or beyond open access.
On the other hand, our own communication defined additional attributes for grey literature that are not necessarily linked to open access, such as intellectual property, quality and interest for collections. Institutional repositories are an interesting and important vector for dissemination of grey literature but they are not enough. Based on literature review and survey data, we made a proposal for a new definition of grey literature ("Prague definition") with four new essential attributes: "Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body" (Schöpfel, 2010).
Concerning open archives, we added that "institutional repositories have started to take over some of the traditional roles of library holdings. In terms of function, they bear some equivalency with grey literature itself, as their main role consists in dissemination and, to a lesser extent, preservation" (ibid). Institutional repositories are important for grey literature but they are not the only option, and they have to satisfy some minimum requirements in order to offer an adequate home for grey literature.
Institutional repositories and grey literature can become a fertile and profitable encounter for scientific communities. But open is not enough. Here are the reasons.

2. Background: A review of grey literature and institutional repositories

Institutional repositories (IR) became a significant channel of digital scientific communication.[1] Part of the open access movement and alongside with subject-based repositories, research repositories or national repository systems (Armbruster & Romary, 2009), they focus on "serving the interests of faculty – researchers and teachers - by collecting their intellectual outputs for long-term access, preservation and management" (Carr et al., 2008).

They can be seen as "tools (…) for collecting, storing and disseminating scholarly outputs within and without the institution" (Jain, 2011), as "a set of services (…) for the management and dissemination of digital materials created by the institution and its community members" (Lynch, 2003) or as an "organisational commitment to the stewardship of these digital materials" (ibid.).

One of their main characteristics is their great diversity. There is not *one* model but multiple possibilities, not *one* path but a multiplicity of options. Yet it is crucial for their success that the institution clearly defines the objective of its repository, in line with its own strategy and environment. "Each of the reasons for setting up a repository carries implications for the content, design and funding of a repository, and the institution needs to be clear about the implications of different roles for a repository, while being prepared to change or add roles as the scholarly communication environment develops" (Friend, 2011).

Institutional repositories have different policies, procedures, functionalities, services and metadata, they have different business models and funding strategies (Swan & Awre, 2006), and their content may include more than current output from faculty. Smith (2008) details a "wide variety of materials in digital form, such as research journal articles, preprints and postprints, digital versions of theses and dissertations, and administrative documents, course notes, or learning objects." Other repositories include datasets, multimedia or cultural and scientific heritage.

Of course, grey literature as unpublished, special or not-for-profit documents is part of the repositories' content. But what is its place in institutional repositories, and what is the relevance of institutional repositories for grey literature?

2.1. The place of grey literature in institutional repositories

Some empirical studies contribute to a realistic vision on grey literature in institutional repositories. Luzi et al. (2008) estimate the part of grey materials eligible for the institutional repository of the Italian National Research Council at about 1/3 of all items, even if not all of these documents are freely available.

In our survey on French repositories, grey literature represents 18% of all documents (Schöpfel & Prost, 2010). Another survey on Spanish repositories reveals that at least 23% of the deposited items with full-text are grey (Melero et al., 2009). Both studies confirm, too, that the number of grey documents in repositories is rapidly growing.

Vernooy-Gerritsen et al. (2009) report results from the EU-sponsored DRIVER project on institutional research repositories. They separate full-text records (33%) from metadata only records and records of non-textual and other materials; 62% of the full-text records are grey literature. This percentage corresponds to 20% of the whole content.

Most of all these grey items are theses, dissertations, proceedings, unpublished papers (working papers) or reports. Up to now, course material is less important.

The part of 20-30% of repository content is somewhat higher than the average percentage of grey literature in citation analyses (see Schöpfel & Farace, 2010).

So far, there is but little evidence on usage of grey items in institutional repositories. Yet, recent studies on access statistics suggest that downloads per item are often higher for unpublished theses or reports than for published articles (Schöpfel et al. 2009, see also Kroth et al. 2010).[2] One reason may be that these items can't be viewed elsewhere.

2.2. The relevance of institutional repositories for grey literature

To which extent are institutional repositories the place for grey literature? According to the information of the OpenDOAR directory of open archives, 82% of all institutional repositories contain grey literature.

Type of documents	Nb IR with these items	% of all IR (n=1,978)
Theses, dissertations	958	48%
Unpublished	616	31%
Proceedings	572	29%
Learning objects	245	12%
Special items	235	12%
Total	1,628	82%

Table 1: Grey items in institutional repositories (source: Open-DOAR, June 2011)

The OpenDOAR figures are comparable to results from France and Spain. In France, 94% institutional repositories hold grey documents while their part is significantly lower in subject-based repositories (37%) or national or research repositories (23%) (Schöpfel & Prost, 2010). In Spain, more than 80% repositories contain theses, and at least 60% have unpublished working papers and/or proceedings (Melero et al., 2009).

For some of this material, especially for specific types of unpublished items like slides, posters or other, supplementary material, it is surely true that "this is academic output that would not likely be otherwise captured and made freely available were it not for publication in an IR" (Kroth et al., 2010).

Some papers praise the impact of institutional repositories for grey literature. On the word of Luzi (2010), they provide "a natural home for GL" because they amplify its dissemination. Open access makes grey literature "less grey and more white" (Gelfand, 2004); the "distinction between GL and conventional literature is becoming increasingly blurred" (Luzi, 2010; see also Swan 2008 and 2011).

Yet, this "blurring" only applies to potential usage, not to value or quality. Banks (2005) believes that even if the hierarchy between grey and white may shift into a continuum of scholarship, this hierarchy will not completely disappear insofar institution and faculty generally prefer published and peer reviewed documents. A recent study on content recruitment and usage in an institutional repository confirms this belief (Connell, 2011).

2.3. Grey issues

Studies on grey literature in institutional repositories recurrently point out six critical aspects for the success and development of such initiatives.

Community: Describing a conference proceedings repository at Cornell, Rupp & LaFleur (2004) plead for "a specific workflow (…) for the identification and gathering of proceedings" that includes public relations, "one-to-one marketing" and communication with faculty to create awareness and get the documents from the author's desktop into the repository. Without community, no repository.[3]

Quality control: A repository that is "'all things to all people' lacks focus" (Westell, 2006). Specific action from the very beginning of the workflow is required to guarantee a minimum quality of content, data and services. Control procedures and workflow technology should ensure quality of item selection and overall project management (Luzi et al., 2004).

Metadata: Grey literature in institutional repositories has need of specific metadata for identification and bibliographic description. For instance, Ruggieri et al. (2009) propose a table with mandatory and optional metadata fields, including a note field, for conference papers, oral presentations, reports and in-house publications. Jeffery (2007) adds that "the syntax must be formal and precise; the semantics must be present, formal and precise (…); the relationships form a fully-connected graph; (…); the relationships require an annotation richer than the triples of RDF (…)." Yet, unfortunately the reality is that "current metadata elements (of electronic theses and dissertations in IR) have a significant level of inconsistency and variation" (Park & Richard, 2011), and often "individual institutions (decide) locally how metadata elements should be defined (ibid).

Interoperability: Institutional repositories are hardly ever stand-alone systems. They should be interoperable or at least three reasons: maybe because their institution is part of a network (Dijk, 2007), maybe because they are connected and exchange data and items, maybe simply because the OAI initiative

stipulates interoperability. Pejsova (2011) describes a national system for grey literature that is interoperable with local repositories for documents, metadata and workflow.

Integration: Some authors insist on the integration of institutional repositories and grey literature into current research information system (CRIS) infrastructure. “An institutional repository, being a central point within the organisation for literature and data, is a component of the integration of processes, which promises benefits both to the organisation itself and to the researchers within it” (Lambert et al., 2005).

New item formats: Jeffery (2007) calls for a linkage between CRIS and e-repositories for grey literature on the institutional level, and he suggests that they should be associated to repositories for research datasets and software, via the CRIS. More recently, Doorenbosch & Sierman (2011) focus on the changing nature of scholarly publications, e.g. enhanced publications with both documents and datasets, outline the challenge of these new items for long term preservation in institutional repositories, and suggest the creation of “collaborative virtual research environments are considered to be the new workspaces for researchers”.

3. Case study: The IRIS audit – grey literature at home at Lille

The IRIS repository, hosted by the Lille 1 university, successor to Grisemine, the first French open archive for grey literature. Its development and usage have been presented at the GL5 and GL12 conferences (Claerebout, 2003; Prost et al., 2010). The following case study provides a short overview on the Grisemine/IRIS history and illustrates some conditions that are favourable or not for the deposit and dissemination of grey literature in institutional repositories.

3.1. General remarks

When Grisemine was launched in 2001, it was one of the first open archives in France, a pioneer especially in the academic sector. Its notoriety and popularity among academic librarians was immediate and without doubt superior to its real impact on scientific communication.

Since 2001, Grisemine underwent deep changes. This “Grisemine/IRIS decade” demonstrates the coming out of the hybrid digital library with service marketing rather than collection building. Nearly all has changed – the name, software, architecture and workflow, content, strategy, policy and institutional positioning.

The story of Grisemine/IRIS is not over. In fact, it just began, again. But which may seem, ex post, logical and necessary often was trial and error, searching for opportunities, benchmarking, exploration and adaptation to a moving context.

3.2. Rise and decline of Grisemine (2001-2005)

Grisemine’s purpose was to collect, preserve and disseminate French[4] grey literature, such as theses and dissertations, communications, notes, working papers, preprints, exam topics or educational programs. Grisemine was developed with the CinDoc electronic content management software (Cincom). Its workflow was compliant with the Dublin Core metadata standard and the MARC format.

Even as a prototype, the Grisemine project was technically viable, except for the technical maintenance and development of the CinDoc software. But it had no real institutional recognition, was a "librarians’ toy" rather than a labelled, validated and accepted repository for the scientific community. Yet, its content (1,300 documents in late 2005) was widely consulted, in particular from French-speaking countries.

It became obvious, too, that the initial goal – a deposit for all French grey literature – was too ambitious and disproportionate to the allocated resources.

3.3. From Grisemine to IRIS (2006-2010)

In 2006, the French government published a decree on the processing, preservation and dissemination of electronic PhD theses and launched a national network for ETDs called STAR. Grisemine was not able to support the new workflow. For this and other reasons mentioned above (maintenance), the Lille library team considered Grisemine as a technical and documentary dead-end. The next four years were a period of transition.

The most important decision was to migrate from CinDoc to DSpace, and then make the system dialogue with STAR. The migration was operational in 2007. With the migration, Grisemine became IRIS.

Why DSpace? At the time the Lille team took the decision to migrate (2004-2006), DSpace was the most common software for open archives, and it was easy to install. Yet, DSpace is designed for self-deposits, not for an encyclopaedic-like collection (scientific heritage) or an institutional and/or national workflow

(theses). Without a dedicated information technology (IT) staff, the Lille library decided to maintain DSpace at best until the new ORI-OAI software became available[5]. "At best" meant keeping the archive alive, continued uploads but no development. For instance, an early project to separate PhD theses and scientific heritage was put on ice.

The deposit of e-theses became mandatory on the Lille 1 campus in 2008, because of STAR. IRIS was able to provide an operating OAI platform for their dissemination but didn't offer a solution for their management or preservation. The open dissemination of Lille ETDs became the main function of the IRIS repository. In December 2010, IRIS had 625 theses and 711 other documents. Their long-term preservation is supported by the academic data centre CINES at Montpellier[6].

With the move from Grisemine to IRIS, the site abandoned its initial strategy as an open repository for French grey literature. The self-deposit of grey items ceased completely. Instead, the library team made another use of the IRIS platform and developed, together with a historical research centre and the academic digitization centre at Lille, a digital library with a collection of copyright cleared documents (articles, papers, books) on the history of sciences. Alongside with the PhD theses, this heritage collection was made freely available on the IRIS platform and is very appreciated by the scientists.

When the university decided the mandatory deposit of e-theses in 2008, it also acknowledged IRIS as the official Lille 1 institutional repository. Yet, this decision was not accompanied or followed by a mandatory policy for the whole scientific production of the faculty. Except some professorial habilitation theses and learning objects, IRIS never received any self-deposits from Lille researchers.

3.4. Rebirth (2010-2011)

At the end of the first decade, the strategic positioning of IRIS was atypical and confusing. The university administration considered IRIS as the official institutional repository. Yet, there was no promotion, communication, incentives or mandate, and the only open archive with a significant number of self-deposits from Lille 1 faculty was (and always is) the French national research repository HAL with 16,143 items.[7]

The library team regarded IRIS as a digital library, more like GALLICA or PERSEE than ArXiv or HAL, yet used the IRIS server for the dissemination of PhD theses, a service usually considered to be a key element of academic institutional repositories, and made some tests with other scientific output from Lille faculty, especially in the context of an emerging learning centre project.

In 2010, with the installation of the ORI-OAI system the Lille 1 repository took a new start. Why ORI-OAI? At least for four reasons: compliance with French metadata standards for theses (TEF) and learning objects (SupLOMFR), interoperability with the nationwide infrastructure for ETDs (STAR) and the national research repository HAL, a French community of software developers and end-users, quality of development and product.

Today Lille 1 hosts a composite repository with two systems accessible through two different interfaces:

- ORI-thèses with theses, habilitations and learning objects.[8]
- IRIS with the collection of history of sciences.[9]

In fact, IRIS became a digital library without input from current scientific production.

A third platform for the self-deposit of scientific production (pre- and post-prints, communications, reports…) is under construction, on the model of the Toulouse OATAO[10] repository or the Luttich ORBi[11] site, and will be launched in 2012 probably with a new name.

3.5. Concluding remarks

As we said above, the story of Grisemine/IRIS is not over and it may be premature to debrief. Yet, we tried to highlight some main characteristics of this project and then to identify the factors in favour of grey literature and success.

The development of the Lille 1 repository was non-linear, dependent on the evolving local and national context, on technology (software) and standards. The library team's quest for legitimacy was complicated by the pluridisciplinarity of their academic community and by the fact that in France, the open archives for scientific information were initially hosted and managed by the public research organisations (CNRS, INRA, IFREMER…).

On the other hand, the national infrastructure for electronic theses (STAR system with TEF metadata standard) and the library's experience with preservation and dissemination of cultural and scientific heritage items – a traditional library function - facilitated the legitimacy and positioning of the project.
So which were the critical key elements for success or failure? Briefly:[12]

- Institutional support and recognition of the project team and the repository.
- Institutional strategy and policy in the domain of open archives and deposit mandate.
- Human resources with sufficient IT and LIS capacities.
- Metadata standard(s) for a careful and precise bibliographic description of the deposited content.
- Software fitting with local needs and IT environment as well as with national infrastructure and standards.
- A solution for perennial preservation of deposits (at least for the theses).
- Added value services for legal aspects and usage statistics.
- Knowledge of the scientific community's information needs and behaviours, and integration into the larger academy.

The Grisemine/IRIS case shows also a close link between grey literature typology, IT solutions (software) and workflow features. The repository must cope with specific conditions, such as (for the Lille repository) the national STAR system for theses or the digital university environment (UNT) for the learning objects. The need to align deposit with existing workflows was highlighted by Westell (2006) and Troll Covey (2011). This, together with the different software solutions, argues for a differential approach to grey literature in institutional repositories. Some grey documents may be at home in some open archives, while others in different ones.

4. SWOT analysis: Grey literature in institutional repositories

Based on the review of literature and standards and including the IRIS experience, our evaluative synthesis will take the form of a strategic SWOT diagnostic, keeping apart internal and external factors that are favourable or unfavourable for grey literature in institutional repositories. However, our analysis does not take into account more general aspects that are not directly related to grey literature (for instance, such as Pinto & Fernandes, 2011).

4.1. Strengths

The internal factors in favour of grey literature in institutional repository models are:

1. Grey literature amplifies the content of institutional repositories.
2. Free availability, dissemination, visibility and referencing act as incentives for grey deposits.
3. What's more, relatively high usage of unpublished items may also act as an incentive for grey deposits.
4. Institutional repositories guarantee more security and long-term accessibility of unpublished material than a personal web site.
5. Compared to published articles, there are fewer problems with copyright for grey literature.

4.2. Weaknesses

The internal factors unfavourable for grey literature in institutional repositories are:

1. The bibliographic control of grey literature, especially of conferences and reports, remains often mediocre or poor because of flawed or incomplete metadata format (non qualified Dublin Core).
2. Most often, the hosting institution doesn't provide any solution for the digital curation of metadata.
3. Deposit is time consuming.
4. Deposit of grey literature needs, more than published documents, incentives and support from institution. This support may be missing.
5. Without institutional support or incentives, self-deposits will not have the same quality as a library collection.

4.3. Opportunities

The external factors in favour of grey literature in institutional repository models are:

1. Universities need a solution for the processing, disseminating and archiving of electronic theses and dissertations (ETD). Institutional repositories offer an interesting solution and may at least be an element in the global academic information system for ETD.

2. Institutions want control on research output and content, and this includes unpublished documents.
3. Institutions want to improve presence and impact on the web. Grey literature in repositories adds to both, due to broader dissemination and increased use of grey items, increasing prestige and visibility for the institution.
4. The open access initiative is not limited to published documents.
5. The evolution from "collection development" to "content recruitment" in academic libraries may act in favour of deposit of grey literature in institutional repositories.

4.4. Threats

The external factors unfavourable for grey literature in institutional repositories are:

1. Funding and evaluation agencies put priority on published documents (articles, books) and at least partially neglect grey items. Grey literature is not indexed in the scientometric databases Web of Science and SCOPUS.
2. If institutions introduce self-archiving mandates in order to generate content, researchers may react negatively to any suggestion of compulsion. Most faculties do not respond to the invitation to "add stuff to the IR" (Jain 2011). Another side-effect is the creation of metadata only records, without full-text. This should be limited to published documents with copyright problems but it isn't.[13]
3. Alternative models, e.g. generating content through deposit by publishers (PEER project) will not impact grey items.
4. Open access through institutional repositories requires funding from particular institutions to set up and maintain a repository (Friend 2011). Poor knowledge on grey literature will make it more difficult to sustain continuous support and commitment from the management and academic staff.
5. A significant part of the scientific community lacks awareness of open access and grey literature.

5. Findings based on four scenarios

Are institutional repositories the future of grey literature? Maybe. But because of the great variety of institutional repositories, we can distinguish at least four different scenarios.

Jain (2011) makes some recommendations for the development of institutional repositories, in particular, promotion and publicity to the faculty, provision of clear policies on ownership, contents, quality and copyright, and an adequate provision of resources. This is in line with the IRIS audit and applies to all scenarios. Therefore, our description is limited to specific criteria for grey literature within this environment.

The differences are with mandatory deposit, strategic vision, services, selection procedure, quality issues, collection management and metadata. Our description is partly based on studies on objectives and business models of institutional repositories (Friend, 2011; Swan & Awre, 2006). We don't describe real cases but potential homes – a kind of ideal archetypes of institutional repositories. The reality will be more complex and composite.

5.1. Scenario 1 – Publishing grey literature

In the first scenario, the institutional repository serves essentially the initial function of open archives, e.g. communication and publishing of scientific papers. Focus is laid on rapid and direct access to full-text, for the scientific community. For grey literature, the strategy is to become less greyish and more white, through institutional digital publishing outside of usual sales channels.

The strategic objective by the institution may be twofold (cf. Friend, 2011):

- "To increase the impact of particular research or teaching programmes through exposure of publications and other outputs on open access.
- To reduce the cost and increase the benefits from the dissemination of the institution's research and teaching outputs."

The most appropriate business model for repository provision and preservation will be institutionally-supported, perhaps with a contribution by community (learned societies).

Selection procedures for a minimum content and formal quality level (through validation or "labelling") probably will be more important than mandatory issues. Self-deposit of full-text (preprints, postprints but also conference proceedings, unpublished reports and papers…) and institutional workflows for electronic

theses, perhaps also for master and habilitation theses, in-house collections of working papers or reports are essential for content recruitment while mandatory deposit policy or incentives are not.
Also, metadata are critical (only) insofar they facilitate content retrieval and access. This means that they are probably of mediocre quality and not very specific for different types of documents, except for ETD.
The primary function of this repository is communication and access to the full-text, via search engines and/or the repository's search and browse interface. The key elements are a high rate of full-text, worthy scientific content, and unrestricted access, followed by a high and representative number of deposits. Other services may be less crucial but would add value to the site:

- usage statistics services,
- preservation services,
- publishing services.

5.2. Scenario 2 – Special items container

In the second scenario, the institutional repository is a container or storehouse for all kind of material produced by faculty, staff and students. In this container, ETD, reports and conference proceedings stand next to images, learning objects, articles, datasets, presentations, posters etc.
The focus is laid on availability and visibility of all kind of materials, "institutional stuff", rather than on selection of scientific relevant results. Quality control through validation or labelling is not an issue.
The strategic objective may be "to collect together all the publications and other research and teaching outputs as a permanent record of the institution's achievements but without any specific use in mind" (Friend, 2011).
Again, the appropriate business model for repository provision and preservation is institutional support. The institution may also decide to establish a mandatory deposit, and/or incentives for self-deposit.
The underlying idea is to "dig out" hidden material, find a solution for digital dissemination and preservation, together with other published or unpublished documents.
As for quality control or editorship, metadata probably are not an important issue. Most likely, services will be limited to preservation, publishing, resource discovery and perhaps research assessment and monitoring. It is also possible to add social indexing and data mining. There is no clear vision on collection and acquisition. But the most promising perspective may be the linking of the deposits to research data.

5.3. Scenario 3 – Scientific heritage

The third scenario the institutional repository is a showcase for the past and present scientific production, with grey literature alongside with published documents and other material.
Again, the strategic vision will be "to collect together all the publications and other research and teaching outputs as a permanent record of the institution's achievements but without any specific use in mind" (Friend, 2011). The difference with scenario 2 is the heritage character of the collection, the inclusion of older material in the public domain.
But there may (also) be other motivations:

- "To increase the impact of particular research or teaching programmes through exposure of publications and other outputs on open access.
- To make a contribution to the world-wide movement for open access to publicly-funded research" (ibid.).

The definition of an acquisition or content recruitment policy is crucial, together with an institutional strategy for the digitization of older, copyright cleared material (theses, journals, books, papers, images, maps…). This may imply a more thoroughly prepared and pondered indexing and metadata policy. The outcome may be 100% access to full text, as for the IRIS repository.
The appropriate business model is institutional support. But there may be other resources, public funding for scientific heritage or thematic or special collections. For this specific case, it may be possible to experience a subscription-supported model, appropriate for access and authentication, preservation and resource discovery services.
Also, the local presence of a digitisation centre may allow those repositories to populate content more rapidly, especially grey literature, and to attract usage (Westell, 2006).

The underlying idea is digital preservation of heritage collection, together with making these collections available to scientists, students and all interested people. This may be complementary to publishers' backfiles.
This scenario is probably the closest scenario to traditional library collection building, with issues such as quality, indexing, classification etc. Evaluation, scientometrics etc. may be less important, at least not in the heart of the project.

5.4. Scenario 4 – Institutional deposit

The last scenario for grey literature in institutional repositories is mandatory institutional or self-deposit in the way it is promoted by Stevan Harnad: green road (self-deposit) to free online full-text access to peer-reviewed literature, through an explicit and institutional mandatory policy in order to obtain commitment by close to 100% of the authors.
This scenario is meant to demonstrate the value of the institution itself through a kind of quasi-legal deposit showcase, to facilitate control over scientific production and evaluation procedures, and corresponds to one or more institutional strategies, e.g.

- "To report the publications and other research and teaching outputs to funding agencies in support of new grant applications.
- To report the publications and other research and teaching outputs to funding agencies as part of an audit of expenditure.
- To demonstrate to governments or taxpayers the impact of the institution outside its walls (a purpose which will require the compilation of metrics).
- To increase the impact of individual members of the institution's staff through the exposure to potential academic and commercial users of the individual's publications and other outputs on open access" (Friend, 2011).

The business model will surely be institutionally-supported and may include services such as usage statistics, research assessment and monitoring, bridging and mapping, and technology transfer/business advice. Also, a connection to a current research information system (CRIS) should be possible.
The impact on grey literature in this environment is triple:
Peer-reviewed publications will play a major role in this environment, and in comparison, grey literature will be less valued or appreciated. This may have a negative impact on metadata.
The institutional policy of mandatory deposit generate a relatively high rate of metadata only records without access to full-text because of embargo, sensitive content, missing authorization by co-authors etc. Paradoxically, this "collateral damage" also impacts grey literature (see above, footnotes 7 and 12). Only 40% of the HAL grey literature records are with full text.
The number of grey documents will be significant but more or less limited to specific categories evaluated by agencies, such as theses and dissertations, conference proceedings and project reports.
The main interest of these repositories is not collection building but evaluation. Insofar grey literature enters evaluation procedures it will be valued and welcome in this environment.

6. Results and concluding remarks

Our paper started with Luzi's (2010) statement that "grey literature is at home in open archives". This may be right but as we tried to demonstrate, open archives not only offer one but at least four different homes that may be complementary, at least to some extent.
Mapped on two dimensions, policy (evaluation vs. communication) and quality (library vs. container), the four options clearly occupy different positions (figure 1).

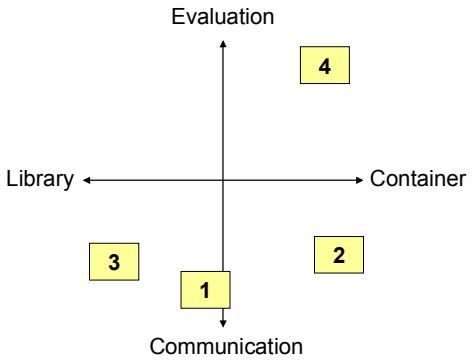


Figure 1: A map of four scenarios for institutional repositories with grey literature

In scenario 1, the political priority is laid on communication of research results, full-text, community, scientific value. Grey literature is part of the content insofar the depositing authors consider it worthy enough for direct communication and preservation. But there is no real control or selection.

In scenario 2, the main objective is the container function, the deposit of all materials produced by faculty, students and staff. Again, the institutional policy is communication-centred but without selection or validation criteria. Grey literature has its home here – in a (too) large sense and together with a lot of other stuff.

Selection or validation criteria are introduced by the 3rd scenario. Here the institution applies a policy of showcase and scientific heritage, most likely accompanied by digitization programs. The place of grey literature depends on the institution's acquisition policy and digitization program.

The 4th scenario reflects the institutional policy in favour of evaluation and ranking. Full-text and communication are secondary goals while metadata and a minimum quality control are necessary. Deposit of grey literature will be welcome insofar it enters evaluation.

Now, which is the most adequate option for grey literature? The response depends on institutional policy, library goals and professional viewpoint. For the scientific community, end-user and consumer of scientific information, perennial open access to validated items in full-text format is priority. This priority implies at least five minimum requirements:

Access to full-text. Open archives with metadata only records are like libraries with empty shelves.

Quality through selection, validation and/or labelling. Even without peer-review or other, web-based reviewing procedures, grey deposits should meet with some basic quality criteria. Incite deposit of all kind of uninteresting stuff is like keeping waste paper on the desktop. Self-deposit is not collection building.

Openness without restriction and/or embargo. Confidential, classified or non-copyright cleared material should not be part of open archives but should be managed via catalogues, databases or other systems.

Metadata quality. Repositories should guarantee a minimum level of metadata quality, e.g. compliance with standards and curation. This requirement is necessary for information retrieval, interoperability and the semantic web.

Long-term conservation. Institutional repositories should offer a solution for the ephemeral nature of grey literature, via a clear statement on and investment in perennial content preservation, if necessary also via outsourcing or "in the clouds".

For the scientific community, the best option for grey literature may be a mix of scenarios 1 "publishing grey literature" and 3 "scientific heritage". Other elements will add value (standard format and metadata, usage statistics, discovery functions, scientometrics) or increase sustainability (institutional support, integration in research community, promotion and communication, interoperability). But they are not specific to grey literature.

We didn't speak about format and legal matters; yet, they may be critical matters for the future of repositories. With the words of Swan (2011), "we (can't) relax (and) watch repositories fill with articles and datasets". Or as Anderson (2011) put it, "accessibility is not access."

The IRIS case should raise awareness that the same solution may not be appropriate to all kind of grey literature and disciplines and that the system should be evolutionary and flexible enough to easily adapt to and keep up with new conditions and opportunities.

A last and rather paradox remark: the success of institutional repositories may become a problem for grey literature, especially when the institution implements a mandatory deposit policy that gives priority to

evaluation and control and not to publishing and communication. Anna Clements, a data manager from St Andrews University, described the problem some time ago on the JISC-Repositories listserv: libraries create institutional repositories with full-text or full objects as the main content, and they are then asked by the institution to look at hosting citations without full-text as well.

A library with empty book shelves may be interesting to research managers but not for scientists. In this case, grey literature would definitively not be at home in institutional repositories. Open is not enough.

6. References

K. Anderson (2011). `Does Access Create New Types of Scarcity?'. *The Scholarly Kitchen* Aug 31.

M. A. Banks (2005). `Towards a Continuum of Scholarship: The Eventual Collapse of the Distinction Between Grey and non-Grey Literature?'. In *Seventh International Conference on Grey Literature: Open Access to Grey Resources, Nancy, 5-6 December 2005*.

L. Carr, et al. (2008). `Institutional Repository Checklist for Serving Institutional Management'. In *Third International Conference on Open Repositories 2008, 1-4 April 2008, Southampton, United Kingdom*.

M.-F. Claerebout (2003). `Grisemine, a digital library of grey university literature'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information, 4-5 December 2003*.

T. H. Connell (2011). `The Use of Institutional Repositories: The Ohio State University Experience'. *College & Research Libraries* **72**(3):253-275.

E. Dijk (2007). `Accessing grey literature in an integrated environment of scientific research information'. In *Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, Antwerp, 10-11 December 2007*.

P. Doorenbosch & B. Sierman (2011). `Institutional Repositories, Long Term Preservation and the changing nature of Scholarly Publications'. *Journal of Digital Information* **12**(2).

F. Friend (2011). `Open Access Business Models for Research Funders and Universities'. Tech. rep., Knowledge Exchange, Copenhagen.

J. Gelfand (2004). `"Knock, Knock:" Are Institutional Repositories a Home for Grey Literature?'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, New York, 6-7 December 2004*.

P. Jain (2011). `New trends and future applications/directions of institutional repositories in academic institutions'. *Library Review* **60**(2):125-141.

K. G. Jeffery (2007). `Greyscape'. In *Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, Antwerp, 10-11 December 2007*.

P. J. Kroth, et al. (2010). `Institutional Repository Access Patterns of Nontraditionally Published Academic Content: What Types of Content Are Accessed the Most?'. *Journal of Electronic Resources in Medical Libraries* **7**(3):189-195.

S. Lambert, et al. (2005). `Grey literature, institutional repositories and the organisational context'. In *Seventh International Conference on Grey Literature: Open Access to Grey Resources, Nancy, 5-6 December 2005*.

D. Luzi, et al. (2004). `The integration of GL documents with a research information system on occupation safety and health'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, New York, 6-7 December 2004*.

D. Luzi, et al. (2008). `Towards an Institutional Repository of the Italian National Research Council: A survey on Open Access experiences'. In *Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society, 8-9 December 2008*.

D. Luzi (2010). `Grey Documents in Open Archives'. *The Grey Journal* **6**(3):137-144.

C. A. Lynch (2003). `Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age'. Tech. Rep. 226, ARL Association of Research Libraries.

C. Marzi, et al. (2010). `A terminology-based re-definition of grey literature'. In *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010*.

R. Melero, et al. (2009). `The situation of open access Institutional repositories in Spain: 2009 report'. *Information Research* **14**(4).

B. Mukherjee & M. Nazim (2011). `Open access institutional archives: a quantitative study (2006-2010)'. *DESIDOC Journal of Library & Information Technology* **31**(4):317-324.

E. G. Park & M. Richard (2011). `Metadata assessment in e-theses and dissertations of Canadian institutional repositories'. *The Electronic Library* **29**(3):394-407.

P. Pejsova (2011). `Czech National Repository of Grey Literature'. *The Grey Journal* **6**(3):117-122.

M. J. Pinto & S. Fernandes (2011). `Gaining a sustainable IR: thinking SWOT'. In *QQML2011: 3rd International Conference on Qualitative and Quantitative Methods in Libraries (DLMC 2011: Digital Library Multi-Conference), Athens, Greece, 24-27 May, 2011*.

H. Prost, et al. (2010). `Usage assessment of an institutional repository: a case study'. *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010*.

R. Ruggieri, et al. (2009). `From CNR Annual report to an Institutional repository: Results from a survey'. In *Eleventh International Conference on Grey Literature: The Grey Mosaic, Piecing It All Together, Washington DC, 14-15 December 2009*.

N. Rupp & L. J. LaFleur (2004). `Making Grey Literature Available through Institutional Repositories'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, New York, 6-7 December 2004*.

J. Schöpfel, et al. (2009). `Usage of grey literature in open archives'. In *Eleventh International Conference on Grey Literature: The Grey Mosaic: Piecing It All Together. Washington D.C., 14-15 December 2009*.

J. Schöpfel (2010). `Towards a Prague Definition of Grey Literature'. In *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010.*

J. Schöpfel & D. J. Farace (2010). `Grey Literature'. In M. J. Bates & M. N. Maack (eds.), *Encyclopedia of Library and Information Sciences, Third Edition*, pp. 2029-2039. CRC Press, London.

J. Schöpfel & H. Prost (2010). `Développement et Usage des Archives Ouvertes en France. Rapport. 1e partie: Développement'. Tech. rep., Université Charles-de-Gaulle Lille 3.

J. Schöpfel & H. Prost (2011). `IRIS. Etat des lieux et perspectives. Rapport d'audit'. Tech. rep., University of Lille, Villeneuve d'Ascq.

K. Smith (2008). `Institutional Repositories and E-Journal Archiving: What Are We Learning?'. *Journal of Electronic Publishing* **11**(1).

A. Swan & C. Awre (2006). `Linking UK Repositories: Technical & Organisational Models to Support User-Oriented Services Across Institutional & Other Digital Repositories'. Tech. rep., JISC, London.

A. Swan (2008). `Study on the availability of UK academic "grey literature" to UK SMEs: Report to the JISC Scholarly Communications Group'. Tech. rep., JISC, London.

A. Swan (2011). `Institutional repositories - now and next'. In P. Dale, J. Beard, & M. Holland (eds.), *University Libraries and Digital Learning Environments*, pp. 119-134. Ashgate Publishing, Farnham.

D. Troll Covey (2011). `Recruiting Content for the Institutional Repository: The Barriers Exceed the Benefits'. *Journal of Digital Information* **12**(3).

M. Vernooy-Gerritsen, et al. (2009). `Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe'. *Ariadne* (59).

M. Westell (2006). `Institutional repositories: proposed indicators of success'. *Library Hi Tech* **24**(2):211-226.

Websites visited between September-November 2011.

[1] See the quantitative study from Mukherjee & Nazim (2011).
[2] For instance, the 2010 annual report of the French Research Institute for the Exploitation of the Sea shows that the average usage for theses in their IR is 4x higher than for published articles, see http://wwz.ifremer.fr/institut/L-institut/Documents-de-reference/Rapports-Annuels
[3] See also the disillusioning survey from Seaman (2011).
[4] French means: edited in France and/or in French language.
[5] A document management system compliant with OAI-PMH, designed for the publishing, sharing and dissemination of academic digital resources and supported by the French Ministry of Higher Education http://wiki.ori-oai.org
[6] Preservation of ETDs via STAR, preservation of other deposits on a contractual basis.
[7] But only 13% of these items have full-text, the rest are metadata only records (12 October 2011).
[8] http://ori.univ-lille1.fr
[9] https://iris.univ-lille1.fr
[10] http://oatao.univ-toulouse.fr/
[11] http://orbi.ulg.ac.be/
[12] See also Westell (2006).
[13] For instance, only 45% of the deposited working papers, conferences and ETD in the Belgian ORBi repository provide access to the full-text.

Grey Literature Matters: The Role of Grey Literature as a Public Communication Tool in Risk Management Practices of Nuclear Power Plants

Cees de Blaaij
Library of Zeeland, Netherlands

Abstract
In 2010 about 60 countries expressed interest to the International Atomic Energy Agency (IAEA) in launching nuclear programs. 29 countries with existing programs are planning to expand their nuclear capacity. With the Three Mile Island, the Chernobyl and Fukushima nuclear accidents it has been shown that the consequences of a catastrophic nuclear accident are huge and therefore adequate risk management is crucial.
The general objective of risk management in relation to nuclear power plants is the planning to minimize the impact of any nuclear catastrophe. Risk communication towards the intended public is a vital part in the risk management approach. The question to be answered in this article is to what extent scientific information is provided by the nuclear industry, responsible governments and ngo's like IAEA as a communication tool in order to keep the public informed as to minimize the risks of a nuclear catastrophe. Is there a "best" option? Differences between countries with existing nuclear programs will be evaluated in their public information policies. The method of the study employs a literature survey and a qualitative evaluation.

Keywords: *Grey literature, Risk communication, Nuclear Power Plants.*

Introduction

The issue of the use of nuclear energy is hotly debated in the last fifty years. In 2010, 60 countries expressed interest to the International Atomic Energy Agency (IAEA) in launching nuclear programs. 29 countries with existing programs are planning to expand their nuclear capacity. Despite the Fukushima disaster the IAEA projects a growth from the present 432 operating nuclear power reactors to more than 500 in 2030.[1] Most of the growth will occur in countries that already have operating nuclear power plants, such as China and India.
On one hand there is the promise of an almost inexhaustible source of energy in an age where the demand for energy is higher than ever. On the other hand we have witnessed a reality check: nuclear accidents that have happened at Three Miles Island, Chernobyl and Fukushima.
These events led to the strengthening of public concerns about vulnerability and the heightened awareness of risks on nuclear power. It questions the nature of the relationship between the public perception of risk and the way nuclear organizations deal with risks.
The term risk communication was first used in the United States at the end of the 80s.[2] According to the American Nuclear Regulatory Commission risk communication "is an interactive process used in talking or writing about topics that cause concern about health, safety, security, or the environment."[3] Risk communication according to the US National Research Council can be successful "if it raises the level of understanding of relevant issues or actions for those involved and satisfies them that they are adequately informed within the limits of available knowledge."[4] The focus of this paper is the examination of several conditions under which grey literature is provided by nuclear regulators and industry together with other categories of information constituting a body of knowledge intended to inform the public in the context of nuclear risk communication. Four conditions of nuclear risk communication are considered: the sociological concept of *risk society*, the issue of public awareness, the role of public trust and knowledge in relation to public attitude and nuclear power.

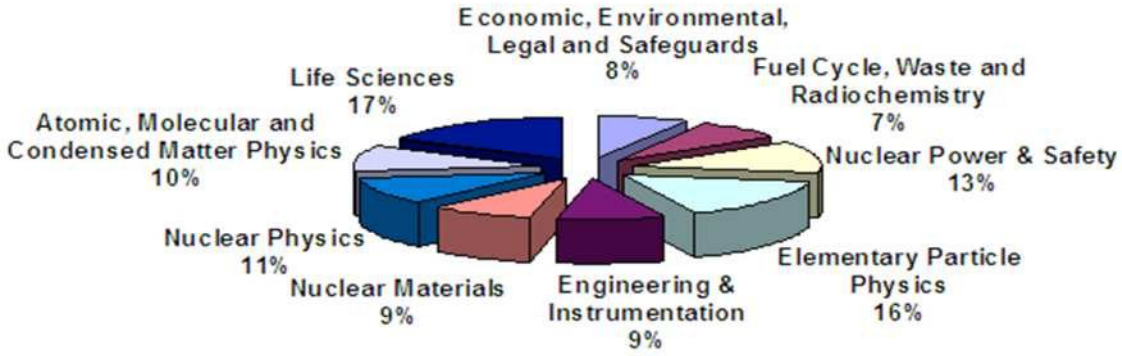
Risk society

In his book entitled *Risk society. Towards a new modernity* the German sociologist Ulrich Beck argues that the developed world finds itself in a transitional phase, evolving from an industrial society towards a risk society.[5] In his view creating some much wealth also means creating risks due to the growing complexity of a technology driven society. The dangers involved concern the effects of modernization and consumption patterns of prosperous societies resulting in risks on different scale and type like those of toxic waste, global warming and nuclear radiation.

In the present situation no scientific agreement or a shared public understanding on the definition of risk exists.[6] In general risk can be defined as: the probability of harmful consequences, or expected loss of lives, people injured, economic activity disrupted (or environment damaged) resulting from interactions between natural or human induced hazards and vulnerable conditions.[7]
To avoid or prevent risks developed societies introduced risk-institutionalized methods and activities for monitoring perceived risks as risk management practices. This development is defined by Beck as a 'reflexive modernization of risk'.[8] As a result risk management as an organizational discipline has grown considerably in the last decades resulting in consideration of risks at a global level with events as Chernobyl, Fukushima, global warming, climate change, terrorism and hazards like tsunamis. These experiences have led to the globalization of the meaning risk and as such have intensified the public awareness for risks. The emergence of the risk society is based on a growing influence of technology. In these circumstances science is becoming more and more important as a tool to inform the public and public policy. Since the formulation of the Millennium Development Goals (MDG)[9] in 2000 by the United Nations to address global problems related to hunger, poverty, and other emergency situations, science explicitly is expected to fulfill an essential contribution in achieving the objectives in order to improve human welfare.[10]
Within this perspective science pursues social objectives and puts itself into the position of being a public interest science. Another relevant example for making the connection between science and its use for public understanding is the way the US Federal Emergency Management Agency (FEMA) makes the connection between science and the objectives to improve the capability to respond to disasters by providing technical guidance and tools targeted for mitigating multi-hazard events by means of The Building Science Branch.[11] This organizational unit takes care of publishing different types of grey literature like the Security Risk Management and Natural Disaster Series Publications.

The issue of public awareness

Already in 90s, during the International Decade for Natural Hazard Reduction, there was a strong focus on raising public awareness in combination with information action towards natural hazards with the objective to reduce vulnerability and making communities more resilient.[12] Policies of raising public awareness about nuclear power are related to technological hazards and as such can be observed sociologically as failures of techno-social systems.
After the accidents at Three Miles Island and Chernobyl there was a building stop of nuclear power plants in a significant number of countries, but as it turned out, the present issue of climate change and energy security was a cause for a 'nuclear renaissance'[13] in search of the ideal energy mix. The Fukushima accident however didn't trigger a cancellation for new build programs. Most countries with a nuclear installed base deem nuclear energy necessary in the future for building energy security. This puts risk communication in a strategic position as a tool to inform the public as the nuclear industry and regulators are in need of public support. From this perspective there is a need to understand the values, preferences and the way the recipients of risk communication comprehend risks in relation to the knowledge they already possess in dealing with vulnerabilities.[14]
The public is especially dependent on media coverage as this is the dominant force in shaping the public perception of nuclear power.[15] The American sociologist Gamson provides a model in which the discussion in the media about nuclear power is perceived as an issue culture produced by general audience media. The discussion on nuclear power is wrapped up in what is called a media package that gives meaning to an issue as how to think about it and reasoning devices that justify what should be done about it.
The question is how can grey literature as scientific information contribute to public awareness?
One of the most important instruments available in the knowledge network of the peaceful uses of nuclear science and technology is the International Nuclear Information System (INIS) operated by the International Atomic Energy Agency (IAEA) in Vienna, the leading information system in the nuclear field.[16] It contains 3.3 million bibliographic descriptions and over 600.000 full text documents. Since early April 2009 INIS has been free to Internet users. It covers a broad range of subjects and issues.



The free availability of scientific information on nuclear power by the Open Access principle from INIS is one step in a positive direction. It cannot to be expected to answer all questions related to risk issues. A higher transparency may improve risk communication but the scientific and technical information is difficult to understand for a large audience as a study in 2008 from the EU Commission has shown.[17] There is a challenge to make the category of grey literature related to nuclear power more understandable for citizens. This doesn't mean any downgrading of quality of information but it shows the need for scientists to be aware of how to communicate to the public.

Both scientists - including scholars of grey literature - and the media have a role and responsibilities in the chain of information that connects author to recipient. Journalists and other public communication specialists need to understand how to frame the technical and social issues related to the use of nuclear power because the public - as the recipient at the end of the chain of information - is a heterogeneous group with different perceptions. These perceptions are based on variables like gender, education difference, and difference in using information channels. From this perspective science has to adapt itself to levels of public communication to be adequate. Social media can be used as a tool for risk communication. The disaster sociologist Quarantelli considers it important that public awareness strategies focus on group levels rather than individuals because of its effectiveness.[18] Within this group level it is important to identify key groups as a reference group for those without knowledge. With the use of social media it is relative easy to focus on communities and groups as a means to mobilize scientific knowledge. However there can be no effective risk communication related to nuclear power if the key ingredient, e.g. public trust is missing.

Public trust

The Nuclear Energy Agency Committee on Nuclear Regulator Activities established[19] in 2000 the Working Group on Public Communication (WGPC)[20] to promote public trust internationally at a time that a majority of the general public opinion in the US and in Europe was opposed to the use of nuclear power. According to L.J. Keen, president of the Canadian Nuclear Safety Commission the regulator image had to be built on public trust and on "credible, unbiased – and frank – source of information".[21] Also it was advised that regulators needed to integrate social concerns into their risk assessments.[22] With this statement WPGC anticipated on research that showed that perceived risks are influenced by the lasting image of historical events and the possibility of future events. The WPGC is also well aware that satisfying public opinion claims as trust and credibility is a priority for nuclear regulators and the associated industry. It is their task for providing the necessary risk information to deal with hazard situations, which has become more difficult after Fukushima. It is the randomness of such nuclear accidents that triggers public fear[23] not the degree of probability that accidents will happen.

A report on Europeans and nuclear safety (2009) found that a large portion of 47 % of the European public opinion considers nuclear risk underestimated.[24] The report states that European citizens are extremely conscious of the importance of safety and protection. It also shows that the influence from past events like Chernobyl as a destructive force of nuclear power had a lasting effect on the discussion on the use of nuclear power plants in the future.[25]

In a Eurobarometer 2008 survey from the European Commission European citizens' attitudes to nuclear waste were examined in the 27 Member States.[26] The knowledge from Europeans how to manage nuclear waste appeared to be limited. It shows that respondents in countries with nuclear power were generally more knowledgeable than those from countries without nuclear power and also tend to be more in favour

of nuclear energy. Most trusted is information from independent sources about the way radioactive waste is managed. Scientists are perceived as the trustworthiest source of information about nuclear waste management (40%) together with non-governmental organizations (38%), followed by international organizations working on peaceful uses of nuclear technology (32%). Information given by the nuclear industry is mainly trusted by respondents in countries that have operational nuclear power plants.
Recent public opinion data however indicates that in general a cultural gap exists between scientists and the public. A number of scholars believe that public confidence in science nowadays is weakening, as the public is skeptic about the ability of scientists and technology "to identify and solve society's fundamental challenges".[27] The fact that not only science itself but also the genre of grey literature is being questioned, as in the case of the publication of the Intergovernmental Panel of Climate Change 2007 report on global warming, is significant.[28]
A number of statements in this report were legitimately challenged as they were based on non-peer reviewed literature. However some statements on the nature of grey literature were incorrect. One scientific journal noted that a statement in the IPCC report, related to the melting of the Himalayan glaciers, "was drawn from non-peer-reviewed findings, known as "gray literature".[29] In another situation an American member of Congress noted the "IPCC's use of "gray literature" to promote a particular objective" indicating it as "non-peer reviewed literature".[30]
These observations about the IPCC report harms the sensitive position of science and grey literature as it becomes politicized instead as it was intended to de-politicize the issue of climate change. Also in combination with observations from a 2009 Pew Research Center report that indicates that 49% of adults in the United States agree that human activity is producing global climate change compared to 84% of scientists makes the gap between the public and science more visible.[31]
The IPCC issue serves as an example how complicated it is to position grey literature in the context of risk communication. Public trust in science is being jeopardized as the public awareness about climate change is experienced as high-profile case and of great concern just as it is in the case of nuclear power.[32]

Nuclear power, the role of knowledge and public attitude

The 1950's and early '60s were characterized by an almost unlimited optimism about the use of nuclear energy in the sense that "electric energy would be too cheap to meter".[33] The reasons for supporting nuclear power in the beginning of the 70s in Europe and the US were the uncertainty related to the first energy crisis in 1973 and the wish to be less dependent on oil.[34] Gradually there was less support for nuclear power. The support dwindled after the accident of Three Mile Island in 1979. Also the fear for a potential nuclear war caused an increase in opposition. These developments challenged the nuclear industry to change public attitude toward more pro-nuclear. In the US national advertising campaigns and other attempts by the nuclear industry and the Committee for Energy Awareness tried to label the use of nuclear power with an idea of technological progress in relation to economic growth. This approach proved not to be very successful.[35].
The first model the nuclear industry used in dealing with the public's perception on nuclear power was the linking of knowledge and public attitude to nuclear risk perception. In the 70's the technical community was convinced that citizens in general were poorly informed on the issue of nuclear power. In their view opposition resulted from ignorance so the more knowledge the public had about nuclear energy the more public confidence would be generated.[36] This hypothesis has been tested in various studies. Kuklinski, Metlay and Daniel found that a majority of the people lacked knowledge about nuclear power[37]. Research dating from the 70s indicates that informed citizens supported nuclear development more than uninformed citizens did.[38] Kasperson et al. concluded in 1980 that there is no decisive empirical evidence that more education and information will change the opinions of those having little or no knowledge about nuclear energy.[39] A European Commission survey from 2008 comes to the conclusion that those with little education are more likely to be opposed to nuclear energy.[40] In general they are not convinced about the manageability of nuclear power.[41]
These studies show that it is not clear how people decide what reasons there are for a calculated choice in favour of or against the use of nuclear energy. Core values like the orientation to technology or political views seem to provide the key for making a choice. Knowledgeable citizens evaluate the issues around nuclear energy more in ideological terms as uninformed are less inclined to see the wider range of consequences. Both groups decide in very different ways.

There are some important other insights. Tichenor, Donohue and Olien observed that the use of information from different channels into society has an uneven effect on knowledge of people.[42] Those who were able to reach a higher level of education show the ability to get a to a more knowledgeable level than those with fewer years of schooling. They have better reading abilities. This leads to "knowledge gaps" between the high and low educated. Providing more information is going to widen the knowledge gap that exists between people with low and high levels of education. In risk communication it is significant to provide for information that can be comprehended by the least educated.[43] Yim and Vaganow report contradictory effects of providing more information towards public acceptance. Only in the situation that provided information matches the core values of the recipients and there is a relation of trustworthiness information provision can be effective.

The accidents at Three Mile Island and Chernobyl nuclear power caused a wave of negative information that was 'social amplified' as the public response was based on fear and distrust.[44] The Chernobyl accident did affect the European public opinion stronger than the American public opinion that already was very negative toward nuclear power. The lack of information emergency plans made it impossible to keep the public and media informed in an adequate way. E.g. in France and Belgium no timely information was provided about possible negative effects of Chernobyl. At that time this was in sharp contrast with the views of some officials from the US and Soviet Union who believed that a serious accident was impossible. This attitude prevented a clear judgment on the situation how much information is needed for estimating risks. The situation in Chernobyl and the information flow was even worse where authorities responded too slow and failed in providing information on radiation levels. The lack of openness causes suspicion with the press and public. Despite the general distrust at the time perception of risks have shifted. From 2005 we can witness an increasing support for nuclear energy[45] as it is connected as a potential solution to the energy security situation and the issue of climate change. This choice by public opinion is based on the notion that nuclear energy has disadvantages but can play a supplementary role in the energy mix to battle climate change. This has been coined 'the reluctant acceptance discourse'.[46] Also this resembles the Gamson model of a media package, a discours strategy reframing the necessity of nuclear power. In the debates since the late 1990s energy security and climate change have been justifications for what has been dimed as a nuclear renaissance with the relativisation of nuclear risks in the light of a dangerous climate change.[47]

The role of grey literature

Fukushima has made the public aware it needs to be better informed by the national Nuclear Regulatory Organizations (NRO). In 2010 as a whole 49% of the European citizens felt 'not very well informed' and 25% say they are not informed at all regarding safety issues.[48] There is a need for more information on topic of nuclear safety. 63% share the conclusion that the general audience media do not offer enough information to make it possible for citizens to draw their own conclusions about risks and safety involved.

This shows the need for greater transparency on the issue of nuclear safety but also the need enhance the availability of grey literature and other types of scientific information through a wider variety of dissemination channels. Another need is the establishing of a legal framework for monitoring international transparency in making different types of information available. The WPGC surveyed member countries on transparency practices. Since 2006 all members of the Organisation for Economic Co-operation and Development (OECD) have Freedom of Information Acts that give the public rights of access but transparency can be different in these countries. The greatest problem is to balance openness and need for security. Other challenges lie in satisfying the public in the need for information. Social media can play an important role as the Fukushima accident proves. On March 11, 2011 when the Fukushima nuclear reactor blew its top Twitter was used as data source for crisis communication after more than 500 million tweets about Fukushima's radioactivity.[49] All tweets with information about *Radioactivity*, *Pollution cloud*, *Fukushima* and similar topics were monitored all over the world through the Google Realtime system and the origin of the tweets geographically located. Can we consider this as an example of sharing grey literature by new information channels? Anyway it shows the enormous potential of effective risk communication through social media in relation to nuclear power. Another interesting observation is made by the WPGC that social media have increased the difficulty for Nuclear Regulators to manage crisis communication in terms of speed and accuracy. Despite Web 2.0 innovations traditional outlets will be

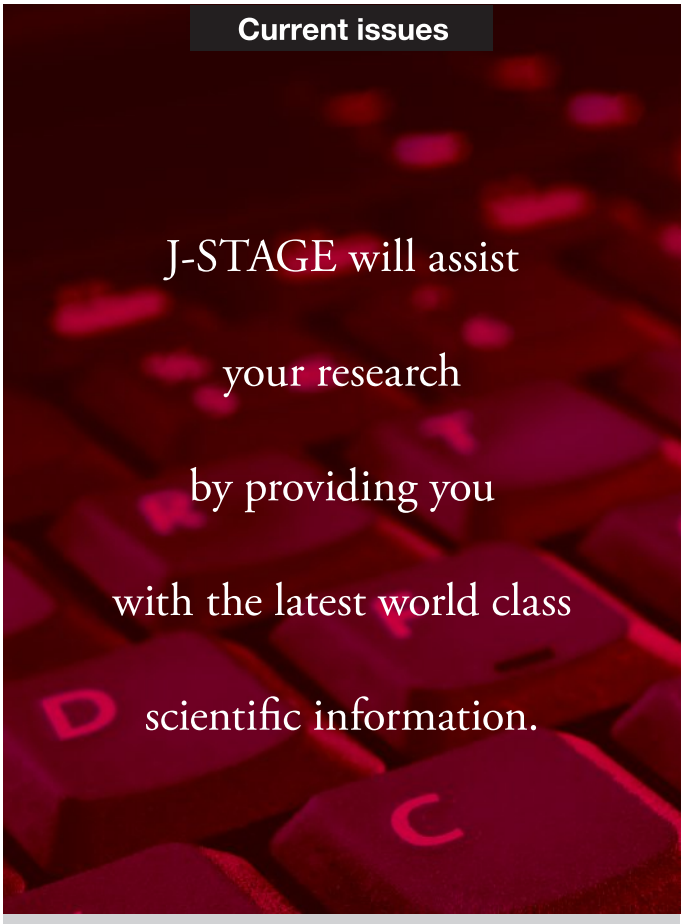
maintained. All NRO's have websites and will use them in the event of a nuclear crisis supplying official information. In the case of an emergency 10 out of 17 countries will continue to use these regulary websites to communicate with the public, updating them with official information and public service messages related to the emergency.

References

[1] Statement IAEA Contribution to International Peace, Security and Prosperity, 6 October 2011, <http://www.iaea.org/newscenter/statements/2011/amsp2011n026.html>.
[2] National Research Council, *Improving risk communication,* National Academic Press, Washington DC, 1989.
[3] OECD/Nuclear Energy Agency, *Building, Measuring and Improving Public Confidence in the Nuclear Regulator: Workshop Proceedings: Ottawa, Canada, 18-20 May 2004*, Nuclear Regulation, OECD Publishing, 2006.
[4] National Research Council (1989).
[5] Beck, Ulrich, *Risk society. Towards a new modernity*, London : Sage, 1992.
[6] Luhmann, Niklas, *Risk: a sociological theory*, New York, A. de Gruyter, 1993. See chapter 1: the concept of risk.
[7] Many definitions of the term risk exist. A more scientific and less descriptive definition would perceive risk as the likelihood of an event multiplied by a series of consequences ranging from mild to catastrophic (risk = probability x consequence).
[8] Beck, Ulrich, Anthony Giddens, and Scott Lash. *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order*. Stanford, Calif: Stanford University Press, 1994.
[9] See: <http://www.un.org/millenniumgoals/>.
[10] The International Council for Science states that: "Each nation must have a source of independent, credible, and timely advice to government policymakers and the public on critical issues involving science and technology."
[11] See: <http://www.fema.gov/rebuild/buildingscience/>.
[12] A/54/497 International Decade for Natural Disaster Reduction: Successor Arrangements. Report of the Secretary-General, 1 November 1999. Available at: <http://www.eird.org/fulltext/SG-report/SG-report-nov1999-eng.pdf.> Also
See: Resolution adapted by the General Assemblee, A/55/L2, United Nations Millennium Declaration, 18 sept. 2000. Par. III, last item. Later the United Nations Millennium Declaration stated explicitly the access to information and communication technology as one of the instruments for raising awareness towards disasters. "To ensure that the benefits of new technologies, especially information and communication technologies, in conformity with recommendations contained in the ECOSOC 2000 Ministerial Declaration, are available to all." <http://www.un.org/millennium/declaration/ares552e.pdf>.
[13] T.W. Wood, W.L. Johnson, B.M. Parker*, Economic globalization and a nuclear renaissance*, October 2001. Prepared for the U.S. Department of Energy.
[14] Paul Slovic, *Perceived risk, trust and democracy*, in: Risk analysis, vol. 13 n. 6, 1993, pp. 675-682.
[15] Gamson, William A., Andre Modgliani, *Media discourse and public opinion on nuclear power: a constructionist approach*, in: American journal of sociology, vol. 95, n. 1 (1989), pp. 1-37.
[16] The International Nuclear Information System (INIS) is available for free <http://www.iaea.org/inis/>. Another source is Nuclear Science Abstracts, a collection of international nuclear science and technology literature for the period 1948- 1976. <http://www.osti.gov/inisnsaview/>. The INIS database started in 1970.
[17] European Commission, *Qualitative study on the image of science and the research policy of the European Union. Study conducted among the citizens of the 27 member states*. Pan-European report, October 2008.
[18] Wisner, B., Disaster *Preparedness and response: why is the phone off the hook*. Invited paper for the European Telecommunications Resilience & Recovery Association Inaugural Conference (ETR2A), Newcastle-upon-Tyne, UK, 11-13 June 2003
[19] The Committee on Nuclear Regulatory Activities (CNRA) is a part of The Nuclear Energy Agency (NEA) a specialised agency within the Organisation for Economic Co-operation and Development (OECD)
The CNRA is responsible for the program of the Agency concerning the regulation, licensing and inspection of nuclear installations with regard to safety. <http://www.oecd-nea.org/nsd/cnra/cnra-mandate.html>.
[20] <http://www.oecd-nea.org/nsd/cnra/wgpc.html>
[21] OECD/Nuclear Energy Agency, p. 14. L.J. Keen, president of the Canadian Nuclear Safety Commission stated: "Public confidence must mean that the regulator's image is one of being a credible, unbiased – and frank – source of information. It means having a willingness to acknowledge uncertainties and the limits of our technical understanding."
[22] Ibidem, p. 16
[23] Wolters, M., M. Haufe, R. Wendte, J. de Jonge en P.Merkx, *'Een hopelijk tijdelijk noodzakelijk kwaad' - Publiekesperceptie kernenergie, rapportage kwalitatief onderzoek*. Amersfoort : SmartAgent Comapny, 2009.
[24] EU/Directorate General Communication, *Special Eurobarometer 324. Europeans and nuclear safety*, March 2010
[25] Koopmans, R. en J.W. Duyvendak, *The political construction of the nuclear energy issue and its impact on the mobilization of anti-nuclear movements in Western Europe*, in: Social problems, vol. 42, n. 2 (1995), pp. 235-251.
[26] European Commisson, *Attitudes toward radioactive waste*. Report, Special Eurobarometer 297, June 2008.
[27] Gordon Gauchat, *The cultural authority of science: public trust and acceptance of organized science*, in: Public Understanding of Science, vol. 20, nr. 6 (2011), pp. 751-770; Allum, N., Sturgis, P., Tabourazi, D. and Brunton-Smith, I. , *Science knowledge and attitudes across cultures: a meta-analysis,* in: Public Understanding of Science vol. 17, n.1, 2008, pp. 35–54.
[28] See: <http://ipcc.ch/>
[29] D. Biello, *Shades of "gray literature": How much IPCC Reform is needed?,* in: Scientific American, 2010, dd. 30 Aug, p. 24.
[30] Letter from Rep. Paul Broun, US House of Representatives, Committee on Science and Technology to Ban Ki-Moon, Secretary-General, United Nations, dd. February 2, 2010.

[31] Leiserowitz, A., E. Maibach, C. Roser-Renouf, N. Smith, *Climate change in the American mind: Americans' global warming beliefs and attitudes in June 2010*. Yale University and George Mason University, New Haven : Yale, Project on climate communication. URL: http://environment.yale.edu/climate/files/ClimateBeliefsJune2010.pdf; Pew Research Center, Fewer Americans see solid evidence of global warming.
URL:< http://www.people-press.org/2009/10/22/fewer-americans-see-solid-evidence-of-global-warming/>
[32] W. Poortinga, N.F. Pidgeon, L. Lorenzoni, *Public perceptions of nuclear power, climate change and energy options in Britain: summary findings of a survey conducted during October and November 2005*. Technical Report (Understanding Risk Working paper 06-02). Norwich: Centre for Environmental Risk, 2006.; Eurobarometer, Attitudes on issues related to EU Energy Policy.
[33] The phrase is from Lewis Strauss, former chairman of the Unites States Atomic Energy Commission. It is from a speech held in 1954 before an audience of the National Association of Science Writers.
[34] Rosa, E.A. and R.E. Dunlap, *Poll trends: Nuclear power: three decades of public opinion*, in: The public opinion quarterly, vol. 58, no. 2 (1994), pp. 295-324
[35] Bert Useem and Mayer N. Zald, *From pressure group to social movement: efforts to promote use of nuclear power*., in: Social movements in an organizational society, New Brunswick, New Jersey : Transaction Publishers, 5th ed., 2009, pp. 273-288.
[36] Kasperson, Roger E, Gerald Berk, David Pijawka, Alan B. Sharaf, James Wood, *Public Opposition to Nuclear Energy: Retrospect and Prospect*, in: Science, Technology, & Human Values, Vol. 5, No. 31 (Spring, 1980), pp. 11-23
[37] Kuklinksi, James H., Daniel S. Metlay, W.D. Kay, *Citizen knowledge and choices on the complex issue of nuclear energy*, in: American Journal of Political Science, vol. 26, no. 4 (1982), pp. 615-642.
[38] Hensler, Deborah R., and Carl P. Hensler. *Evaluating Nuclear Power: Voter Choice on the California Nuclear Energy Initiative*. Santa Monica, CA.: Rand, 1979; Melber, Barbara D., and Stanley M. Nealey. *Nuclear Power and the Public: Analysis of Collected Survey Research*. Seattle, Wash: Battelle Memorial Institute, Human Affairs Research Centers, 1977.
[39] Kasperson et al. (1980).
[40] Eurobarometer 2008.
[41] Pligt, J. van der, *Cognition and affect in risk perception and risky-decision making*, in: C. von Hofsten en L. Bäckman (red.), Psychology at the turn of the millenium. Social, developmental and clinical perspectives; volume 2 (pp. 247-270), Hove : Psychology Press, 2002.
[42] Tichenor, Philip J., George A. Donohue, and Calice N. Olien, *Mass Media Flow and Differential Growth in Knowledge,* in: Public Opinion Quarterly 34 (Summer), 1970, pp. 159–70.
[43] Gordon Gauchat (2011)
[44] David M. Rubin, Communicating risk: the media and the public. How the news media reported on Three Mile Island and Chernobyl, in: Journal of Communication, vol. 37, no. 3 (1987), pp. 42-57.
[45] Eurobarometer 2010.
[46] Adam Corner et. al., Nuclear power, climate change and energy security: exploring British public attitudes
[47] Teravainen, T., M. Lehtonen, M. Martiskainen, *Climate change, energy security and risk-debating nuclear new build in Finland, France and the UK*, in: Energy policy, vol. 39, 2011, pp. 3434-3442
[48] European Commisson, *Europeans and nuclear safety*, Special Eurobarometer 324, 2010.
[49] See:<http://www.socialintensity.org/#fukushima>

Largest collection of e-Journals in Japan



J-STAGE

- All abstracts **FREE** (90% with English)
- More than 570 **PEER-REVIEWED** journals (70% **FREE**)
- **FULL TEXT SEARCH**
- **ADVANCE** online publication
- Personalized management tool – "**My J-STAGE**"

www.jstage.jst.go.jp

Journal@rchive®

- **Over 750,000** articles published by Japanese societies
- **FREE** from 1st issues
- The oldest goes **back to 1877**
- **FULL TEXT SEARCH**

www.journalarchive.jst.go.jp

Japan Science and Technology Agency
ELECTRONIC JOURNALS DIVISION
DEPARTMENT OF ADVANCED DATABASES
contact@jstage.jst.go.jp

Management of Obsolete Grey Literature in Engineering Research Institutions

C.P. Ramasesh - University of Mysore, India.
N. Chowdappa - BMS College of Engineering, Bangalore, India
L. Usha Devi - Bangalore University, India

Abstract

This paper depicts the findings of the survey research covering 65 engineering institutions which are recognised as research centres in the state of Karnataka, India. Responses have been sought from research supervisors and research scholars as to the extent of the use of obsolete grey collections. Opinions have also been captured from the chief librarians of the 65 research institutions as regards the ways of managing the obsolete grey literature. The following are the areas covered in the study in respect of obsolete grey resources available in these 65 research libraries.

- *Demand for GL in engineering sciences and technology*
- *Grey Literature collection in English and Foreign languages*
- *Process of weeding-out of obsolete grey literature*
- *Reasons for weeding out of old GL collections*
- *Extent of the use of old GL collections*
- *Weeding-out vis-à-vis Relegation of obsolete GL collections.*

The study was undertaken during 2007-2010 with the objective of understanding the pattern of obsolescence and the method followed in the libraries to manage the old GL collections in the engineering research institutions. The findings also project the practical methods followed by various libraries as to the weeding out process. Importance of old collections of GL for research vis-à-vis weeding out process adopted by the libraries have been depicted in the study based on the opinions exposed by 1270 researchers, which account for 84.6% of the total population. The summary or core of findings eventually shows that 66.4% of the research supervisors and research scholars hold the view that old collection of GL is of vital source for their research and has to be retained along with active collection in the libraries.

Keywords: Grey Literature, Obsolescence; Grey Literature, Weeding out; Grey Literature, Engineering Institution Libraries

Introduction

Grey literature is gaining importance in the sphere of research and development in engineering sciences and technology. Grey resources are produced in limited number of copies and are circulated in a limited circle of users. However, the researchers rely upon them to a greater extent. This trend of producing limited number of copies and not marketed through normal book selling channels makes the researchers rely upon regional libraries (RL).

In the present study, an attempt is made to study the use pattern of GL, especially the use of old collections of GL available in the engineering institutions in the State of Karnataka. 65 institutions have been recognized as research centres by the Visvesvaraya Technological University. The study also aims at identifying the necessity of weeding out process followed in these libraries and the method of maintaining the weeded out / relegated GL collections. Questionnaires were administered to the following respondents to obtain relevant data/ opinion on the topic of study.

1. The Chief Librarians of the 65 engineering institutions which are recognized as research centres.
2. 340 faculty members who are the research supervisors engaged in research activity in these 65 institutions and
3. 930 research scholars who are actively engaged in either full time or part time research, leading to the award of doctoral degree.

The percentage of total response in the present study is 84.6.

Table 1 here below depicts 9 disciplines identified for research and the number of institutions engaged in research in these disciplines.

Table 1: Institutions Engaged in various Disciplines of Research

Sl. No.	Disciplines of Research	No. of Institutions	Percentage
01	Civil Engineering	26	19.6
02	Mechanical Engineering	31	23.3
03	Electrical Engineering	17	12.8
04	Electronics and Comm. Engineering	18	13.6
05	Computer Science and Engineering	15	11.2
06	Chemical Engineering	06	4.5
07	Textile and Silk Technology	03	2.2
08	General Science	06	4.5
09	Business Administration	11	8.3
	Total	**133**	**100**

Among the disciplines, mechanical engineering and civil engineering departments are more in number representing 23.3% and 19.6% respectively. However, textile and silk technology subject is offered for research only in 3 institutions representing 2.2%.

Table 2: GL collection in the Libraries

Sl. No.	Category of GL	Number	Percentage
01	Theses	793	0.6
02	Dissertations	79845	60.6
03	Institutional Publications	9499	7.2
04	Trade Literature	3021	2.3
05	Technical Reports	22181	16.8
06	Proc. of CSW	16456	12.5
	Total	**131795**	**100%**

Table 2 shows the collection of GL in the 65 engineering institutions. Six categories of GL is identified for the study and the table depicts that the total collection in these institution libraries is 131795. The collection of dissertations being the highest in number account for 60.6%. Technical reports and proceedings of conferences, seminars and workshops are in considerable percentage representing 16.8% and 12.5% respectively in the total holdings of the grey literature covering the 65 institutions. The theses volumes are 796 in number representing just 0.6% because prior to the year 2003-04 the theses were submitted to the state universities for the award of doctoral degree. Since the year 2004 the theses volumes have been submitted to VTU. Hence, the number is the least among the categories of GL considered for the study.

Table 3: Libraries Possessing GL Collection in Foreign Languages

S. No	Libraries Possessing GL in Foreign Languages	Govt. Institutions	Aided Institutions	Private Institutions	Total
01	Positive Response	2 33.3%	Nil	9 18.0%	11 16.9%
02	Negative Response	4 66.7%	9 100%	41 82.0%	54 83.1%
	Total	6 100%	9 100%	50 100%	65 100%

Table 3 depicts the number of libraries which possess GL collection in foreign languages. 11 institutions possess GL in foreign languages which account for 16.9%. Among the government institutions, 33.3% possess GL in foreign languages. From this it is clear that GL collection in foreign languages is not given priority in majority of the institutions. It is only Indian Space Research Organization and Indian Institute of Astro-physics which come under government organization possess adequate and relevant collection of GL in foreign languages. In rest of the institutions, the GL in foreign languages is not the felt need. There are 9 private institutions which possess GL in foreign languages. However, the relevancy and up- datedness is not appreciated.

Table 4: User-wise Frequency of Access and use of GL

S. No.	Frequency	Research Supervisor	Research Scholars	Total
01	Most frequently	20 (5.9%)	67 (7.2%)	87 (6.8%)
02	Frequently	189 (55.6%)	519 (55.8%)	708 (55.8%)
03	Moderately	98 (28.8%)	279 (30.0%)	377 (29.7%)
04	Occasionally	31 (9.1%)	64 (6.9%)	95 (7.5%)
05	Not at all	2 (6.0%)	1 (0.1%)	3 (0.2%)
	Total	340 (100%)	930 (100%)	1270 (100%)

Table 4 projects the extent of access and use of GL collection available in the engineering institutions. It is really encouraging to note that 62.6% of the researchers use GL collections either frequently or most frequently. Just 7.5% of the researchers have expressed that they need GL occasionally, only 3 respondents among 1270, representing mere 0.2% opine that they seldom need grey literature for their research. From this it can be concluded that a large segment of respondents rely upon GL for research purposes.

Table 5: Response on the Use of Old Collection of GL

S. No.	Response on use of old GL collection	Research Supervisor	Research Scholars	Total
01	Positive, yes, useful	277 (81.5%)	661 (71.1%)	938 (73.9%)
02	Negative, No. not useful	63 (18.5%)	269 (28.9%)	332 (26.1%)
	Total	340 (100%)	930 (100%)	1270 (100%)

As regards use of old collection of GL in the engineering institutions, it is highlighting that 81.5% of the research supervisors and 71.6% of the research scholars access old collections. Among the entire population of respondents, 73.9% have offered positive response stating that old collection is of use for research. A small segment comprising 26.1% opines that old collection of GL is not at all useful. From this it can be inferred that old collections of GL is useful according to the opinion expressed by the researchers.

Table 6: Extent of Obsolete GL collections in the Libraries

Category of GL	Extent of Obsolescence				
	100%	75%	50%	25%	Nil
Theses	5	6	21	25	8
Dissertations	4	18	20	11	12
Institution Publications	4	15	19	17	10
Trade Literature	7	21	21	13	3
Technical Reports	2	7	24	22	10
Proc. of CSW	3	7	18	25	12

Table 6 furnishes the extent of obsolete collection according to the librarians working in the 65 libraries of the research institutions. The data in the table depicts that dissertations and trade literature collection are outdated to a considerable extent. 42 librarians hold the view that dissertations available are outdated in the range between 50%, 100%, and 49 librarians are of the opinion that the trade literature collection is out dated in the range between 50% and 100%. It is also clear that theses, technical reports and proceedings of conferences, and seminars and workshops are not outdated in the higher range according to the contention of the respondent librarians. Therefore, it is worthwhile to weed out/ relegate dissertations and trade literature from time tot time in order to make provision for easy access to the active collection (recent additions).

Table 7: Weeding out / Relegation of Obsolete GL

S. No	Response on weeding out	Govt. Institutions	Aided Institutions	Private Institutions	Total
01	Positive Response	Nil	4 (44.4%)	9 (18.0%)	13 (20.0%)
02	Negative Response	6 (100%)	5 (55.6%)	41 (82.0%)	52 (80.0%)
	Total	6 100%	9 100%	50 100%	65 100%

Table 7 projects response of librarians on the weeding out process. 13 librarians representing 20.0% have offered positive response stating that they periodically weed out old / obsolete collections of GL. The remaining 52 librarians representing 80.0% have responded negatively. Again, among the librarians of the aided institutions, 44.4% weed out old GL collections periodically. From this it is clear that considerable percentage of institutions which are getting government aid, weed out the old collections of GL.

Table 8: Weeding out of Obsolete GL
(as per year of establishment of the institutions)

S. No	Response on weeding out	Year of Establishment			
		1980	1981-2000	2001-2010	Total
01	Positive Yes weed out	9 (32.1%)	2 (11.8%)	2 (10.0%)	13 (20.0%)
02	Negative Don't weed out	19 (67.9%)	15 (88.2%)	18 (90.0%)	52 (80.0%)
	Total	28 (100%)	17 (100%)	20 (100%)	65 (100%)

Table 8 shows response of librarians on weeding out process. The data in the table is furnished is as per the year of establishment of the institutions. It is clear from the table those 9 libraries which were established prior to 1987 weed out old collection of GL periodically. The institutions which were established prior to 1981 possess large number of GL collections and hence the need to weed out from time to time. As the recently established institutions possess limited number of GL collection, the necessity of regular weeding out process does not arise.

Table No. 9 : Opinion on Weeding out of Obsolete GL

S. No	Response	Govt. Institutions	Aided Institutions	Private Institutions	Total
01	Yes, weed out	12 (41.4%)	105 (28.4%)	302 (34.7%)	419 (33.0%)
02	No. Don't weed out	17 (58.6%)	266 (71.6%)	568 (65.3%)	857 (67.0%)
	Total	29 (100%)	371 (100%)	870 (100%)	1270 (100%)

As regards opinion of researchers on the weeding out process, a large segment representing 67.0% has responded negatively. According to them the old collection of GL is of value for research. 33.0% have given positive response favouring periodical weeding out process. The researchers from government institutions are in favour of regular weeding out process. The respondents from government institutions who have offered positive response account for 41.4%. On the other hand, the respondents from aided institutions who have offered positive response account for just 28.4%. Hence, majority of the researchers opine that old collections of GL need to be retained and they should not be either relegated or weed out as they also form vital source of information for research.

Table 10: Periodicity of the Weeding out Process

S. No	Periodicity	Govt. Institutions	Aided Institutions	Private Institutions	Total
01	Annually	Nil	Nil	1 (11.1%)	1 (7.7%)
02	Once in 3 years	Nil	Nil	3 (33.3%)	3 (23.1%)
03	Once in 5 years	Nil	1 (25.0%)	2 (22.3%)	3 (23.1%)
04	Once in 10 years	Nil	3 (75.0%)	3 (33.3%)	6 (46.1%)
	Total	Nil	4 (100%)	9 (100%)	13 (100%)

Table 10 depicts periodicity of the weeding out process in the 13 institutions. One of the institutions representing 7.7% weed out obsolete GL collection annually. 23.1% weed out once in 3 years. Another segment of 23.1% weeds out once in 5 years. Further, a large segment of the institutions representing 46.1% weed out once in 10 years. From this it is clear that the necessity of weeding out of old collection of GL is not noticed among the large number of libraries of engineering institutions.

Table 11: Reasons for Weeding out of GL collections

S. No	Reasons for weeding out of GL	Govt. Institutions	Aided Institutions	Private Institutions	Total
01	Lack of space	Nil	4 (100%)	8 (88.5%)	12 (92.3%)
02	Out dated contacts	Nil	2 (50%)	1 (11.1%)	3 (23.0%)
03	Unused by the members	Nil	3 (75.0%)	1 (11.1%)	4 (30.7%)

With reference to the reasons for the weeding out process, 20% of the librarians quote the reason of lack of space in their libraries, 23.0% state the reason that the contents of the old volumes of GL are out dated. Yet another segment of respondent librarians representing 30.7% opines that the old collections are

unused by the library members. Yet another reason expressed by the librarians is that there is need to make provision for browsing active collections (recent additions) in the library, which account for 72.3%. Therefore, a large majority of the librarians opines that weeding out is on account of making adequate space for browsing active collections of grey literature.

Table 12: Response on Scatter of GL at Different Places

S. No.	GL is scattered and difficult to trace	Research Supervisor	Research Scholars	Total
01	Yes	56 (16.5%)	182 (19.6%)	238 (18.7%)
02	No	284 (83.5%)	748 (80.4%)	1032 (81.3%)
	Total	340 (100%)	930 (100%)	1270 (100%)

Table 12 shows response of researchers on the scatter of GL in the libraries. Only a small percentage of researchers representing 18.7% opine that the GL is scattered in their libraries and difficult to trace. However, a large segment of researchers representing 81.3% find it comfortable as regards the organization of GL and there is no difficulty in locating the needed GL collections. Hence, it can be inferred that a large majority of the respondents find it easy to locate GL in the engineering institute libraries.

Table 13: Training Requirement Projected by Librarians

S. No.	Response on training need	No. of Librarians
01	Positive	57 (87.7%)
02	Negative	8 (12.3%)
	Total	65 (100%)

As regards training requirement of librarians, an overwhelming majority representing 87.7% intends to attend training in order to efficiently manage GL collections. A small portion of librarians amounting to 12.3% does not need any training at all.

Table 14: Extent of Training Needed for Librarians

S. No.	Year of Establishment of Engineering Institutions.	Highly	Moderately	Slightly	Total
01	Till 1980	20 (80%)	4 (16%)	1 (4%)	25 (100%)
02	1981 to 2000	9 (64.3%)	4 (28.6%)	1 (7.1%)	14 (100%)
03	2001 to 2010	12 (6.7%)	6 (33.3%)	Nil	18 (100%)
	Total	41 (100%)	14 (100%)	2 (100%)	57 (100%)

Further, table 14 depicts the extent of training needed by the librarians. However, 20 librarians of the institutions which were established prior to 1981 have expressed that they need intensive training. This group of respondents account for 80% among old institutions. Further, 14 librarians need training to a moderate extent. Therefore, it can be deduced that a large majority of the librarians need intensive training.

Summary of Findings and Recommendations:

- 62.6% of the researchers frequently access GL
- Trade literature and dissertations form major part of obsolete collections in the libraries
- 73.9% of the users also use old collections of GL for research
- 33% of the researchers recommend for weeding out process
- Only 20% of the libraries weed out old GL collections periodically. Majority of these libraries were established prior to the year 1981
- 72.3% of these libraries weed out to accommodate active collection. 20% move obsolete collections to secondary sequence.
- Dissertations of students and trade literature lose value in time and may be weeded out periodically.
- Theses, technical reports and proceedings of CSW form vital source for research. Sharing of the Proceedings of CSW is the felt need among the researchers. These sources may be taken up on priority for digitization and resource sharing programs.

Bibliography

1. Farace, Dominic J (2004). Grey literature: Grey matters in the world of networked information. *Publishing Research Quarterly,* 20 (1):3-91.
2. Gelfand, J. M. (2000). Grey literature poses new challenges for research libraries. *Collection Management,* 24 (1-2):137-147.
3. Gelfand, Julia (2007).Grey Literature: Taxonomies and structures for collection development. *The Grey Journal*, 3 (1): 7-16.
4. Gokhale, P.A. (1999). Information system for Grey literature in India: A model. *Library Science with Slant to Documentation and Information Science,* 36 (4): 229-236.
5. Laloo, Bikika T (2002). *Information needs, information seeking behavior and users*. New Delhi : Ess Ess Publications.
6. Lee, Lafleur and Rupp, Nathan (2005). Making grey literature available through institutional repositories. *The Grey Journal*, 1 (2): 67-72.
7. Manorama, Tripathi and Jeevan, V.K.J (2007): Grey literature archiving in open universities: A model for India. *The Grey Journal*, 3 (2) : 100-106.
8. Needham, Paul A, et.al (2002). *The MAGIC Project : Managing Access to Grey literature Collection,* Final report. London ; The British Library.
9. Ramadevi, Tella (2006). *Access and use of grey literature in social sciences :Problems and prospects.* New Delhi: Ess Ess Publications.
10. Ranger, Sara L (2005). Grey literature in special libraries : Access and use. "Work on Grey in Progress". Proceedings of the Sixth International Conference on Grey Literature, New York,Dec.6-7, 2004: (http://opensingle.inst.fr).preprint. *Publishing Research Quarterly*, 12 (1) : 53-63.
11. Vijayakumar, J.K.et.al,(2006). Introducing electronics theses and dissertations in universities: An Indian perspective. "Open access to Grey Resources*". Proceedings of the Seventh International Conference on Grey Literature,* Nancy, France. 5-6 December 2005. (http://opensigle.inist.fr). Preprint
12. Chowdappa, N. (2010). Organisation and use patterns of grey literature by faculty and researchers in research centres affiliated to Visvesvsraya Technological University: A study. Mysore, University of Mysore (Doctoral Thesis).

Audit DRAMBORA for Trustworthy Repositories: A Study Dealing with the Digital Repository of Grey Literature

Petra Pejšová, National Technical Library, Czech Republic
Marcus Vaska, University of Calgary, Canada

Abstract
The credibility of a grey literature digital repository can be supported by a specialized audit. An audit of credibility declares that the digital repository is not only a safe place for storage, providing access and migrating to new versions of document formats, it also asserts the care components required of a digital repository environment, including the mandate, typology, policy, team, etc. This audit is very important in showcasing to participants and users the quality and safety of the data process.
This paper will present DRAMBORA (Digital Repository Audit Method Based on Risk Assessment), a methodology and tool for auditing a trustworthy digital repository of grey literature. DRAMBORA is an online instrument which helps organizations develop documentation and identify the risks of a digital repository. DRAMBORA is accessible from http://www.repositoryaudit.eu. The paper will also summarize prevailing advantages and disadvantages of DRAMBORA.
The second part of this paper will describe the audit of the National Repository of Grey Literature (NRGL) as a trustworthy digital repository using DRAMBORA as part of creating a digital repository of grey literature in the National Technical Library (NTK). The most important outcome of the audit was represented by the identified risks connected to the repository and potentially endangering its operation, quality, image, and other features. The main principle of the DRAMBORA audit and, at the same time, its main contribution, is its iteration (i.e. its repetition after a certain time period in new conditions when the original risks are reassessed; the measurements adopted for solution are assessed and new risks are identified).

Keywords*: audit, credibility, gray/grey literature, methodology, repository, trustworthiness*

Introduction: Audit for Trustworthy Repositories

> "One of the central challenges to long-term preservation in a digital repository is the ability to guarantee the authenticity and interpretability (understandability) of digital objects for users across time" (Susanne Dobratz and Astried Schoger, 2007)

In our technologically-enhanced environment, managing, preserving, and storing material for posterity is essential, regardless of whether the material in question is a paper file or a digital object (Ambacher, 2007). In fact, efforts at maintaining a stronghold over digital records has been attempted since the 1960s, however, awareness surrounding the true digital repository has only existed for the past decade. This has lead to a number of organizations, most notably the Research Libraries Group (RLG)/U.S. National Archives and Records Administration (NARA) to establish an audit for certifying and enhancing the credibility of grey literature digital repositories. As with any marketing campaign, creating awareness of an initiative and gaining the public's trust is fundamental to ensure success. The Audit Checklist developed by RGL and NARA in 2005 supports this notion with its goal to "develop criteria to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections…a method by which...customers could gain confidence in the authenticity, quality, and usefulness of digitally archived materials" (Ambacher, 2007, p. 2).
Long-term preservation of the material contained within digital repositories functions similarly to the storing of paper documents in a traditional index file within an archive. Ever since institutional repositories arose and began gaining acceptance in the 1990s, efforts at sustaining the material within these storage banks for generations to come have been explored. The first such effort occurred in 1996 when the Task Force on Archiving of Digital Information drew attention to the need for a certification program for the long-term preservation of digital repositories, proclaiming that repositories "must be able to prove that

they are who they say they are by meeting or exceeding the standards and criteria of an independently-administered program for archival certification." (Dobratz and Schoger, 2007, p. 210).
While traditional publishing ventures often result in a considerable time lag between an author's manuscript submission, peer-review by a panel of experts, and subsequent publication in a leading journal within a particular discipline, digital libraries, and in particular digital repositories, allow an author to submit a presentation, thesis, report, etc., as soon as it is written. Further, the author is able to choose from a number of creative commons licenses, maintaining control over his/her data, and deciding how and by whom the data can be accessed (Ambacher, 2007).

Credibility of Grey Literature Digital Repositories

As with any research pursuit, guidelines must be followed and adhered to in order to gain credibility and reputation that a chosen research path is indeed the right one. The same holds true when evaluating the trustworthiness of institutional repositories. Although researchers caution that the approaches used in a national repository could well transcend boundaries and apply to international pursuits, it does not necessarily lead to only one universal tool for preserving digital material over the long-term (Dobratz and Scholze, 2006). Rather, the major task of any repository should be "evaluating and disseminating examples of good or best practice and by initiating and intensifying regional, national, and international collaboration" (Dobratz and Scholze, 2006, p. 583).
In order for a repository to be deemed trustworthy, it must operate "according to its objectives and specifications (it does exactly what it claims to do)" (Dobratz and Schoger, 2007, p. 212). Further, a repository must contain information that is complete and control for any unplanned changes, whether these changes are accidental technological glitches or deliberate sabotage. It therefore becomes essential that any edits to any part of a record, once it has already been placed in the depository, is meticulously noted.
Dobratz and Schoger (2007) also make mention of groups of users whose particular interests lie in ensuring that the trustworthiness of repositories is maintained. These include users who wish to access the information, data producers and content providers, and funding agencies. In addition, repositories that wish to remain functional, trustworthy, and in business for many years down the road must "fulfill legal requirements…to survive in the market" (p. 212). A trustworthy digital repository puts the author's mind at ease, knowing that their information is secure, and will be preserved with the utmost integrity (Dobratz and Scholze, 2006). As previously mentioned, the RLG/NARA audit checklist and the Nestor certificate may be the most well-known means to prove the validity and trustworthiness of a repository, but they are by no means the only methods in existence.

What an Audit Represents

A question that should weigh heavily on the minds of any institution containing a digital repository is to assess what an audit represents to establishing criteria and trustworthiness, and what decisions must be made in order to either carry along the same work, or guide the repository in a different direction. Further, in order for a repository to be deemed trustworthy, it must meet its objectives, and contain information and material according to its mandate. There is certainly a strong tie between a trustworthy repository and its information technology infrastructure, dependent upon a number of competing factors. These include integrity, authenticity, confidentiality, and availability (Dobratz and Schoger, 2007). Authenticity precludes that the repository meets its objectives, containing information and material that, by its mandate, it is supposed to contain.
As Dobratz and Schoger (2007) explain, "availability is a guarantee of access to the repository…and that the objects within the repository are interpretable" (p. 212). This is essential for ensuring a repository's survival: repeated difficulties encountered with retrieving a specific item within a repository, or continuous maintenance resulting in repository downtime will result in clients choosing to deposit and/or access their material elsewhere. Allowing the owners of the repository to determine who should be granted permission to access the repository's contents instills a higher level of confidence for the depositing author, as he/she is able to upload and tag his/her own publications. Nevertheless, this level of access can be difficult to maintain. (Dobratz and Scholze, 2006).
Hou, Wojcik, and Marciano (2011) provide a voice that many institutions housing digital repositories can relate to: "integrity is an essential component of a trusted digital repository…all of the functional areas will

have an audit trial" (p.182). Thus, establishing an audit for trustworthy repositories represents evidence gathered (usually by means of a checklist) measuring whether or not the repository adhered to pre-determined established evaluation criteria. Further, as digital repositories are primarily web-based programs relying on a server housed in the home institution, these repositories must have "a succession plan or escrow arrangements in place in case the repository ceases to operate." (Ambacher, 2007, p. 6). Ambacher also posits that data loss, whether accidental or intentional, will inevitably occur, a potential weakness that can be exploited. Therefore, maintaining a sustainable repository with a firm foundation, along with establishing a back-up alternate route in the event of a digital disaster, is essential.

While gathering appropriate hardware, establishing a reliable and secure network connection, and ensuring that a digital repository is utilized to its full potential are all essential components of certification; having the appropriate software to run the repository cannot be overlooked. The *Audit Checklist for the Certification of a Trusted Digital Repository*, jointly created in 2005 by the RLG and NARA, comments on the framework used to evaluate such common repository software packages such as DSpace, Eprints, and Greenstone (Kaczmarek et al., 2006). Regardless of the software package that is chosen, it must be applicable and adaptable, in order to "facilitate data transfer…easily…to take advantage of future, unforeseen developments in computer software and technology" (p.2).

The goal of the RLG/NARA Audit Checklist is "to develop criteria to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections" (Kaczmarek et al, 2006, p. 4). Adhering to the three key areas of digital preservation (namely, technology, resources, and management), the Audit Checklist consists of four key sections: organization; repository functions, processes, and procedures; designated community and the use of information; technologies and technical infrastructure (pp. 4-5).

Reasons Why an Audit is Done

If a digital repository is mapped out appropriately, it can have tremendous benefit to both the author depositing research material, and the institution responsible for its upkeep and maintenance. Therefore, an audit need not necessarily be seen as a negative or patronizing activity, but rather as a means of establishing credibility, and educating the repository owner as to any changes that may be required in order to help the repository gain trustworthiness among its users. Of the numerous reasons for why an audit is undertaken, the following are considered to be the core criteria that is often adhered to: an audit should maintain a sustainable, secure repository, with a user-friendly interface; it should establish and maintain a policy that will result in a long-term repository for data producers; it will benefit from a solid management foundation, ensuring that high-quality information is continuously deposited; finally, an audit must identify weaknesses and risks, and establish a process to overcome these challenges (Prieto, 2009).

As the recent copyright issues in Canada indicate, particularly the current Access Copyright befuddlement that exists at some academic institutions, there are a number of legal ramifications that must be taken account when depositing material into any repository. The repository ownership must allow material to be uploaded, stored in an archive, and modified, as required, for posterity (Dobratz and Scholze, 2006, p. 587). Additional challenges faced by these institutions result from the speed in which some repositories have been established. As Downs and Chen (2010) explain, methods for storing and preserving digital content have not yet reached the level of organization used to house non-print material. This can raise doubts about the content of a digital repository, as "trust encompasses not only the integrity of the digital data, but also the authenticity of the links between the data and the data sources and documentation" (Downs and Chen, 2010).

Security of the contents within a repository will always play a prominent role. Repositories should be accessible around-the-clock, and include digital signatures as well as digital object identifiers (DOI) to be able to easily retrieve a requested file. In addition, the establishment of a consistent archiving format will ensure that documents are preserved for many years into the future. In fact, "the minimum availability of a document [should] be no less than five years" (Dobratz and Scholze, 2006, p. 590).

While supporters of the Open Access Movement would declare that the full contents of a repository should be freely and publically available to all (and indeed, this is the case with a number of institutional repositories, including DSpace at the University of Calgary), there are nevertheless a number of interest groups for whom trustworthiness holds particular merit. These include users who wish to access reliable information immediately and well into the future, content providers who rely on the audit of a repository

to support their effort at ensuring high-quality information in a repository is maintained (i.e. a warranty for data producers), and corporations that determine whether or not a repository will receive adequate funding and for how long. Finally, as previously mentioned, entering the digital repository environment is indeed a competitive venture, and all repositories are therefore required to "fulfill legal requirements" (Dobratz and Schoger, 2007, p.212) in order to survive.

One methodology posited by Kaczmarek and colleagues (2006) is the creation of a matrix to function as a tool which will aid in the decision-making process of certifying a repository as a trustworthy source of information. Kaczmarek et al (2006) explain that settling on which software package best suits a particular repository will lead to a rubric "to determine how critical each particular point of functionality is and if that point is absolutely required" (p.2). Such steps were taken by the Exploring Collaborations to Harness Objects in a Digital Environment for Preservation (EXCHO DEPository) project, a joint effort between the National Digital Information Infrastructure and Preservation Program (NDIIPP) at the Library of Congress, and the University of Illinois at Urbana-Champaign.

While the above examples of digital repositories comment on the importance of establishing policies that are firmly adhered to in order to establish trustworthiness and acuity, repositories must also be established in such a way that they can be easily customized if necessary. Such is the case with DCAPE, the Distributed Custodial Archival Preservation Environments project, originating out of the University of North Carolina Chapel Hill (Hou, Wojcik, and Marciano, 2011). Adhering to the three key preservation policies, namely "management of archival storage, validation, and trustworthiness" (p. 181), DCAPE supports one of the fundamental reasons why an audit of a repository is undertaken. Ensuring that high quality material is continuously deposited is certainly one way of ensuring a repository's livelihood, however without a user-friendly interface, authors and researcher's alike may become frustrated and choose to deposit their publications elsewhere, which, in turn, reflects negatively on the purpose of sustaining the repository for generations to come.

Existing Audit Methodologies and Tools

DINI, the Deutsche Initiative für Netzwekinformation, is aimed at supporting the Open Access movement in Germany. The aim of this guideline is to enhance the cooperative partnership between German educational institutions with a goal to "provide a tool for repository operators that could be used to raise the visibility, recognition, and importance of the digital repository within the university." (Dobratz and Scholze, 2006, p. 584).

As exemplified in many repositories, DINI criteria are based on two categories, the first of which explains the minimum requirements that must be captured in order for the repository to be deemed credible. These requirements include visibility and server policy, support for authors, legal issues, authenticity and integrity, indexing, impact and access to statistics, as well as long-term availability (Dobratz and Scholze, 2006, p. 585). Nevertheless, despite these rather strict requirements, Dobratz and Scholze comment on the challenges involved in deeming a repository to be both trustworthy and credible, hence the need for an audit. These include the establishment of a server policy, creating a visible service for authors, and implementing persistent identifiers (p. 586).

In addition to the aforementioned repository requirements, DINI also supports the need for creating open access to archived materials, and posits that a policy needs to be established to allow for each repository to be registered and recognized by large-scale collectives, namely the Directory of Open Access Repositories, OpenDOAR. (Dobratz and Scholze, 2006). As DINI proclaims, creating an open access policy showcases "a clear commitment to support the 'green way' to open access" (p. 587).

Originally created with cultural heritage organizations in mind, the Nestor Catalogue of Criteria for Trusted Digital Repositories serves as a guide for planning and maintaining digital repositories well into the future (Dobratz and Schoger, 2007). The criteria raised by Nestor include the following key concepts which can be applied to virtual any repository framework: compliance with terminology created by the Open Archival Information System (OAIS), abstraction, adequate documentation, transparency (essential to gain trust), adequacy, and measurability. As Dobratz and Schoger (2007) proclaim, these criteria will function as "indicators showing the degree of trustworthiness" (p. 214). The organizational structure for Nestor is divided into three top-level categories, each with a number of subdivisions. These are depicted as follows: organizational framework (defined goals, adequate usage, legal and contractual rules, organizational form, quality management), object management (integrity, authenticity, strategic plan for technical preservation,

acceptance from producers adhering to established criteria, archival storage, usage, data management system), and infrastructure and security (adequate IT infrastructure, protecting the repository and the objects contained within it) (pp. 215-216). [See Appendix 1].

(DRAMBORA): A Methodology and Tool for Auditing a Trustworthy Digital Repository

DRAMBORA description: tool and methodology

Launched in 2008, as the result of a joint effort between the Digital Curation Centre and Digital Preservation Europe, the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) functions as a toolkit "to make the self-auditing process easier and more efficient for repository managers" (Donnelly et al., 2009). Although digital repositories had already been in place for some time prior to the establishment of DRAMBORA, there was no standard guideline for determining the key components required for successfully implementing, initiating, and sustaining a digital archive. This issue led to the Centre for Research Libraries (CRL), widely credited as the developers of DRAMBORA to produce a list of 10 core requirements that all digital repository owners must be made aware of and should follow to preserve their archival storehouses for generations to come (see Appendix 3). As can be seen from this list, technological infrastructure plays only one part in ensuring that the data within a repository is adequately stored and maintained over time. Creating a manageable process and action plan, along with accounting for any legal ramifications that may manifest themselves along the way, is equally important.

While DRAMBORA is a relatively recent phenomenon, it nevertheless underwent a series of pilot tests in the two years prior to its official unveiling. More than merely serving as another toolkit, it is primarily responsible for presenting "a methodology for self-assessment, encouraging organizations to establish a comprehensive self-awareness of their objectives, activities, and assets before identifying, assessing, and managing the risks implicit within their organizations" (Donnelly et al., 2009). Undoubtedly, attempting to maintain any form of electronic storage method implies a certain amount of risk, perhaps even more so than a traditional print collection. DRAMBORA has attempted to ease this risk process by positing a series of stages: authors are required to develop an organizational profile, describe and document a mandate, and list objectives/goals, activities, and assets (Donnelly et al., 2009). These stages are, however, only meant to serve as guidelines; the DRAMBORA team cautions that the entire purpose of this audit method is to serve as a living document, with revisions being made along the way as the need arises.

How to Use DRAMBORA

There are presently 18 institutions that rely on the DRAMBORA toolkit to conduct self-assessment audits of their digital repositories. A number of these organizations also hold strong ties to the grey literature community. Before describing how to use this methodology it is necessary not only to discuss the purpose of DRAMBORA (see Appendix 2), but also its three primary applications: as a web-based tool, DRAMBORA can assess the effectiveness of a repository infrastructure, and offer suggestions for its improvement; it acts as a preparatory resource for external auditors who may wish to serve as aggregators of the DRAMBORA movement; finally, it anticipates any potential weaknesses or challenges, and subsequently adjusts its plans to overcome these boundaries (Donnelly et al., 2009).

Existing in both an online and offline format, DRAMBORA is a user-friendly program that guides the user through four phases. First, the user is encouraged to register for a personal account, as well as provide details regarding the repository at his/her institution. This allows DRAMBORA to present a customized self-assessment profile for each user. Further, additional staff members from the institution in question will be identified, where they will be able to contribute to the self-assessment process. The subsequent phase of using DRAMBORA refers to the actual self-assessment audit. The goal of this stage is to ensure that the repository undergoing an audit establishes clear objectives with documented sources. The organization's mandate and/or vision/mission statement, along with any potential legal or technical issues should be listed here.

Following the actual self-assessment, any potential risks must be identified and assessed. For evaluative purposes, all of the risks identified are categorized and assessed according to their potential impact, accounting for the frequency or probability that any potential negative effects and subsequent risks could appear. (Donnelly et al., 2009). Finally, once all risks have been assessed and identified, a plan should be established to develop counter-measures, anticipated outcomes, and a timeline to reassessment the

repository, to ensure that any issues have been resolved. The careful mapping of a repository using the DRAMBORA tool may already give an overview of what is finished and what is not, which documents, procedures, tools, and measurements are missing and where most critical risks reside.

DRAMBORA: Advantages and Disadvantages

Undoubtedly, as a web-based self-audit tool, DRAMBORA far surpasses a number of competitors in this field, and thus it would not do the tool justice without mentioning some of its key benefits. First and foremost, the online version of this program allows the user to view the internal activity of a repository, identifying any potential problems, and rectifying them as quickly as possible. In addition, the user is able to interact with the content on the screen, navigating to sections of interest without having to flip through an entire text. Second, the methodology and tools are well implemented, clearly identifying the organizational role and structure of each institution taking part in the audit. Third, the descriptions used and examples presented are pertinent, intuitive, and applicable to the task at hand. This includes a clearly-defined mission statement, complete with key aims and objectives. Finally, the scale of evaluation compared to the risks is adequately assessed: "an internal understanding of the successes and shortcomings of the organization [enable it]…to effectively allocate or redirect resources to meet the most pressing issues of concern" (Donnelly et al., 2009). "For inspiration and possible direct help, the DRAMBORA tool contains a number of links to supporting documents and a range of practical examples of completed entries, whether in the preparation phase or the audit phase. In the area of risk identification, predefined risks can be directly used and modified or unique risks may be formulated." (Karlach, 2010, p. 127)

Despite the obvious benefits of DRAMBORA, there are nonetheless a few disadvantages that must also be considered. Namely, at present, the implementation and methodology is only available in English. Although English is seen as the universal language of communication, it is important to note that the majority of DRAMBORA users hail from European countries. Thus, offering the DRAMBORA interface is a variety of languages is a task that the developers of this toolkit would be wise to consider. An additional disadvantage relates to the technical, albeit programming aspect of this product. DRAMBORA functions perfectly fine on a standard Windows or Mac interface, but is not compatible with the Czech Windows operating system (i.e. it does not support the Czech character set – iso-8859-2/windows-1250). While it is understandable that this program cannot comply with every possible computer operating system, perhaps a text-based (DOS) version, in addition to the current HTML version, should be brought forward to the development team. The final two arguments surrounding the negative aspects of DRAMBORA center on access issues. Presently, read-only access is not permitted (a user must be fully registered and log in to a created account, in order to make use of al of the program's features). In addition, exporting records (either via e-mail or to a bibliographic management program) is currently not possible.

Audit of the National Repository of Grey Literature (NRGL)

Introduction to NRGL

The NRGL project, *The Digital Library for Grey Literature – Functional Model and Pilot Implementation*, started thanks to the support from the *Ministry of Culture of the Czech Republic* as part of both research and development programs. The project is divided into three phases, lasting from 2008 to 2011. Its main goals are the systematic collection, long-term archiving and provision of access to specialized grey literature, pertaining specifically to research and development, civil service and education, as well as with the business sphere and "open access" at the national level. To support this goal, the NTK created a functional network of partner organizations, a working model, and a pilot application. In addition, on the basis of verified technology and methods defined under the project, recommendations and standards are created for other institutions electing to build their own digital grey literature repositories. Recommendations and standards consist mainly of a preferred metadata format, exchangeable formats and templates, examples of licensing models and of legal issues resolved, preservation methodology, archives, and the provision of access to digital data.

NRGL Audit

The first audit of the NRGL as a trustworthy digital repository using the toolkit and methodology of DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) was performed at the end of 2009 as a part of creating a digital repository of grey literature in the National Technical Library (NTK). The audit results and experience from its course were summarized in a final report, and published in the book *Grey Literature Repositories* in 2010. The most important outcome of the audit was presented by identifying risks connected to the NRGL and potentially endangering its operation, quality, images, and other features; these risks were eliminated or moderated by the NRGL team during 2010. The main principle of the DRAMBORA audit and, at the same time, its main contribution and iteration, resulted in its repetition after a certain time period in new conditions when original risks were reassessed, the measurements adopted for solution assessed, and new risks identified.

The second audit of the NRGL digital repository was performed after one year. In this audit, the actual state of the repository was assessed, with progress achieved during 2010. New potential risks were identified, as well as possible ways to eliminate them or to reduce their impact. The NRGL documentation, the description of the whole project, its processes, procedures, and related documents developed significantly during 2010, which, in turn, made a solid basis for the audit.

Work with the DRAMBORA Tool went on without any of the issues and problems experienced with the previous audit. On the basis of lessons learned from 2009, the NTK communicated with the authors of the tool and methodology, and proposed some improvements and modifications. As a result, the web tool DRAMBORA appeared more stable after a one-year pause; however, the most unexpected modifications have not been introduced yet, especially regarding the elimination of the rather unpleasant fact that the online version of DRAMBORA does not support languages other than English at this time. Nevertheless, as the results of the audit are intended to be presented in the international field, namely in the area of grey literature projects, we will continue to use English.

NRGL Audit: Preparation and Definition

DRAMBORA Interactive was used during the preparation phase in addition to the audit phase (Karlach, 2010, p. 126-127). The preparation phase consisted of acquiring all relevant information and documents on the status of the repository, its description, standards, procedures, staff, material, budget, etc. (Donnelly et al., 2009). This information served as input data for the preparation phase of the audit and was entered into the DRAMBORA Interactive in the section *Before the Assessment*. Here, the repository was described, the scope of relevant areas (Functional Classes) of the audit were defined, and the repository staff, including a detailed description of individual team members and their roles, was listed. The definitions of staff roles were especially important since, at the subsequent risk identification stage, it was necessary to relate risks to respective roles. Even during the preparation phase, a substantial contribution might be made to an audited repository. This helps the staff see the repository from a global vantage point, to map and accumulate the most important descriptive data about the repository, and to point to possible deficiencies and defects, offering the opportunity for problems to be remedied and missing materials to be completed. The audit was run using the portion of the DRAMBORA tool called the Assessment Centre (Donnelly et al., 2009). Here, the repository mandate, including its mission, purpose, founders, etc., was defined. Other repositories were also identified that influenced its activity, both external (e.g., legislative) and internal (e.g., organization, content type restrictions, etc.). The audit continued by defining repository goals, activities, and the means used to achieve these goals.

NRGL Audit: Identified Risks

In addition to the mapped repository and the relevant environment, the producers of the methodology and the tool consider the most important output to be the analysis of identifiable risks endangering the repository, its quality, readiness, reputation and position in the eyes of both specialists and ordinary users. In the 2010 audit, 16 risks from the previous audit were assessed, primarily regarding the progress in their elimination, and an additional 8 new risks were identified. The NRGL repository is still in the pilot project stage; however, it is run on final software versions and real data are being stored. Identified and reassessed risks mainly refer to the description of activities and procedures of the repository, the state and development of the staff, project funding, hardware and software sources, including their backup and relationship to the NRGL environment (see Appendix 5).

After all necessary information is entered, the Reporting Centre function helps to create output reports on the identified risks for the repository, with respect to their relationship and plausible solutions. Two types of output report formats are available, either PDF or HTML. Other saved descriptive information cannot be exported easily, however, it is possible to copy saved snapshots of the audit page. Besides the mapped repository and its relevant environment, the producers of the methodology and tool consider the most important output to be the analysis of identifiable risks endangering the repository, its quality, readiness, reputation, and position in the eyes of both specialists and ordinary users. Since the DRAMBORA tool does not provide read-only access, it is regrettably not feasible to allow free access to the audit at this time.

Generally, risk elimination is much easier in a case where the respective area is fully under control and in charge of the NRGL management and team. If the risk relates to the cooperation within or even outside the NTK, the situation is considerably more complicated. The creation of a knowledge database NRGL Wiki indicates great progress; this database should be further developed and strictly adhered to, as the continuous documentation of procedures, activities and results of the NRGL team is of crucial importance for the elimination or minimization of the impact of most risks. Such progress may be seen in the development of the NRGL repository since the last audit along with the new activities and goals that have been added. Therefore, the documentation of activities and analysis of risks are most important. A large portion of risks reflect the topic of building the NRGL partner network, i.e. the partner network of providers of the repository content. Consequently, this area should be of priority especially for promotion and education. In relation to NRGL partners, sufficient attention should be paid to legal issues connected to the Author Act.

Conclusion

A yearly repetition of the audit under new conditions, identification of new or modified risks, and creation of another action plan make the audit an iterative process that contributes to the trustworthiness of the NRGL. Despite the valiant efforts of libraries, information technology specialists, and researchers, who devote considerable amounts of time and effort to maintain credible digital repositories, it can seem like a tall barrier to overcome. While Downs and Chen (2010) caution that “no organization can absolutely guarantee long-term preservation and access”, efforts to establish methods of audit and recognize trustworthy digital repositories must continue. As DRAMBORA and the subsequent audit of the National Repository of Grey Literature have shown, the task at hand may not yet been complete, but it is certainly moving in the right direction. It is thus perhaps fitting to conclude with the mission statement of Columbia University, which reflects not only on the goals of this particular institution, but which speaks to the efforts of raising awareness of grey literature in all topic fields and venues. Namely, one must “advance knowledge and learning at the highest level and…convey products of its efforts to the world” (Columbia Mission Statement, 2011). We therefore recommend that an audit be undertaken on an annual basis, identifying any associated risks, and creating an action plan to make the audit an iterative process that contributes to the trustworthiness of the digital repository.

References

Ambacher, B. (2007). Government archives and the digital repository checklist. *Journal of Digital Information, 8*(2), 1-10.

Columbia University. (2011). *Mission Statement*. Retrieved November 20, 2011 from http://www.columbia.edu/content/mission-statement.html

Dobratz, S., & Schoger, A. (2007), Trustworthy digital long-term repositories: The Nestor approach in the context of international developments. *Research and Advanced Technology for Digital Libraries, Proceedings, 4675,* 210-222.

Dobratz, S., & Scholze, F. (2006). DINI institutional repository certification and beyond. *Library Hi Tech, 24*(4), 583-594.

Donnelly, M., Innocenti, P., McHugh, A., & Ruusalepp, R. (2009). *DRAMBORA Interactive User Guide*. Glasgow. Retrieved November 22, 2011 from http://www.repositoryaudit.eu/help/

Downs, R. R., & Chen, R.S. (2010). *Self-assessment of a long-term archive for interdisciplinary scientific data as a trustworthy digital repository.* Retrieved October 22, 2011 from journals.tdl.org/jodi/article/download/753/642

Hou, C.Y., Wojcik, C., and Marciano, R. (2011). Trusted digital repository design: A policy-driven approach. *Archiving, 7*, 181-186.

Kaczmarek, J., Hswe, P., Eke, J., & Habing, T.G. (2006). Using the Audit Checklist for the Certification of a Trusted Digital Repository as a framework for evaluating repository software applications. *D-Lib Magazine, 12*(2), 1-10. Retrieved August 3, 2011 from http://www.dlib.org/dlib/december06/kaczmarek/12kaczmarek.html.

Karlach, P. (2010). An audit of the National Repository of Grey Literature using the DRAMBORA tool. In Pejšová, P [ed.]. *Grey Literature Repositories*. Zlin: VeRBuM, p. 126-127. Available as an E-book at: http://nrgl.techlib.cz/images/Book.pdf.

National Technical Library. (2008). *Audit of the National Repository of Grey Literature (NRGL) in the NTK using the DRAMBORA tool: Second audit, 2010.* Retrieved August 25, 2011 from http://nrgl.techlib.cz/images/DRAMBORA_2010_EN.pdf

National Technical Library (2008). *National Repository of Grey Literature: An audit of the NRGL as a trustworthy digital repository.* Retrieved August 25, 2011 from http://nrgl.techlib.cz/index.php/Audit

NUŠL. (2011). *National Repository of Grey Literature [NRGL].* Retrieved November 20, 2011 from http://nrgl.techlib.cz/index.php/Main_Page

Pejšová, P. (2010). The development of grey literature in the Czech Republic. In Pejšová, P [ed.]. *Grey Literature Repositories*. Zlin: VeRBuM, p. 34. Available as an E-book at: http://nrgl.techlib.cz/images/Book.pdf.

Prieto, A.G. (2009). From conceptual to perceptual reality: Trust in digital repositories. *Library Review, 58*(8), 593-606.

Ross, S. (2006). The role of evidence in establishing trust in repositories. *D-Lib Magazine, 12*(7-8). Retrieved November 18, 2011 from http://www.dlib.org/dlib/july06/ross/07ross.html

Appendix 1: TRAC-Checklist

(adapted from: Dobratz, S., & Schoger, A. [2007]. Trustworthy digital long-term repositories: The Nestor approach in the context of international developments. *Research and Advanced Technology for Digital Libraries, Proceedings, 4675,* 210-222.)

Common principles:

1. Continued maintenance of digital objects
2. Organizational fitness
3. Acquires and maintains contractual legal rights; fulfills responsibilities
4. Effective and efficient policy framework
5. Acquires digital objects based on criteria, corresponding to commitments and capabilities
6. Maintains and ensures integrity, authenticity, and usability of digital objects over time
7. Creates and maintains metadata about actions to take on digital objects during preservation, as well as relevant production, access support, and usage process context before preservation.
8. Dissemination requirements
9. Strategic program (preservation planning and action)
10. Technical infrastructure adequate (maintenance and security)

Appendix 2: Purpose of DRAMBORA Toolkit

(from Donnelly et al., 2009)

1. Defining the mandate and scope of functions of the repository
2. Identifying the activities and assets of the repository
3. Identifying the risks and vulnerabilities associated with the mandate, activities, and assets
4. Assessing and calculating the risks
5. Defining risk management measures
6. Reporting on the self-audit

Appendix 3: The Ten Core Requirements for Digital Archives

(from Donnelly et al., 2009)

1. Mandate & Commitment to Digital Object Maintenance
2. Organizational Fitness
3. Legal & Regulatory Legitimacy
4. Efficient & Effective Policies
5. Adequate Technical Infrastructure
6. Acquisition & Ingest
7. Preservation of Digital Object Integrity, Authenticity & Usability
8. Metadata Management & Audit Trails
9. Dissemination
10. Preservation Planning & Action

Appendix 4: NRGL Audit - Examples

DRAMBORA Interactive creates connections among individual parts of audit. See connections marked bold.

1. Mandate and equipment

NTK status: To build national repository of grey literature and to make the information and findings contained in the repository accessible for NTK users using modern information technology.
http://www.techlib.cz/default/files/download/id/1747/dodatek-c-1-k-zl-ntkpdf.pdf
Repository Hardware
Hardware used to run the repository software and database - SUN SUNXFIRE 4500 server, OS SOLARIS 10

2. Functional classes

Supporting Functions:
Legal & Regulatory Legitimacy
Functions and characteristics corresponding to legislative, regulatory or common law rights and responsibilities of the repository.
Operational Functions:
Acquisition & Ingest
Functions and characteristics corresponding to the repository's negotiation, submission, receipt and ingestion of data from creators and suppliers.

3. Staff

Position: Manager
Unique Staff ID: 1
Telephone: +420232002485
Staff Email: petra.pejsova@techlib.cz
Address: NTK, Technicka 6/2710, 160 80 Praha 6
Status: Coordinator
Username: petrapej
Name: Miss Petra Pejsova
Alt. Email: petra.techlib@gmail.com
Roles: Management

4. Roles

Role Name: Management
Description: Establishes strategy and objectives of the repository, Establishes strategy of the repository content provider network...
Corresponding Staff Members: Manager
Activity Responsibilities: Budget Management, Cooperation Network, Team Management, NUSL Publicity
Risk Responsibilities: Loss of Staff Members, Pilot Project End, Disaster Recovery, Partner Network Voluntary, Backup Tapes...

5. Constriction

Name: Documents Publication Status
Description: The Repository is devoted to grey literature, so it accepts only unpublished or semi-published documents
Type: Policy
Functional Class(es): Supporting Functional Classes - Acquisition & Ingest
Web Links:
http://nusl.techlib.cz/index.php/Typologie_dokument%C5%AF_NU%C5%A0L

6. Goals
Name: Best Practices
Description: Best practices for building similar cooperating institutional repositories are one of the planned outputs of the project
2010: Best practices for partners created in the technical and methodical areas, see section **Constraints - Methodology of the cooperation with NUSL etc.**
Functional Class(es)*: Supporting Functional Classes - Efficient & Effective Policies

7. Activities
Activity Name: Repository Backup
Activity Desc: To create a backup copy of the system to preserve the current setup and of the repository database to preserve it's content
Activity Role(s): Administrator
Related Assets: Repository Hardware, Repository Software
Related Objective(s): Main Function
Functional Class(es)*:Preservation of Digital Object Integrity, Authenticity & Usability
Related Risks: Backup Tapes Storage

Appendix 5: NRGL Audit – Identified Risks

NRGL was analyzed in all functional classes, however, only the most severe and most obvious risks were recorded according to Pareto's rule of 80/20 – 20% of risks are responsible for 80% of the danger.

16 originally identified risks in 2009:

Risk Number 1: Loss of Staff Members
Risk Number 2: Pilot Project End
Risk Number 3: Disaster Recovery
Risk Number 4: Partner Network of Volunteers
Risk Number 5: Backup Tapes Storage
Risk Number 6: Financial Shortfall
Risk Number 7: Budget for Services
Risk Number 8: FAST Trial Version
Risk Number 9: Weak Mandate
Risk Number 10: No Ingest Policy
Risk Number 11: Document Formats
Risk Number 12: Software Administration
Risk Number 13: Undocumented Policies
Risk Number 14: Long Term Preservation Strategy not described
Risk Number 15: Staff Skills insufficient
Risk Number 16: Deliberate System Sabotage

Risks newly identified in 2010:

Risk Number 17: Duplicate project
Risk Number 18: Partners do not supply full text
Risk Number 19: Slow growth of partner network
Risk Number 20: Sample partner contract has limited usability (applicability)
Risk Number 21: Migration to new HW platform
Risk Number 22: New CDS Invenio version
Risk Number 23: Legal Risk - Authors Act
Risk Number 24: Termination of legal support

Federal Information System on Grey Literature in Russia: a New Stage of Development in Digital and Network Environment

Aleksandr V. Starovoitov, Aleksandr M. Bastrykin,
Anton I. Borzykh and Leonid P. Pavlov
Centre of Information Technologies and Systems of Executive State Authorities, Russia

Introduction

Since the late nineties when the Russian grey literature (GL) system in the sphere of scientific and technical information was first presented to the international GL community in Luxembourg we have had several opportunities to describe one or another aspect of the Federal Information System on GL in Russia [1,2,3,4,5]. This time we would like to dwell upon the system as a whole following its development from the past through present times to the prospective view all the more as this year a new ambitious project is started with the aim of renovating the system in accordance with the up-to-date requirements. The project has received a sufficient government funding for the coming three years.

The Russian Federation has inherited the federal-level information system on grey literature from the Soviet Union. The system covers the most informative kinds of grey literature - scientific research and development reports and post-graduate theses as the sources of scientific and technical information being centrally collected at the Centre of Information Technologies and Systems of Executive State Authorities (abbreviated in Russian as CITIS) in accordance with the Federal Law "On the obligatory copy of documents". The law obliges all the organizations – the collective authors of reports and persons – the individual authors of dissertations to give a free full-text copy of the documents to CITIS. In turn, the Centre is obliged not only to complete and permanently store the collection but also to disseminate the information on its content.

In the course of the past decades the system experienced several modifications in order to get adapted to the changing organizational and technological reality. In its present state the federal system combines the following three functionally separate systems run by CITIS: the traditional system for collecting, processing, storing and providing access to R&D reports and theses called the computerized information system on science and technology (abbreviated in Russian as ASINIT) that has recently been improved to store the full-text reports and dissertations in a digital form and provide full-text search and retrieval; the system for self-funded research projects registration and monitoring that was put into operation in mid-2000 to reflect a growing trend in funding R&D projects from research organizations' own financial resources; the federal register for the results of scientific and technical activities also created in mid-2000 with the idea of monitoring the life-cycle of patentogenic findings documented in scientific reports.

All the three systems are operative under the name "United Federal Database on Research and Development" (UFD R&D) and fulfil their functions however rapidly changing digital and network technologies create new environment to increase the systems' efficiency and improve its services. A new project in the process of development at CITIS is under the auspices of the newly-started State Programme of the Russian Federation "Information society (2011 – 2020)". The project is aimed at the creation of the Integral state information system on scientific research and development that is supposed to unite the three systems using unified forms of input documents so that users were to fill in the similar information only once and in interactive network conditions. The integral system will use the instruments of full-text digital documents analysis and web-technologies so that to improve data-mining and to avoid plagiarism.

The past

The computerized information system on science and technology (ASINIT) has been operating since 1975. It was then created as the grey literature part of the national library-information fund of the USSR and the part of the State System for Scientific and Technical Information (abbreviated in Russian as GSNTI). ASINIT consisted of two divisions: the full-text R&D reports and dissertations (the so-called primary documents)

stored on microfiches and the bibliographic cards with abstracts (the so-called secondary documents) stored in the mainframe computer in a database format. There are two kinds of the secondary documents: the registration cards that are filled in when a new R&D project is started and the information cards that accompany full-text reports and dissertations.

Later on, in the early eighties ASINIT became the host core of the computer network called AIST in Russian for "computerized teleprocessing information network". AIST connected distant smart terminals, a prototype of personal computers situated all over the country, to the ASINIT host-computer with the grey literature databases situated in Moscow. The network operated in a dial-up mode through the public telephone lines. The distant users could conduct online searches in the centralized databases on reports and theses and order copies of documents from the System GL collection. The network throughput provided for more than 500 search, retrieval and copy-ordering transactions per day. That was the first information computer network of the pre-Internet epoch commercially working in the country.

No matter how obsolete the soft- and hardware of the System may seem now from the very beginning ASINIT met the main complex of requirements for completing the obligatory copy grey literature collection (R&D reports, candidate and doctoral dissertations – theses, descriptions of algorithms and computer programs), federal registration of the documents, the database support, online search and retrieval, abstract journals publishing, permanent storing and archiving the documents [1].

This system's configuration existed for several decades with the technological changes from mainframe computers to PCs, database and network servers and the information migrating from magnetic tapes through diskettes and CDs to the modern electronic data stores.

The present

At present ASINIT is still the heart of the United Federal Database on Research and Development (UFD R&D) along with other two systems appeared several years ago. All the systems are functioning on the technological platform of the Centre of Information Technologies and Systems of Executive State Authorities (CITIS). In 2004 by the Decree of President of the Russian Federation ASINIT was included in the list of strategically important systems. Since 2010 the System has been listed in the Federal Register of the State Information Systems.

By the end of 2010 the system supported the following information resources:

- the retrospective bibliographic database with abstracts for R&D projects registration cards and R&D reports information cards containing nearly 2,5 million documents with the depth of retrospective 30 years (each card consists of more than 30 information fields) including
  - registration cards – nearly 1,2 million;
  - information cards – nearly 1,3 million;
- the retrospective bibliographic database with abstracts for dissertations containing more than 640 000 documents with the depth of retrospective 30 years (each card consists of 35 information fields) including
  - candidate dissertations information cards – nearly 560 000;
  - doctoral dissertations information cards - more than 80 000;
- the abstract journals database – nearly 3,0 million documents;
- the database for information cards translated into English – more than 80 000 documents;
- the algorithms and computer programs database – more than 15 000 documents;
- full-text R&D reports (since 1984) – nearly 800 000;
- full-text dissertations (since 1984) – nearly 600 000 including
  - doctoral dissertations - nearly 80 000,
  - candidate dissertations – more than 500 000;
- the database for scientific organizations submitting R&D reports – more than 6 000 organizations.

The report and dissertation information cards databases are placed on a CITIS server with online network availability for the users. The databases serve as an electronic catalogue for the full-text collection providing a fast means of registration and search. The arriving full-text paper reports and dissertations are being scanned and PDF stored. At the same time the earlier documents are retroconverted (now backwards

to the year of 2000) from the microfiches to PDF format. About 11 000 full-text R&D reports are entered into a full-text database. For the beginning of 2011 the total size of the electronic document store is 5 TByte. The total size of the information fund – more than 7 million documents.

The desk-top publishing system allows for issuing both electronic and paper abstract journals but now only the electronic versions are commercially disseminated by subscription. 51 titles of the journals by 25 subject series are published, totally 236 issues per year.

There are two government level documents which form a legal ground for the operation of the system: the Federal Law "On the obligatory copy of documents" of December 29, 1994 № 77-FZ (in the wording of March 26, 2008 № 28-FZ) and the Government Decision of March 31, 2009 № 279 that delegated all the functions of running ASINIT to CITIS.

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 and supports monitoring and controlling the situation (both in financial and subject respect) in *the state funded* scientific research and development activities covering extensively all the territory of the Russian Federation [2,3]. The system's collection is an indispensable source for government agencies with an interest in the latest Russian contributions to science and technology.

At the same time it is evident that 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 and the diversification of funding is inevitable. So, there is a growing trend in scientific research that more and more R&D projects are being funded from *research organizations' own financial resources.* Those organizations are commercial 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 were out of centralized monitoring and hence were not taken into account when updating the lists of priority development directions in science and critical technologies of the Russian Federation.

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 [4]. The system was designed in the years of 2007 – 2008 and now is in operation as the second part of the United Federal Database on Research and Development (UFD R&D). Based on the output information from the system the Annual Summary Report for the Government is prepared. 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.

Now there is a four-year retrospective database (with the report documents of 2007- 2010 years – totally about 1,5 thousand documents), next year (2012) the documents of 2011 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. The self-funded research monitoring system is evidently grey because its input form and output Annual Summary Report are typically grey documents. Since 2010 the System has been listed in the Federal Register of the State Information Systems.

The grey literature sources contain a bulk of findings to be commercialized and/or claimed as intellectual property objects. The registration of reports and dissertations that is carried out in ASINIT now is rather document- than result-oriented. It would be useful to follow all the lifecycle of a scientific result beginning with the idea and basic research outcome through feasibility study findings to industrial implementation of the result in the form of innovative products and services [5].

In 2005 a Government Decision was issued (No. 284, now it functions in the wording of August 18, 2008 No. 622) on the development of the United Register for the Results of Scientific and Technical Activities (UR RSTA). In 2006 the Register was put into operation with its separate input forms designed to register the objects of intellectual property (patents, databases, computer programs, etc.) obtained in the course of state-funded research. Now the Register database contains the information on 50 ministries – the state R&D projects customers, 15 000 state contracts concluded by them to carry out the projects and 6 000

intellectual property objects. The Register is the third component of the United Federal Database on Research and Development (UFD R&D) operating in CITIS.

Though functionally satisfying the main requirements of the Law and scientific community the existing UFD R&D suffers from several shortcomings that are supposed to be eliminated in the course of the further system's development:

a) all the databases (DB) on R&D — state contracts DB, reports and dissertations DB, the Register DB – are formed independently one from another, so the user has to fill in several similar forms wherein the information is redundant and duplicated;
b) the lack of effective customers and executors control mechanisms, so to say, a feedback from the System to the customers in order to provide the completeness of R&D reports registration and presence in the System;
c) the total computer power of the System is insufficient to implement the modern web-technologies of forming the information resource and providing a comfortable access to it;
d) the limited analytical means of textual information processing;
e) there is no online interaction with other state information systems such like the Computerized information system for scientific research of the Russian Academy of Sciences.

The future

The newly-started State Programme of the Russian Federation "Information society (2011 – 2020)" has opened a real chance for the state financial support of the System's development in the context of rapidly changing digital and network technologies. Under the auspices of the Programme a competition was announced for a state contract to realize the System's development project. CITIS won the competition and the contract was concluded for three years to fund the project named "The development of the United R&D projects registration system (UPRS R&D) for the projects carried out in the civil sphere with the state budget funding".

In general, the project is aimed at the creation of the integral state information system on scientific research and development that is supposed to unite the three systems functioning on the platform of CITIS.

There are some main problems to be solved within the project:

- The development and implementation of effective mechanisms for securing the information completeness in the System that is all the R&D reports must be registered and present in the System. This is very important because the experience of the latest decades exposed a low executive discipline of scientific and scholar institutions that perform R&D projects.
- The elimination of redundancy and duplication in the inputted and stored information due to the existing database conducting independence. Different databases have different forms of records with many coinciding information fields.
- The registered data must include not only the subject of research information but also the data on the state contracts, size and structure of the state funding, patentogenic results of the research project.
- The development of analytical instruments to expose the innovating projects and estimate the results of conducted research.
- The development of the legal basis for the UPRS R&D. A new Government Decision regulating the procedures of the System's operation must be prepared and approved.

From a technological point of view the system's modernization must develop in the direction of network computing and digital documents processing. The essential points of novel approaches are the following.

1. The unified forms of the secondary documents – information cards – are developed so that on the one hand to eliminate the duplication of the same fields in different cards and on the other hand to include more detailed financial data on the size and sources of funding and data on the life cycle of the intellectual property objects (patents, computer programs, databases, etc.).
2. The online mode of filling the new forms of information cards is provided for the authors of R&D reports and dissertations who are able to address the CITIS site on the Internet (www.rntd.citis.ru), click "the online form 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.

3. The formation of digital full-text databases for all the arriving documents (reports and dissertations) with the effective means of full-text search and analysis. The digital documents are entered into the single electronic repository that allows four modes of documents entering: scanning and recognizing the paper documents; inputting the documents arriving on CDs; online arrivals entering; retroconversion of the documents stored on microfiches. Now, in accordance with the existing legal acts, the full-text documents arrive on paper and must be scanned and digitized before being PDF stored. The evident tendency is to pass on to digital input documents.
4. In order to introduce exclusively digital input forms of both the full-text and metadata documents it is necessary to implement an electronic signature technology and to make alterations in the legal acts (laws etc.) currently in force.

There are two kinds of subsystems envisaged in the technical assignment for the new system: those existing and being modernized and those newly designed and implemented.

The modernized ones are:

- the subsystem for reports and dissertations collecting, processing and registration;
- the digital documents repository and archiving subsystem;
- the search and retrieval subsystem;
- the abstract journals publishing subsystem.

Among the newly designed ones are:

- the system's common Internet portal subsystem;
- the R&D projects in progress monitoring and content analysis subsystem;
- the subsystem for interaction with international scientific and technical information systems;
- the subsystem for integration with other Russian state information systems on science and technology.

In the framework of the integral system a new complex of analytical and search instruments is to be designed using artificial intelligence technologies for linguistic text processing and semantic analysis, context and fuzzy search algorithms, subject area structuring, new knowledge and data-mining, antiplagiarism and experts activity support. This will allow to create analytical information not only about a separate scientific work but also about scientific trends, scientific groups and schools, the information for updating and systematizing scientific classification schemes. A linguistic support of these possibilities suggests that computer glossaries and dictionaries, thesauri, ontologies and classifiers should be developed and maintained.

Concluding remarks

During the next three-year stage of development it is supposed to implement the advantages of digital and network technologies and significantly improve the system's characteristics. The system is designed to provide a complete R&D documents collection, a fast access to full-text documents and relevant information. It will allow to monitor the situation in the sphere of R&D works and projects all over Russia, to support the federal level administrative decisions in the sphere of science and technology, to prognosticate its development, to improve the distribution of financial means for scientific R&D, to reduce the unjustified duplication and overlapping of R&D projects and dissertations.

References

1. 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.

2. Pavlov, L.P. Literatura gris rusa en un mundo digitalizado e informatizado. - "Ciencias de la Informacion." – La Habana: IDICT, 2002, agosto, v.33, N 2, pp. 39-42.

3. Pavlov, L.P. Legal Foundations of the Scientific and Technical Grey Literature Development in Russia. - "The Grey Journal". Internat.Journal on Grey Literature, Amsterdam. Spring 2007,vol. 3, N 1, pp.37-43.

4.. Starovoitov A.V., Bogdanov Yu. M., Bastrykin A. M., Pavlov L.P. The Grey System for Monitoring Self-Funded Research. - GL11 Conf. Proc.– GL-conf. series, N 11. 11th Internat. Conf on Grey Literature, 14-15 Dec.2009, Washington D.C.- Amsterdam:TextRelease, Dec. 2009.- 142 p. P.13.

5. Pavlov, L.P. The Commercialization of Research Findings Documented in Grey Literature. - Proc. 5th Internat. Conf. on Grey Literature: Grey Matters in the World of Networked Information. – 4-5 December 2003 Amsterdam/ GreyNet.- Amsterdam: TextRelease, January 2004.VI. – Pp.64 -68.

Enhancing diffusion of scientific contents: Open data in Repositories

Daniela Luzi, Rosa Di Cesare, and Marta Ricci, IRPPS-National Research Council, Italy
Roberta Ruggieri, Senato della Repubblica, Italy

1. Introduction

The free availability of data gathered during research activities is becoming one of the new challenges facing the Open Access Movement. New scientific instruments and technologies used in highly collaborative fields such as molecular biology, astronomy and environmental sciences, make it possible to collect a great amount of data in different formats. Moreover, data are often associated with tools that can aggregate them as well as with direct references to the publications – conventional or non-conventional – that report the results of their analysis. The benefits of the availability of these data are evident, and include assessment of research results, along with the reproduction and re-utilisation of data, potentially to draw new insight for future research.

According to the National Science Foundation: "digital data are the currency of the data collection universe, which, like currency in the financial realm, comes in many different forms". They are different in nature, generally depending on the very specific field of study; they are produced for different purposes using varying methods and/or instruments; they have their own lifecycle before they are "translated" into scientific results and diffused in scientific publications. Understanding all these aspects makes it possible to determine whether to preserve them and how, who is responsible for their curation and/or diffusion, what type of archive, or better infrastructure, should be developed. This in turn implies issues related with data ownership, as well as funding resources, types of institutions and services to be involved.

Several policy papers (NSF, 2005, OECD, 2007, US National Research Council, 1995; 1999) are advocating free access of datasets and are outlining recommendations to coordinate efforts for the development of successful data repositories and infrastructures. What is clear is that "one-size-fits-all approach to policy development is inadequate" (NSF, 2005).

That is why debate on data ranges from the analyses on issues related to data sharing (Gold, 2010, Piwowar et. al., 2010, Piwowar, 2011) to studies in specific scientific fields (NIH, 2003, 2007, Karasti, et al., 2006, Baker et al., 2009, Waaijers et al., 2011) including surveys on usage patterns (Brown, 2003, Piwowar et al., 2007,) and researchers' attitude to make them available (Savage CJ, Vickers AJ (2009).

A few studies deal with the analysis of the existing dataset archives and compare their different characteristics (Marcial, 2010). Our paper intends to follow this type of survey, but with a different approach. In fact, the decision to use data archives listed in OpenDOAR enabled us to select a random sample given by the providers that had registered their archives in OpenDOAR. This approach throws light on an emerging reality such as IRs, that theoretically at least, have started to include datasets along with other digital objects. Insight can also be gained into archives of large scale and well-established datasets. Clearly, the adoption of a random sample affected our survey. In contrast to Marcial's empirical survey mentioned above, that found a cluster of elements common to different archives, our study revealed elements of dataset archives listed in OpenDOAR, in order to bring them into line with traditional archive classification (Armbruster & Romary, 2010). This enables the tracking of possible trends in dataset archive expansion policy.

In this paper we present the result of an exploratory analysis of a dataset archive in OpenDOAR. After the dataset definition given in paragraph 2., we describe the method used to select the sample and their main variables. In the fourth paragraph the results of our survey are reported.

2. Dataset definition

Research data are complex objects (Borgman, 2010) and that explains why there is no common agreed definition. They are very generally described and definitions, especially those reported in policy documents, include a very broad variety of digital objects (see Box 1). This is evident if we consider the definition given by the U.S. National Research Council in 1995, where research data are exclusively associated with numerical quantities. Following definitions encompass a wider range of digital objects (for instance images, sounds, etc.), thus representing research outputs in all scientific fields.

A more general agreement is reached when it comes to the definitions of dataset, considered as a meaningful and systematic representation of the subject being investigated. What is importantly stressed here is the importance of its re-use for validation and future investigations.

In this paper the term dataset is used to denote the digital collections managed in data archives.

Box 1. Data definitions

Research data definitions

- *National Research Council (1995):* Data are numerical quantities or other factual attributes derived from observation, experiment or calculation *(http://www.nap.edu/catalog.php?record_id=4871)*
- *National Research Council (1999): Data* are facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors. *(http://www.nap.edu/openbook.php?record_id=9692&page=15)*
- *National Science Foundation (2005):* The term 'data' is used in this report to refer to any information that can be stored in digital form, including text, numbers, images, video or movies, audio, software, algorithms, equations, animations, models, simulations, etc. Such data may be generated by various means including observation, computation, or experiment. *(http://www.nsf.gov/pubs/2005/nsb0540/start.jsp)*
- *OECD (2007):* Research data are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. *(http://www.oecd.org/dataoecd/9/61/38500813.pdf)*
- *PARSE.INSIGHT (2009):* Digital research data is used for all output in research. In practical terms, raw data, processed data and publications are all covered by the same term. A distinction between these sorts of research data is only made when necessary (for example when policies for publications are compared with other data). *(http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf)*
- *HLWIKI Canada (2011)* Research data is often defined as the information (e.g. data sets, microarray, numerical data, clinical trial information, textual records, images, sound, etc.) generated or used as quantitative evidence in primary biomedical research. This research data is distinguished by the fact that it is accepted by the research community as a means to validate research findings, observations and hypotheses. *(http://hlwiki.slais.ubc.ca/index.php/Data_curation)*

Dataset definitions

- *ODLIS (Online dictionary for library and information science):* A logically meaningful collection or grouping of similar or related data, usually assembled as a matter of record or for research, Also spelled *dataset*. *(http://www.abc-clio.com/ODLIS/odlis_A.aspx)*
- *OECD (2007):* A research data set constitutes a systematic, partial representation of the subject being investigated *(http://www.oecd.org/dataoecd/9/61/38500813.pdf)*
- *DOE (Department of Energy):* No-text scientific and technical information *(http://www.osti.gov/data/index.shtml)*
- *University of Edinburgh:* A set of files containing both research data and documentation sufficient to make data re-use. *(http://datashare.is.ed.ac.uk/)*

3. Methods

The information source of our analysis was the directory OpenDOAR (The Directory of Open Access Repositories) that currently lists more than 2000 repositories worldwide providing a detailed description of each of them. OpenDOAR categorizes and provides access to Institutional and Subject-based repositories, but also includes open access archives developed by funding agencies, governmental institutions and digital libraries.

The inclusion of different types of archives allowed us to analyse:

- Types of archives that collect datasets;
- Types of providers;
- Relationship between dataset characteristics and types of archives.

Moreover, OpenDOAR archive description, built on the information submitted by their providers, are then categorized, allowing users to sort the listed archives according to different criteria. We used the option "dataset" reported in the OpenDOAR content type categories to identify our first sample of analysis.

The purpose of our analysis was to track current trends in the development of data archives in the general framework of open access repositories, using the random sample provided by OpenDOAR listed archives. For these reasons, the OpenDOAR archive classification needed to be supplemented with the additional categories: Directory and Digital library. This was necessary in order to group archives with features different from "traditional" IRs or Subject-based repositories. Moreover, the category Digital library was introduced, even if limited to a single case, to show trends in data archives provided by libraries that may make their collections available and re-usable in digital forms.

The second step of our analysis concerned the identification of datasets provided in each archive of the OpenDOAR sample, which was performed searching for dataset, if the archives had this search option, or analysing the archives' collections manually.

According to the NSF definitions for data origin and digital data collections the sampled archives were analysed in terms of:

- Data origin (experimental, observational, computational)
- Types of Data collection (Research data, Resource or community, Reference data collections)

Moreover, datasets were classified as follows:

- Dataset content (Numeric, Scientific image, Image of artifacts, Maps, text-image)
- Dataset format
- Contextual information associated with datasets ("traditional documents", project descriptions, etc.).

Archives with a limited number of datasets (i.e. > 4) were excluded, as they were considered not representative of a stable commitment to dataset collection. Moreover, archives containing video, audio or other multimedia, were not considered in our analysis, this can be the subject of further analysis.

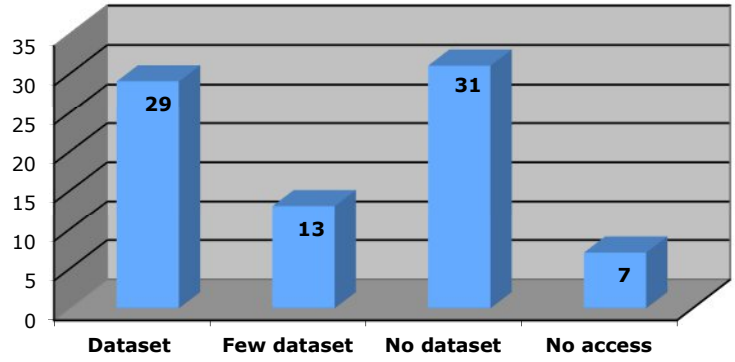
The latest update of our analysis was completed in October 2011.

4. Results

4.1. The sample

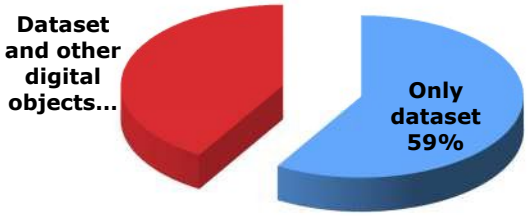
In OpenDOAR there are 80 archives that claim to contain datasets in their content type. The analysis of each of the selected OpenDOAR archives showed that only 29 out of 80 actually contain datasets, while 13 archives were discarded for the limited number of datasets available. In 33 archives no datasets at all were found, whereas the remaining 7 archives were not accessible (fig. 1).

Fig. 1. - Dataset archives listed in OpenDOAR



Our sample of analysis consequently numbers 29 archives. Given the variety of archives listed in OpenDOAR, it should be noted that 59% of them (equal to 17 archives) exclusively contain datasets, while 41% (equal to 12 archives) contains both datasets and other digital objects, such as journal articles, reports, theses, etc. (fig. 2).

Fig. 2 Sampled archives by type of digital objects (n= 29)



4.2. The data archives' providers

In the analysis of the type of providers that insert datasets in their archives we also wanted to verify whether they are single institutions or have built consortia. Our hypothesis is that consortia may have developed internal rules, specific metadata and or format to describe and exchange data to be shared within a specific scientific community. Results are reported in table 1.

Table 1. Dataset providers by type of organisation

Research Institution	
Single (15)	**Consortium (5)**
• *Chiba University, JP* • *Spanish National Research Council, ES* • *Cambridge University Library and Computing Service, UK* • *Data Archiving and Networked Services (DANS), NL* • *University of Southampton (Soton), UK* • *Data Library, University of Edinburgh, UK* • *Inter America Institute for Global Change Research (IAI), BR* • *International Food Policy Research Institute (IFPRI),US* • *University of Minnesota ,US* • *Monash University Library - Australia* • *Scripps Institution of Oceanography (SIO),US* • *University of Delaware Library, US* • *University of Hull ,UK* • *Centre de Données astronomiques de Strasbourg (CDS), FR* • *Marine Biological Laboratory & Woods Hole Oceanographic Institution (MBL & WHOI) Library, US*	• *Mineralogical Society of America, Mineralogical Association of Canada, University of Arizona, Schweizerbart Science Publisher, INT* • *COD Consortium, INT* • *Center for Research Libraries (CRL), US* • *Alfred Wegener Institute for Polar and Marine Research (AWI), Center for Marine Environmental Sciences (MARUM), University of Bremen, DE* • *Department of Geosciences, University of Arizona University of Arizona (UA), CALTECH (California Institute of Technology), US*
Indexing abstracting service	
Single (3)	**Consortium (1)**
• *Archaeology Data Service, UK* • *National Center for Biotechnology Information (CBI), US* • *National Library of Medicine (NLM), US*	• *Ontario Council of University Libraries, CA*
Publisher	
Single (1)	**Consortium (1)**
• *FigShare, UK*	• *Dryad, INT*
Government	
Single (3)	**Consortium (0)**
• *Coordenação de Biblioteca / CGDI / SAA / SE, Minisétrio da Saúde, BR* • *U.S. Department of Energy (DOE), US* • *Deutschen Zentrum für Luft- und Raumfahrt (EDINA), DE*	---

The majority of the providers of dataset archives are research institutions (20 out of 29), among which 8 universities and 7 research institutes.

Datasets are also available in archives developed by Indexing/abstracting services, governmental institutions and publishers. Such providers reflect the growing interest in datasets to be diffused for different purposes. At governmental level, for instance, the request for open data has been met by different countries that are progressively diffusing data collected within their institutions. In our sample we

found the Brazilian Health Ministry, a German governmental agency for transport, and the U.S. Department of Energy that has a long tradition in the diffusion of technical information. Further, the presence of publishers represents the tendency to request datasets together with journal articles. In our sample a consortium of scientific journal publishers has developed Dryad that allows authors to submit their data and connect them with peer-reviewed articles. Similar features are provided by the publisher FigShare that provides citations of the datasets downloaded by authors.

4.3. Types of archives

In the analysis of type of archives we have adopted the traditional distinction between IR and subject based repositories. This classification is influenced by the information source we have chosen for our analysis, that has the advantage of exploring small dataset collections and verifying whether IRs are also beginning to consider datasets in their research results.

We introduced the category Directory to group heterogeneous types of archives, websites of governmental institutions, large databases that provide access to different data sources.

In OpenDOAR we also found an archive in the form of a Digital library, which we included in our analysis because we consider it a good example of providing a re-usable dataset from digitalised documents. In fact the South Asian Digital library not only digitalised an old text containing statistical data from the colonial period, but also provided an excel file that reported the datasets of the document. In our opinion this is a good example of making datasets re-usable, even if they are not digitally born.

Table 2. Archives by type

Subject-based Repository (15)
American Mineralogist Crystal Structure Database – http://rruff.geo.arizona.edu/AMS/amcsd.php
Archaeology Data Service - http://archaeologydataservice.ac.uk/
Crystallography Open Database (COD) - http://www.crystallography.net/
EDNA-the e-depot for Dutch archaeology - http://www.dans.knaw.nl/en/content/categorieen/projecten/edna-e-depot-dutch-archeology
eCrystals - Southampton) - http://ecrystals.chem.soton.ac.uk/
IAI Search - http://mercury.ornl.gov/iai/
Metropolitan Travel Survey Archive - http://www.surveyarchive.org/
PubChem - http://pubchem.ncbi.nlm.nih.gov/
PANGAEA® (Publishing Network for Geoscientific and Environmental Data) – http://www.pangaea.de/
RRUFF Project - http://rruff.info/
ShareGeo Open - http://www.sharegeo.ac.uk/
SIOExplorer Digital Library Project (SIOExplorer) - http://siox.sdsc.edu/
Verkehrsmodelle – http://modelle.clearingstelle-verkehr.de/
VizieR Catalogue Service - http://vizier.u-strasbg.fr/
Woods Hole Open Access Server (WHOAS) - https://darchive.mblwhoilibrary.org/
Institutional repository (7)
Chiba University's Repository for Access To Outcomes from Research (CURATOR) - http://mitizane.ll.chiba-u.jp/curator/
Digital.CSIC - http://digital.csic.es/
DSpace @ Cambridge - http://www.dspace.cam.ac.uk/
Edinburgh DataShare - http://datashare.is.ed.ac.uk/
Monash University ARROW Repository - http://arrow.monash.edu.au/vital/access/manager/Index
University of Delaware Library Institutional Repository - http://dspace.udel.edu:8080/dspace/
University of Hull Institutional Repository - https://hydra.hull.ac.uk/
Directory (6)
Biblioteca Virtual em Saúde - http://bvsms.saude.gov.br/php/index.php
Dryad - http://www.datadryad.org/
FigShare - http://figshare.com/
IFPRI Publications (Int. Food Policy Research Institute Publications) - http://www.ifpri.org/publications
OSTI (Office of Scientific & Technical Information) - http://www.osti.gov/
OZone (OZone provided by Ontario Scholars Portal) - https://ospace.scholarsportal.info/
Digital Library (1)
DSAL (Digital South Asia Library) - http://dsal.uchicago.edu/

4.4. Which science area?

The majority of dataset archives in our sample cover hard science (52%), but there is also a meaningful percentage of archives that provide datasets in Humanities and social sciences (fig. 3).

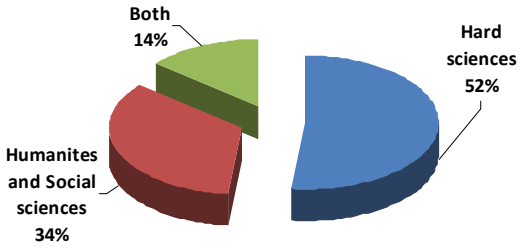


Fig. 3 Distribution of archives by science area (n=29)

If we group them in broad disciplinary fields, the most prevalent are Environment (21 %) and Demography (21%) (fig. 4).

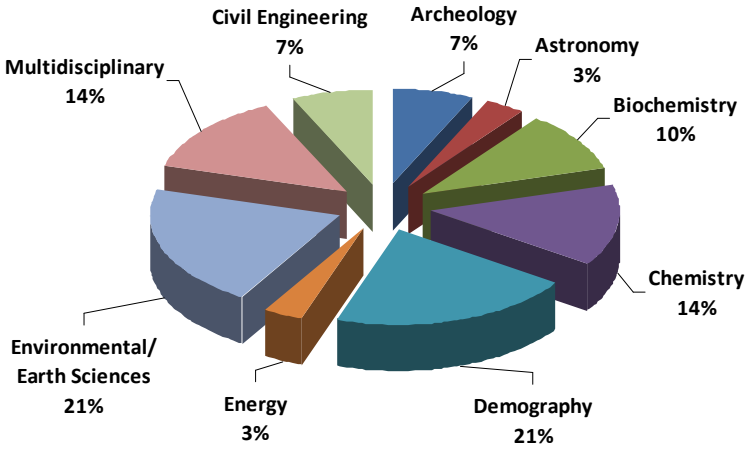


Fig. 4. - Distribution of archives by disciplinary fields (n=29)

Important criteria for the analysis of datasets depend on their origin, that is whether they are produced measuring specific phenomena at a given time, or are generated by experiments, or developing computational models or simulations to predict certain phenomena. Further, these variables are important when deciding whether it is important to preserve the data, considering that some of them cannot be so easily reproduced and/or collected. Figure 5 shows this variable linked with the scientific area.

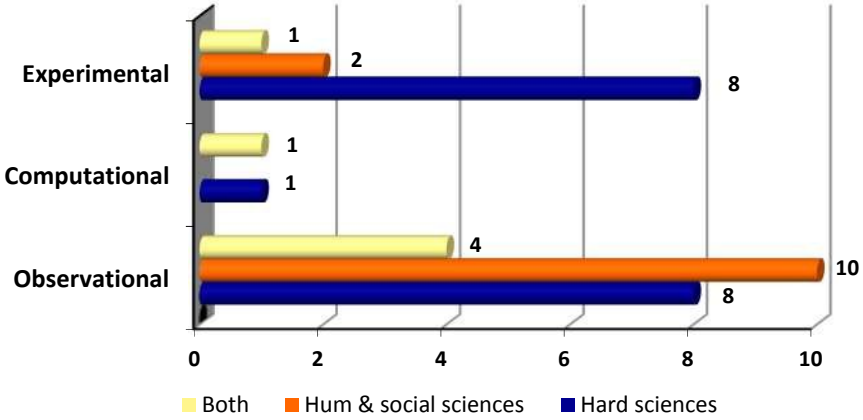


Fig. 5 Archives by data origin

The dataset archives in our sample are predominantly observational, and this is true in all science areas. Experimental data are collected mainly in hard sciences.

4.5. Functional categories of digital data collections

The National Science Foundation introduced three functional categories to analyse data collections referring to databases, infrastructures and organisations and individuals essential to managing this collection (NSF, 2005). This classification aims to distinguish between research data collected within a project of a certain size and budget as well as with different types of funds and funding sources. This distinction is also made to evaluate efforts necessary to preserve and diffuse datasets. Of course, a Research data collection can progressively become a Resource or Reference data collection, this was the case for instance of the well-known Protein data bank.

We applied these categories to the archives listed in OpenDOAR and compared them with the type of archives (fig.6).

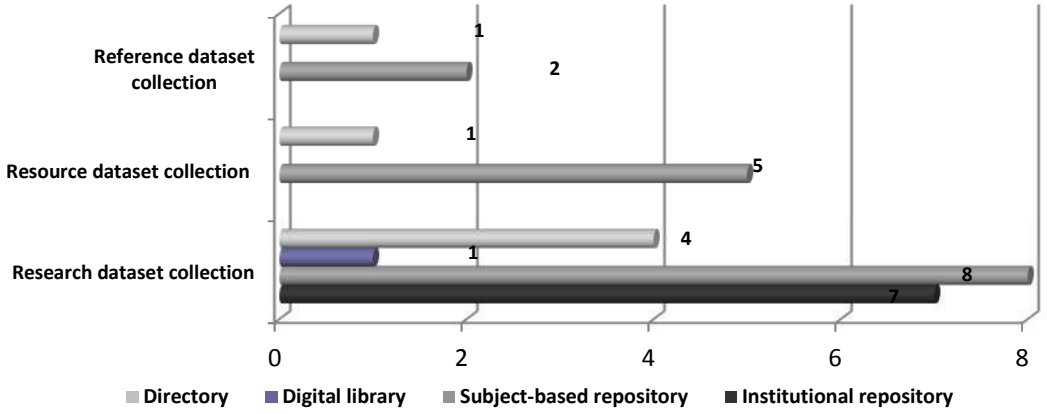


Fig. 6. – Archives by digital data collection

IRs exclusively contain datasets that fall into the category of Research data collections. Subject-based repositories contain datasets in all 3 categories, with a prevalence of Research data collections, while directories contain 1 Reference data collection.

4.6 Dataset content

For each archive in our sample we examined datasets with a view to analysing their content.

Figure 7 shows that the majority of archives contain numeric data, followed by scientific images, maps, text-images (i.e. digitized text) and images of artefacts.

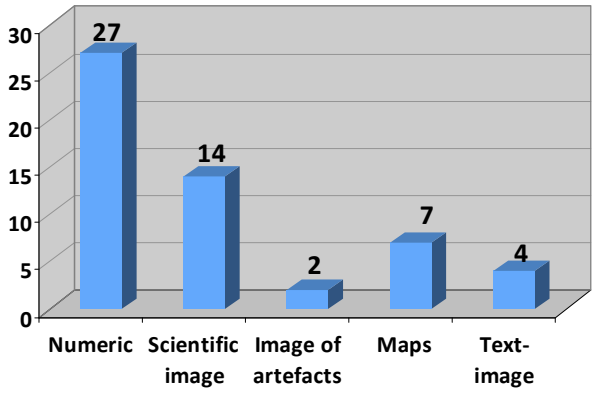


Fig. 7 Dataset content in our sample

Considering that the results of scientific observations, experiments or computational models can be expressed using different representations, not limited to numeric values, we associated the numeric content with other types of representation. Figure 8 shows the number of archives that only contain numeric datasets and/or images and those that associate numeric data with other digital objects.

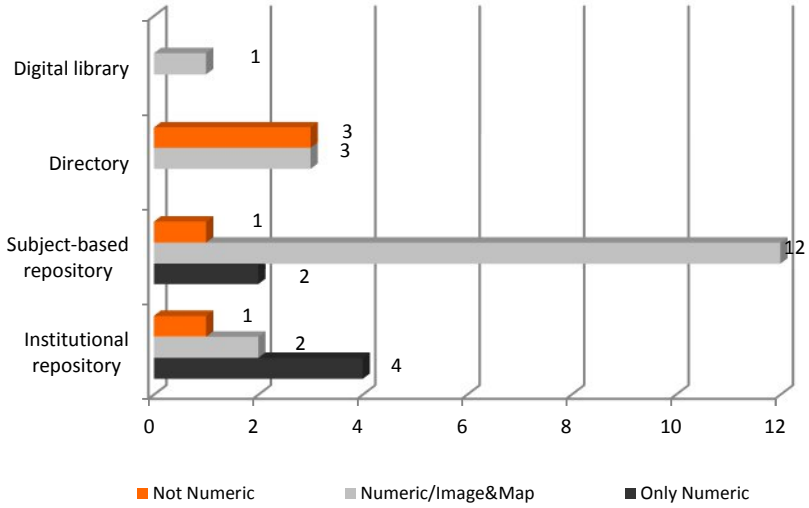


Fig. 8 Dataset content by type of archive

If we relate the content of a dataset with the type of archive, we see that there is a tendency to represent research results through numeric data associated with images. This is evident particularly in the case of Subject-based repositories (12 archives), while Institutional repositories tend to collect only numeric datasets (4 out of 7). Of course depending on the subject, some datasets are represented only by images (i.e. Not numeric) and this is present in all the kinds of archives in our sample.

The research results in the Subject-based repositories of our sample seem to provide a richer representation of datasets as a whole. For instance in the case of crystallography, the crystal structure described in the CIF format (see below) is combined with the graphical representation of its chemical structure, adding value for both crystallographers and chemists. (Cragin et al., 2010)

4.7 Dataset format

On the one hand file formats give evidence of the content of dataset (formats used to view images, texts and/or to store structured data already recognisable from their format extension). On the other, they also show how easily datasets can be exchanged. For instance the use of flat files, that is files that transform a record of a database into text, can be easily exchanged because they are not connected with proprietary systems. The disadvantage of using this format is that one needs to have additional information to interpret the data. In the archives listed in our sample we found different formats and sometimes the same archive provides the dataset in different formats so that users can easily access the data in the format he/she prefers. It follows that data format can also be considered an indicator of sharing and re-use. The formats more commonly used in our sample are reported in table 3.

Table 3. Dataset formats in our sample

File type category	File type/extension
Flat files	.txt, .ascii, .csv
Word processor	.doc, .pdf
Image	.tiff, .jpeg, .gif, .jmol
Spreadsheet	.xls
Statistical analysis	SPSS

As sharing and re-use are crucial for the dataset environment, we also looked for other file formats that facilitate their exchange. We found that some datasets were associated with the so-called readMefile, that contain important information, such as copyright, or how to install the database. We found readMefile especially in IRs (57%) and in directories related to Research data collections (67%).

It is of course the development of a specific standard format to exchange datasets that assures the highest degree of exchange and re-use. Their application indicates that a certain scientific community has a

tradition in data sharing and has already agreed upon an exchange format that has a specific structure and meaning. An example of this standard is the CIF format used in crystallography to describe crystal structures or the standard used in astronomy to describe latitude, longitude and size of astronomical objects. It comes as no surprise that such exchange formats were present in Subject-based repositories (53%) and both in Resource and Reference data collections.

4.8 Datasets and "traditional documents"

Usually datasets are not self-describing, we need to know the context in which they are produced, how, and in which period, etc. Moreover their analysis can be described in other "traditional" documents, such as journal articles, reports, and theses (fig. 9).

In our sample we found that in the majority (72.4%) of archives, datasets are linked with traditional documents, and this is true for all types of archives.

Some archives also connected the dataset with the description of the project in which datasets were collected: this we found especially in large Subject-based repositories. Other archives also described the entire collection and this was the case especially in IRs.

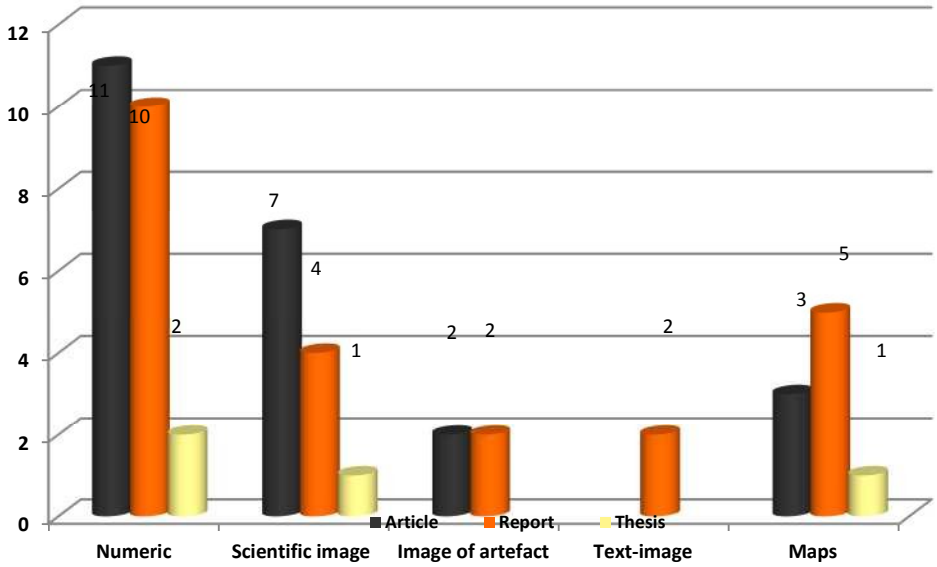


Fig. 9. – Dataset content by document type

5. Conclusions and discussion

Our sample enabled us to determine some common features, but also some characteristics that while not so widespread, may indicate possible trends in the development of dataset archives.

Considering the providers, our sample shows that along with research organisations and data services, governmental institutions and publishers too are developing archives making datasets available to the public. This is in line with the policy on open data announced in some countries as well as with the tendency of publishers to require datasets together with articles they are going to publish. Consortia are also frequently involved in building dataset archives, confirming the importance of collaboration in this field. Further analysis on the types of consortia (based on scientific collaboration, funding resources, and/or organisational models) should be carried out.

Datasets are collected in IRs along with other digital objects, while the majority of Subject-based repositories of our sample contain exclusively datasets. The introduced category Directory represents another way of organising data archives, combining different databases and linking various information sources. In our sample we also had an example of a Digital library that made datasets re-usable and an IR that contained only datasets.

Archives specifically focused on datasets and on specialized sub-disciplinary fields provide a richer environment in terms of data representations, of development of specific formats that facilitate data exchange, and of re-use and links to other digital objects and/or documents. This was evident in the Subject-based repositories and in Directories in our sample. This does not exclude that IRs cannot

contribute to the collection and diffusion of datasets. Certainly, given the variety of datasets and their close relationship with the sub-disciplinary field in which they are collected, this poses different issues, such as self-archiving procedures and attitudes, ownership and copyright of data as well as their updating and maintenance. In this respect, the data collection categories proposed by the NSF provide a useful interpretative key and also suggest procedures to adequately construct and store data collections according to the type of archive and the mission of the archives' provider. In our sample, we had a prevalence of Research data collection in IRs, representing the outcomes of specific scientific projects with a limited user community and budget. Dataset availability in IRs along with other scientific results provide a more complete description of the research activities carried out in scientific institutions, while efforts concerning their visibility and usability should be further improved.

References

Anderson W.L. (2004). Some Challenges and Issues in Managing, and Preserving Access to, Long-live Collections of digital Scientific and Technical Data. *Data Science Journal,* 3.

Arms William Y., Larsen Ronald L. (2007). The Future of Scholarly Communication: Building the Infrastructure for Cyberscholarship. URL: http://www.sis.pitt.edu/~repwkshop/NSF-JISC-report.pdf

Armbruster Chris, Romary Laurent (2010). Comparing Repository Types: Challenges and Barriers for Subject-based Repositories, Research Repositories, National Repository Systems and Institutional Repositories in Serving Scholarly Communication. *International Journal of Digital Library Systems*, 1 (4).
URL: http://arxiv.org/abs/1005.0839

Baker Karen S., Yarmey Lynn (2009). Data Stewardship: Environmental Data Curation and a Web-of-Repositories. *The International Journal of Data Curation*, 4 (2).

Borgman C.L., Wallis J.C., & Enyedy N. (2007). Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries. *International Journal on Digital Libraries*, 7 (1-2).

Borgman L. Christine (2010). Research Data: Who Will Share What, With Whom, When, and Why?
URL: http://works.bepress.com/cgi/viewcontent.cgi?article=1237&context=borgman>

Brown C.M. & Abbas J.M. (2010). Institutional Digital Repositories for Science & Technology Information: A View from the Laboratory. *Journal of Library Administration Special Issue: Emerging Practices in Science and Technology Librarianship*, 50:181–215.

Brown C.M. (2003). The Changing Face of Scientific Discourse: Analysis of Genomic and Proteomic Database Usage and Acceptance. *Journal of the American Society for Information Science & Technology*, 54(10): 926-938.

Cragin Melissa H., Palmer Carole L., Carlson Jacob R., & Witt Michael. (2010). Data Sharing, Small Science, and Institutional Repositories. *Philosophical Transactions of the Royal Society A*, 368(1926): 4023-4038.

Gold Anna (2010). Data Curation and Libraries: Short-term Developments, Long-term Prospects. *Data Curation and Libraries*, 4.

Graaf Maurits van der, Waaijers Leo (2011). KE Knowledge Exchange Primary Research Data Working Group. A Surfboard for Riding the Wave: Towards a Four Country Action Programme on Research Data.
URL: http://www.voced.edu.au/content/ngv48428>

Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council (2009). Harnessing the Power of Digital Data for Science and Society.
URL: http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

Joint Information Systems Committee (JISC) Managing Research Data (MRD) Programme (2009).
URL: http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx

Karasti Helena, Baker Karen, Halkola Eija (2006). Enriching the Notion of Data Curation in E-science: Data Managing and Information Infrastructuring in the Long- term Ecological Research (LTER) Network. *Computer Supported Cooperative Work (CSCW)* 15: 321-358.

Marcial Laura Haak, Hemminger Bradley M. (2010). Scientific Data Repositories on the Web: an Initial Survey.
URL: http://onlinelibrary.wiley.com/doi/10.1002/asi.21339/pdf

National Institutes of Health (2006). Data Sharing Policy and Implementation Guidance.
URL: http://grants2.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

National Research Council (1995). Preserving Scientific Data on our Physical Universe: a New Strategy for Archiving the Nation's Scientific Information Resources. Washington D.C: National Academy Press.
URL: http://www.nap.edu/catalog.php?record_id=4871

National Research Council (1997). Bits of Power: Issues, in Global Access to Scientific Data. Washington, D.C.: National Academies Press. http://www.nap.edu/catalog.php?record_id=5504

National Research Council (1999). A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases. Washington, DC: National Academy Press.
URL: http://www.nap.edu/openbook.php?record_id=9692&page=14

National Research Council (2003). Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences. Washington, D.C.: National Academy Press.
URL: http://selab.janelia.org/publications/Cech03/Cech03-reprint.pdf

National Science Foundation (2005). Long-lived Digital Data Collections: Enabling Research and Education in the 21st century.
URL: http://www.nsf.gov/pubs/2005/nsb0540/start.jsp

National Science Foundation (2011). Dissemination and Sharing of Research Results.
URL: http://www.nsf.gov/bfa/dias/policy/dmp.jsp

OECD (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding.
URL: http://www.oecd.org/dataoecd/9/61/38500813.pdf

OpenDOAR (2011).
URL: http://opendoar.org/

PARSE.INSIGHT (2009). Insight into Digital Preservation of Research Output in Europe.
URL: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

Pirowar Heather A., Day R.S., Fridsma D.B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3): e308. doi:10.1371/journal.pone.000030
URL: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.00003088

Pirowar Heather A., Chapman Wendy W. (2010). Public Sharing of Research Datasets: a Pilot Study of Associations. Journal of Informetrics 4 (2): 148-156. doi:10.1016/j.joi.2009.11.010.
Savage C.J, Vickers A.J (2009). Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PLoS ONE* 4(9): e7078. doi:10.1371/journal.pone.0007078

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E., Manoff M. Frame M. (2011). Data Sharing by Scientists: Practices and Perceptions; PLoS ONE (6)6
URL: http://dx.doi.org/doi:10.1371/journal.pone.0021101

UK data archive (2011). Managing and Sharing Data. [3rd ed].
URL: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf

Waaijers Leo, Graaf van der Maurits (2011). Quality of Research Data, an Operational Approach.
URL: http://www.dlib.org/dlib/january11/waaijers/01waaijers.html

Whitlock M.C. (2011). *Data Archiving in Ecology and Evolution: Best Practices. Trends in Ecology & Evolution* 26: 61-65. doi:10.1016/j.tree.2010.11.006.

Research product repositories: Strategies for data and metadata quality control

Luisa De Biagi, Massimiliano Saccone, and Luciana Trufelli
CNR Central Library 'G. Marconi'
National Research Council of Italy

Roberto Puccinelli,
CNR Informative Systems Office
National Research Council of Italy

Abstract
In recent years a significant effort has been spent by R&D institutions and scientific information stakeholders in general to enhance and improve the quality of Open Access initiatives and the performance of the associated services. Nevertheless much work is still needed to tackle pending data quality issues.
This paper proposes some functional and organizational solutions, based on the cooperation of all the main actors of the R&D system, which in our view should help improving quality control of data and descriptive metadata stored in research product Open Access (OA) repositories. We think that this strategy could favor a substantial innovation of the document management services offered to the scientific community and to policy makers, ensuring the interoperability between institutional repositories and Current Research Information Systems (CRIS).
Particular emphasis is given to the problem of data and metadata indexing and organization with respect to unconventional research products, which represent an important asset in the field of scientific communication.

Introduction

In Europe, despite the efforts of the scientific community and of many expert groups, effective methods and tools for R&D performance evaluation are still not available. This, in our opinion, is a top-priority issue, since reliable measurements are a pre-condition for credible process and product quality assessment.[1]
In this paper we propose a cooperative organizational approach for tackling some crucial challenges, such as research product metadata quality certification, with particular focus on metadata stored in Open Access repositories (OA).
Currently some national evaluation systems[2] [3] leverage data coming from institutional repositories, which are integrated within R&D Information Systems[4]. Disciplinary and institutional repositories can be used as data sources for R&D performance measurement, also because they keep products which are highly representative of the different scientific communities.[5]
Another interesting (but sometimes neglected) aspect of institutional repositories is that they can collect, index, keep and disseminate grey literature products. The availability of certified data about those products could provide new perspectives to science and technology phenomena investigation.[6] Actually, grey literature products could be used as a significant evaluation set both for bibliometric analysis and for investigations aimed at understanding science and innovation dynamics, change driving ideas, knowledge basis used in particular scientific developments, connections and communication patterns in particular disciplinary contexts.
In general, we think that cooperative systems facilitate the traceability of the different research product life-cycle phases and of the related metadata (versioning, persistent identification, etc.). The cooperative approach should be further extended within the scientific community to quality certification by adopting open and transparent peer-review processes (open peer review, open peer commentary, etc.).

Open Access repositories in R&D information system: strategic role of cooperation

Open Access repositories, whose number has been steadily rising in recent years, are an important component of the global e-Research infrastructure.[7] The real value of repositories lies in the possibility of interconnecting them to create a network that can provide unified access to research outputs and be (re-) used by OA service providers, researchers' communities, management information systems (CRIS)[8],

statistical information systems, bibliographic databases, etc.[9] However, in order to achieve this goal, a *multilevel* interoperability is needed. The purpose of this paper is to provide a broad overview of multilevel interoperability between Open Access repositories and other R&D information systems, identify the major issues and challenges that need to be addressed, stimulate the engagement of the repository community and trigger a process that will lead to the establishment of a cooperative network of R&D information management systems.[10]

Today, Open Access repositories are increasingly being used to collect, archive, and disseminate all types of research outputs such as research articles, conference proceedings, dissertations, data sets, working papers and reports.

Currently, research product data and metadata managed by OA and commercial repositories and databases are not used for official statistics due to several problems, such as the influence of the different national policies and strategies on the scientific production; the lack of a coherent framework of commonly agreed strategies; the different methods, tools and criteria used to collect data within the different public and private organizations; the lack of common classification criteria for product types, semantics and fields of reference; the insufficient reliability of data provided by the main bibliographic data bases (data base structure issues, lack of bibliographic & authority control tools, etc.); and more.[11]

The research process is an international and distributed endeavor, involving a variety of stakeholders such as scientists as authors and grant recipients, policy makers, research institutions, universities, publishers, and research funding agencies – each with their own set of interests. An international collaboration is needed between these stakeholders (actors) in order to develop cooperative and dynamic methodologies and processes for data and metadata quality control.

Interoperability is a pre-condition for a cooperative and widespread infrastructure of R&D information systems and for the value-added services and tools that can be built on top of the repositories.[12] The quality of these services depends on the data provided by repositories/CRIS/other information systems and on the standardization of "quality control processes" (quality of data and metadata collection and management processes).

Given the quantity and complexity of the problems affecting what in a broad sense could be called the R&D international information system, it seems evident to us that the interoperability should be implemented not only at the technical level but also at the political and organizational ones by all the institutions involved in the creation, management and use of the information resources.

Data and metadata model standardization is necessary in order to enable efficient data exchange and to allow researchers to find the desired information in the different research management systems.

From a strategic view point, the development of common logical and organizational data and metadata models i*n the Scientific and Research System* is important for:

- giving a simplified view to describe the specific area of interest;
- allowing for a better communication and multilevel interoperability between different information systems (Current Research Information Systems[13], Institutional Repositories, OA Service Providers, public and commercial Bibliographic databases, statistical databases, etc.);
- supporting *information workflow management;*
- supporting management and evaluation activities.

The aim of such cooperation should be the development of a common multilevel interoperability network and the first step should be a survey about policies and guidelines for organization and workflows, available data and metadata standards, cooperative bibliographic, authority control and subject access systems, formats and access conditions, data use and re-use patterns, in order to gain sufficient insight into the scale of interoperability problems. Only on such basis, that is actual options, effective solutions can be developed and deployed.

The multilevel cooperation is necessary at the following levels[14]:

- Political: effective initiatives are needed at the national and international levels to favor open access to research results achieved through public funding; those initiatives should address and harmonize the different R&D stakeholders' interests;
- Institutional: academic and research institutions should define institutional and operational policies and carry out effective and widespread advocacy actions in their reference communities.

- *"For institutional record-keeping, research asset management, and performance-evaluation purposes, and in order to maximize the visibility, accessibility, usage and impact of our institution's research output"*[15];
- Economic and legal: Open Access is not zero-cost. Economic strategies are needed to sustain open access to public research products, based on the "author/institution pay" model; on the legal side, the adoption of Creative Commons (CC) licenses should protect intellectual property rights while granting open access;
- Technical-organizational: standards and commonly-agreed guidelines (based on a cooperative approach) are needed to certify data and metadata quality;
- Technological: OA greatly benefits from the development and widespread adoption of open standards and protocols and from the development of modular, interoperable and open source-based platforms for the management and diffusion of digital contents.

Green road: institutional and disciplinary archives

> *"…Two roads diverged in a wood, and I --*
> *I took the one less traveled by,*
> *And that has made all the difference."*
> (Robert Frost, The road not taken, 1920).

As a matter of fact, we could poetically say "*two roads to OA diverged in the wood of 'online scientific publishing*":

- the "golden road" of OA journal-publishing , where journals provide OA to their articles (either by charging the author-institution for refereeing/publishing outgoing articles instead of charging the user-institution for accessing incoming articles);
- the "green road" of OA self-archiving, where authors provide OA to their own published articles, by making their own eprints free for all.

In our opinion, the Green Road is the one that could bring more benefits to the scientific community.
One of the main research access/impact problem is that journal articles are not accessible to all potential users, causing a lack of potential research impact. The solution is making all articles really Open Access, granting a free, immediate and permanent online access to the full text of research articles for anyone, anywhere, webwide.
On the other hand we should consider the two roads to OA complementary, as well: the green road, representing the fastest and safest way to reach immediate 100% OA, might eventually lead to gold too.
In fact OA self-archiving is not self-publishing without quality control; nor it is meant to be scientific documentation for which the author could request payment and royalties (e.g. books or magazine/newspaper articles). OA self-archiving is bounded to peer-reviewed research, written only for research impact rather than royalty revenue[16].

The main consequence of a wider OA diffusion is that the whole society could benefit from a faster information spreading and from an accelerated research cycle through channels in which researchers can immediately satisfy their needs. It has been proved that OA articles have a significantly higher citation impact than non-OA articles. Only 5% of journals are gold, but over 90% are already green (the green light to self-archiving is possible and authorized to authors); yet only about 10-20% of articles have been self-archived. To reach easily the '100% OA' goal, self-archiving needs to be mandated by researchers' employers and funders, as U.K. and U.S.A have recently recommended, and universities play a significant role in that. It is crucial that both funders and universities/research-boards mandate Green OA self-archiving, as not all research is funded and repositories are successful in attaining a considerable percentage of self-archiving only where a mandatory policy has been issued and enforced.

The main benefit supplied by OA, in general, and Green Road, in particular, is that researchers can increase visibility, usage and impact of their own findings, as well as their chance to find, access and use results from other researchers. On the other hand, Universities co-benefit from the increased impact of their researchers, because it also gives an excellent return on the investment to research funders, such as

governments, charitable foundations etc. Finally, publishers likewise benefit from the wider dissemination, visibility and higher journal citation impact factor of their articles, and Open Access can generate new metrics to be used for assessing and improving research impact.

OA and grey literature valorization

Grey literature plays a significant role in the context of scientific documentation managed and diffused through Open Access archives, indexed and aggregated by the main service providers. Since the Seventh International Conference on Grey Literature at Nancy in 2006, GreyNet community started increasing its research activities relating to the OA effect on grey literature.

The adoption of open standards and OAI protocols by the International OpenGrey network facilitates the interoperability between OA repositories and OpenGrey (System for Information on Grey Literature in Europe). That's a first important step in developing cooperative networks for data and metadata certification.

The diffusion of the International Open Access initiative might certainly facilitate the development and coordination of cooperative networks, implementing sustainable processes and guidelines for:

- a better quality certification of grey literature products (open peer review, open peer commentary, etc.) and related metadata (adoption of common metadata standards and mappings, cooperative bibliographic and authority control, versioning, persistent identification systems, etc.);
- a better intellectual property protection especially for multimedia materials, containing a significant percent on Education, Learning and Professional Training (Creative Commons License is still weak). Moreover, a significant number of 'grey' production - as pre-prints, fact sheets, standards and working papers, committee reports, dissertation and Phd thesis - , still gets a discontinuous or null visibility due to intellectual property rights[17];
- a better information to users about copyright constraints (when and in which terms could I use it?);
- a wider access to research products, which can improve their visibility and impact.

Integrating Grey and Peer-reviewed literature often hosted in IR would enable a global view of the total available sources in a given scientific field, as well as an enhancement of research output measurements and metrics. Finally, it would also give increased researcher and affiliation visibility and (most importantly) better research outcomes.

Quality control: strategy, methods, processes and tools

Bibliographic standards and authority control tools are not sufficient to assure data and metadata accuracy, completeness and consistency.

Quality management systems are needed to define processes for the production and management of data and metadata (Trusted Digital Repositories) [18], which imply commonly agreed organizational models[19].

Only a shared effort can guarantee:

- Quality certification of the main data and metadata production and management processes;
- Commonly agreed bibliographic and authority control tools for metadata certification[20];
- Highly customizable software solutions, based on open standards and platforms.

In our opinion, after defining policies, strategies, services[21], methodologies and processes, the cooperative effort should be focused on the design and implementation of technical and organizational solutions able to support interoperability between the different R&D information Systems[22]. To achieve this goal it is important to:

- adopt a web service-based architecture (as in the JISC Information Environment Architecture);
- use open source software for information & content management systems (CRIS) and digital repositories (DSpace, E-prints, Fedora, JDIAM, Alfresco, etc.);
- use standard protocols and solutions for harvesting, aggregation, deposit, retrieval, cross-linking and context-sensitive linking (e.g. OAI-PMH and OAI-ORE[23], SRW - Search & Retrieve Web Service, SRU –

Search & Retrieve URL Service, SWORD - *Simple Web-service Offering Repository Deposit,* Open URL[24], *etc.*);

- define an optimal set of context metadata, make sure these metadata are stored in CRISs and create automatic procedures for transferring these metadata to the repositories (CRIS-driven repositories – see also CERIF Metadata Model[25]);
- define common intermediary XML schemas for complex applications, in interoperable semantic and syntax context, for metadata interoperability, which allows flexible granularity[26];
- use interoperable record formats and syntaxes (e.g. SGML, XML, XML-RDF, XML-MARC, XML-MODS, XML-METS, etc.);
- use common standard models for web based interchange (e.g. RDF[27])
- participate to and leverage experiences from the cooperative development and use of Knowledge Organization Systems in the context of the semantic web (thesauri, classification schemes, subject heading lists and taxonomies, etc.) [28];
- enable citation metadata automatic detection within publications; work out/implement various multilingual controlled vocabularies (content international classifications) for the information objects in the Scholarly and R&D Information Domain (work out - or fill - the CERIF semantic layer)[29];
- define and use common research product categories and types (for example, CERIF – result–publication classification); [30]
- develop a cooperative bibliographic and authority control[31] system for Institutional Repositories and CRISs;
- develop cooperative multi-version control systems[32];
- extensively use specific unique and persistent identification codes:
    - for the different research product types (Handle, URN, DOI, Open DOI for dataset[33], SICI, ISBN, ISRN, ISTC, etc.);
    - for the researchers (international author ID, ORCID[34], etc.);
    - for research information space, CERIF entities being the core;
    - for institutions and projects (international Digital Institution Id – DII - and international Digital Project Id - DPI);
- develop a cooperative Persistent Identifiers (PI) resolution system (meta-resolver for PI)[35];
- develop cooperative semantic and meta search and discovery systems and tools[36].

CNR IA: a viable solution

In this section we will describe the situation of CNR research product archives, the current initiative aimed at implementing an Institutional Archive of research products and viable solutions to accomplish this task. A brief description of the CNR library system is given below, in order to allow a better understanding of the IA discussion.

CNR's library infrastructure reflects CNR's organization, featuring a Central Administration in Rome and a Scientific network made up of thematic institutes distributed all over the national territory. A significant percentage of CNR's institutes are hosted inside territorial Research Areas, which provide common services thus increasing efficiency.

CNR's library system features a hierarchical and distributed organization, which includes a Central Library (Biblioteca Centrale), Research Area Libraries (Biblioteche delle Aree di Ricerca), Institute Libraries (about 80). It provides a wide range of services to the entire scientific community and has recently adopted new organizational measures in order to increase the coordination of its different branches and improve the quality of the services provided to the internal scientific community. This effort has already produced some results in terms of process rationalization and digital resource sharing. The medium term objective is to complete the integration between CNR's libraries and to provide new added value services both to the internal and external scientific community.

At present, within our institution there are some research product archives but an Institutional Archive is not available. The existing repositories are based on open source platforms and are all OAI-PMH enabled.

An ad hoc working group has been established in order to define the architecture, standards, workflows and rules of a unified Institutional Archive. This group includes the personnel which has been involved in the development and management of the existing archives. The new architecture will be based on open

standards and open source platforms. Web service interfaces will be provided for the communication with other systems.
From the researchers' perspective, auto-archiving will be implemented and favored. Obviously several levels of control will be enforced, in order to assure content and metadata quality. To this end, we think that the whole CNR library system should be involved, in order to have a first formal control at the local level (institutes and research areas) and a second one at the central level (Central Library). On the other hand, quality control will be automated where possible, leveraging the quality control strategies, methods, processes and tools described in the previous sections.
One of the main benefits for researchers will be the possibility to produce certified lists of their own publications (e.g. for internal career advancement procedures). We think that this could be a good incentive for self-archiving.

IA integration with CNR IS

Thanks to the web service based interfaces, the new system will be integrated with CNR Information system. Figure 1 shows the high level architecture of CNR IS. The new Institutional Archive is positioned in the right bottom corner.

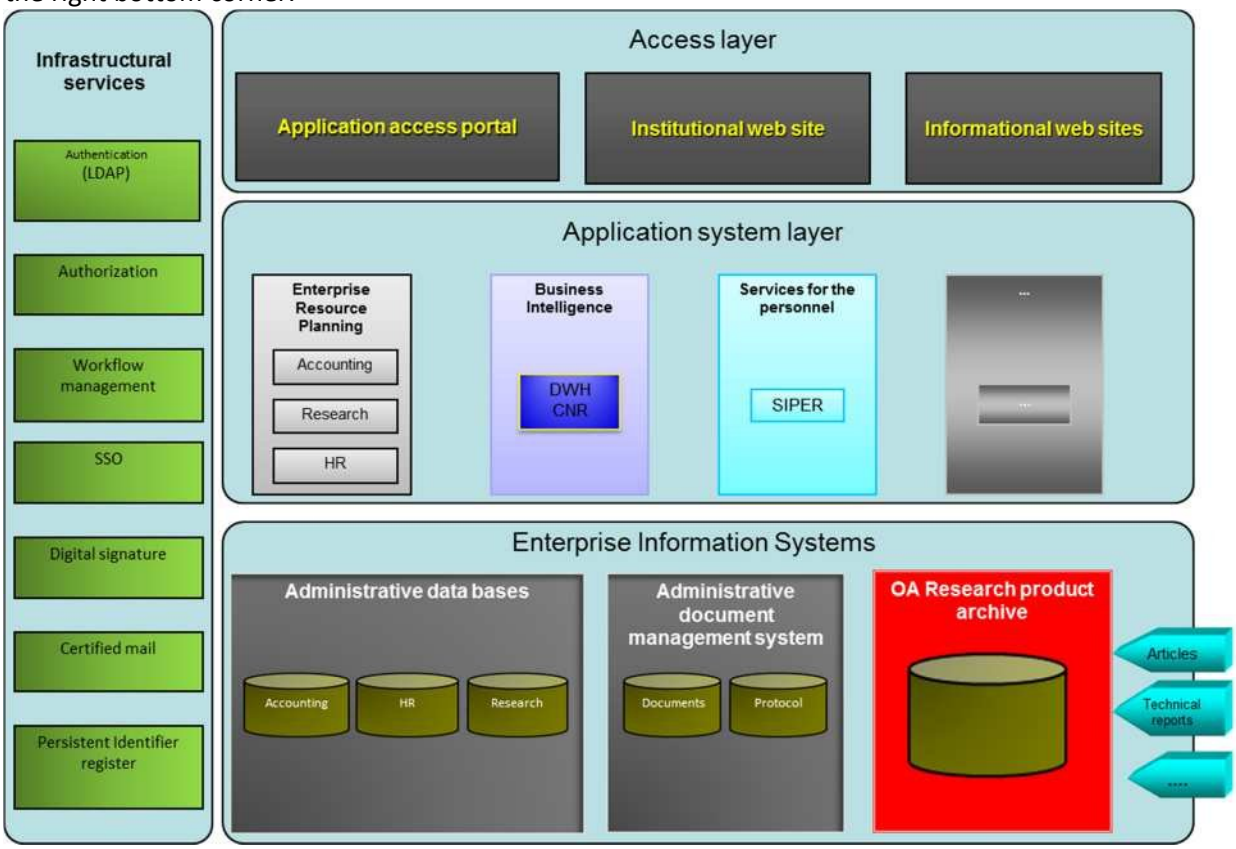


Figure 1: CNR IS high level architecture

At the bottom of this architecture there is the Enterprise Information System layer, which includes the administrative data bases and document management systems. The new IA will be positioned at this level. The Application System layer includes all the systems and applications that manage or analyze the data kept at the underlying level. The Access layer includes all the portals and websites that provide access to services and information residing in the Application layer. Orthogonal to the described layers there is the Infrastructural Services one, which provides cross-application services to the entire IS, such as authentication, authorization, single sign on, etc..
Particular care will be put in implementing an actual interoperability of the new IA with other internal and external systems. The reference schema for interoperability will be the EuroCRIS one, described in Figure 2 (single institution) and Figure 3 (inter-institution interoperability).

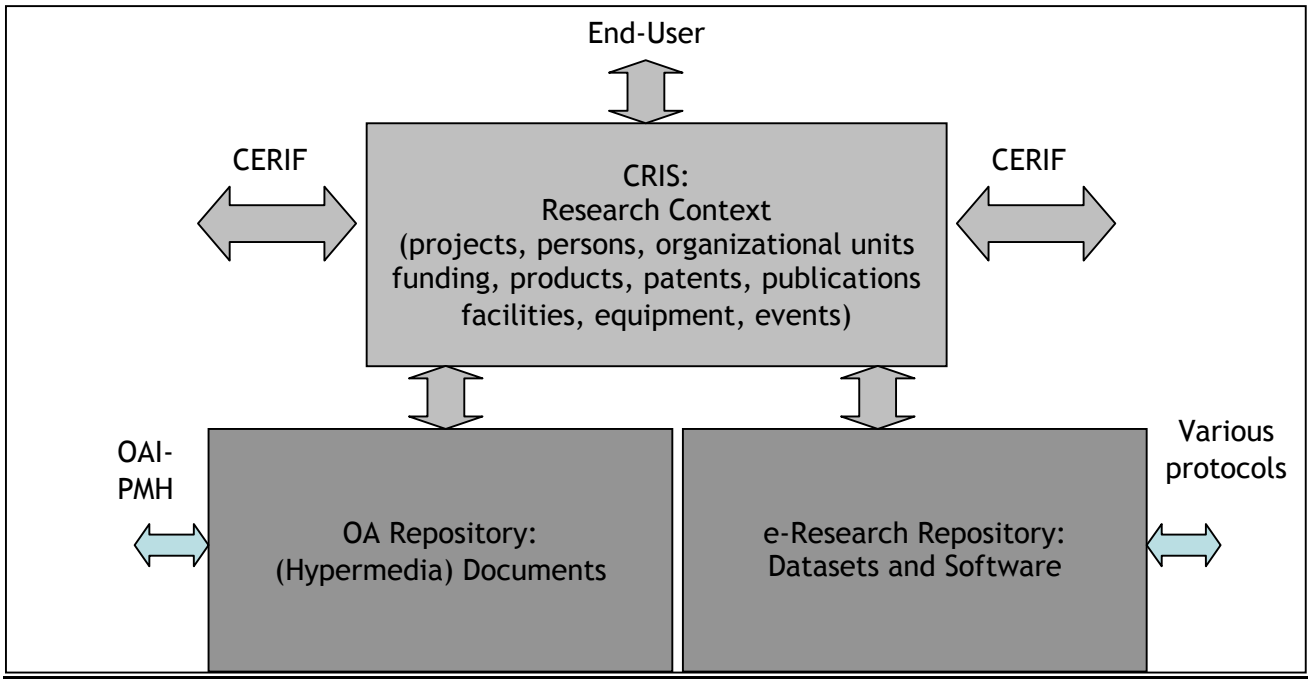


Figure 2: EusroCRIS schema - Architecture for a single institution[37]

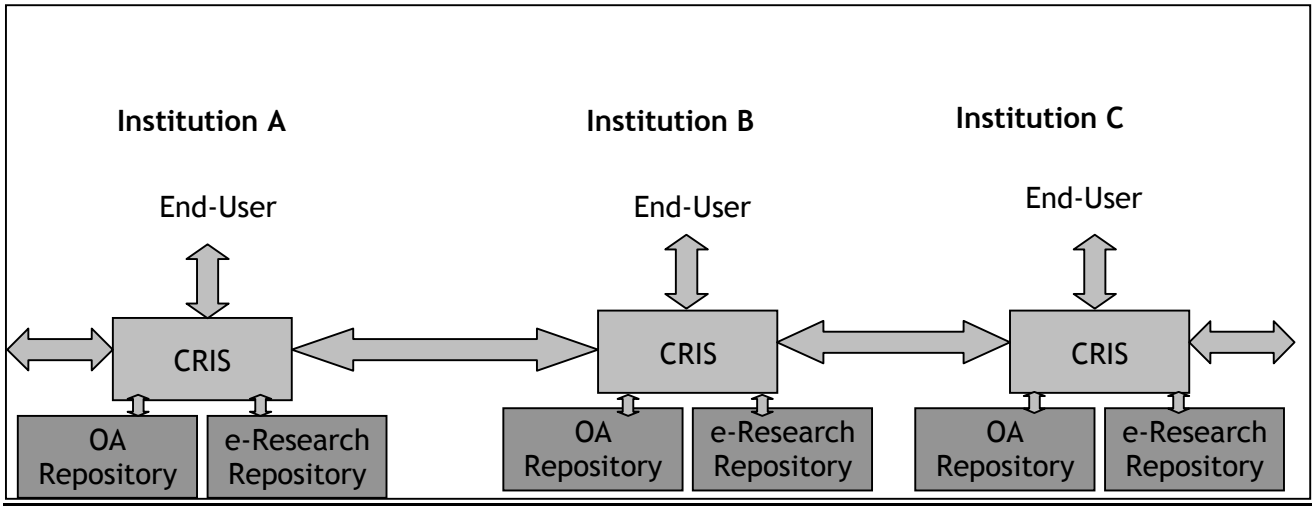


Figure 3: EuroCris schema for CRIS and OA/e-research repositories interoperability[38]

As regards the communication between systems, the figures clearly show that OAI-PMH will play a significant role at the repository level whereas CERIF will be the standard of choice for inter-CRIS communications.
Last but not least, persistent identification of digital resources, authors, institutions, projects, etc. will be taken in due account as well as standards for product classification.

Conclusions and future work
We think that it is important to be aware that the organizational and technical problems regarding multilevel interoperability are currently being discussed and addressed (or have been discussed and addressed in the past) in several other contexts[39], which are partly overlapping with the (digital) library community[40]:

- World Wide Web Consortium (W3C) (communities and working groups for interoperability);
- EuroCRIS – the European Organization for International Research Information (community for Current Research Information System interoperability)[41];
- the OAI (Open Archive Initiative) community (open archives and service providers based on harvested metadata according to the OAI-PMH, OAI-ORE protocols);
- institutional repositories/OA disciplinary repository networks (OpenAire, COAR[42], etc.);
- the Grey Literature Network Service and the OpenGrey - **multidisciplinary European database;**

- scholarly networks for Open Access publishing initiatives (SPARC - *Scholarly Publishing and Academic Resources Coalition, DOAJ* - Directory of Open Access Journals, OAPEN - *Open Access Publishing in European Networks, etc.);*
- Knowledge Exchange[43].

We think that we should learn from these communities and start with them discussions and common developments. The reason is not only the high similarity of data, services and ambitions, but also the fact that scientific products and data will be shared in all of these international contexts, thus requiring basic metadata to be produced only once, close to the source, and be re-used and augmented in other service contexts.

In our opinion, initiatives should be launched at the international level in order to:

- analyze new service scenarios/use cases for records and services or adapt existing ones;
- establish permanent cooperation for on multilevel interoperability involving R&D information system communities[44];
- establish international agencies or cooperative networks[45] for the definition and maintenance of commonly agreed workflow systems, principles, rules and vocabularies.

Within the Italian R&D system we are currently addressing the interoperability issue between the various information systems, also following the stimulus provided by recent laws and rules in the field of research evaluation. Within this context, OA archives are acquiring a great relevance thanks to their role of research product management systems and institutional data sources. In order to assure content reliability, a common effort is required for the development of cooperative certification systems.

References

1. Organisation for Economic Co-operation and Development (OECD), ***Frascati manual 2002 : proposed standard practice for surveys on research and experimental development : the measurement of scientific and technological activities*****,** Paris, OECD - Organisation for economic co-operation and development, 2002, ISBN 92-64-19903-9;
2. Carr, Leslie; MacColl, John, *IRRA (Institutional Repositories and Research Assessment): RAE Software for Institutional Repositories*, IRRA, 2006, http://irra.eprints.org/white/;
   Open Access to research outputs: final report to RCUK, LISU and SQW consulting, 2008, http://www.rcuk.ac.uk/documents/news/oareport.pdf ;
   Open Access to research outputs: annexes: final report to RCUK, LISU and SQW consulting, 2008, http://www.rcuk.ac.uk/documents/news/oaannex.pdf;
3. Dijk, Elly, *NARCIS: linking CRISs and OARs in the Netherlands: A matter of standards and identifiers,* position paper presented at the *EuroCris Workshop on CRIS, CERIF and Institutional Repositories, CNR, Rome, 10-11 May 2010*, http://depot.knaw.nl/6365/;
4. Confederation of Open Access Repositories (COAR). Working Group 2. Repository Interoperability, *The Case for Interoperability for Open Access Repositories. Version 1.0*, COAR, 2011, http://www.coar-repositories.org/files/COAR_Interoperability_Briefing.pdf;
5. Van der Graaf, Maurits; Vernooy-Gerritsen, Marjan (editor), *The European Repository Landscape 2008: Inventory of Digital Repositories for Research Output*, Amsterdam, Amsterdam University press, 2009, DOI: 10.5117/9789089641908 - E-ISBN: 9789089641908;
6. ERA Expert Group 7 - EG 7: Rationales for ERA, *Developing World-class Research Infrastructures for the European Research Area (ERA)*, Luxembourg, Office for Official Publications of the European Communities, 2008, DOI 10.2777/96979, ISBN 978-92-79-08312-9;
   European Commission, *Work Programme 2012 - FP7 - Capacities: Part 1: Research infrastructures*, European Commission, 2011, European Commission, 2011, http://ec.europa.eu/research/infrastructures/pdf/wp2012_research_infrastructures.pdf#view=fit&pagemode=none
7. Jeffery, Keith; Asserson, Anne, *Institutional Repositories and Current Research Information Systems*, New Review of Information Networking, 14, n. 2 (2009), p. 71-83, doi:10.1080/13614570903359357 – OAI Item Identifier: oai:epubs.cclrc.ac.uk:work/ 51773;
8. White, Wendy, *Institutional repositories: contributing to institutional knowledge management and the global research commons*, In *4th International Open Repositories Conference, Atlanta, Georgia* 18th - *21st May, 2009*, http://www.mendeley.com/research/institutional-repositories-contributing-to-institutional-knowledge-management-and-the-global-research-commons/;
   Vernooy-Gerritsen, Marjan; Pronk, Gera. Van der Graaf, Maurits, *Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe*, Ariadne, n. 59 (April 2009), http://www.ariadne.ac.uk/issue59/vernooy-gerritsen-et-al/;
9. Okubo, Yoshiko, *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples*, in *OECD Science, Technology and Industry Working Papers*, 1997/1, Paris, OECD Publishing, 1997, http://dx.doi.org/10.1787/208277770603;

10. Swan, Alma, *Sharing knowledge: open access and preservation in Europe: conclusions of a strategic workshop - Brussels, 25-26 November 2010 - Report*, Luxembourg, Publications Office of the European Union, 2011, doi: 10.2777/63410 - ISBN 978-92-79-20449-4;
11. International organization for standardization (ISO), *Space data and information transfer systems. Open archival information system: Reference model. Standard ISO 14721:2003,*Geneva, ISO, 2003;
12. *Giarretta, David; Harmsen, Henk; Keitel, Christian, Memorandum of Understanding to create* a *European Framework for Audit and Certification of Digital Repositories*, http://trusteddigitalrepository.eu/Site/Memorandum%20of%20Understanding.html.
13. Mauro, Guerrini; Capaccioni, Andrea (a cura di), *Gli archivi istituzionali: Open access, valutazione della ricerca e diritto d'autore*, Milano, Editrice Bibliografica, 2010, p. 33-60, ISBN 9788870756920, http://hdl.handle.net/10760/15609;
Park, Jung-Ran, *Metadata Quality in digital repositories: a survey of the current state of the art*, Cataloging & Classification Quarterly, 47, n. 3-4 (April 2009), p. 213 – 228, DOI: 10.1080/01639370902737240;
Guy, Marieke; **Powell, Andy; Day, Michael,** *Improving the Quality of Metadata in Eprint Archives*, Ariadne, n. 38 (2004), http://www.ariadne.ac.uk/issue38/guy/;
14. DINI - Deutsche Initiative für Netzwerkinformation. Working Group Electronic Publishing, *DINI Certificate Document and Publication Services - 2010: version 3.0*, march 2011, http://nbn-resolving.de/urn:nbn:de:kobv:11-100182800;
15. Bijsterbosch, Magchiel; Brétel, Foudil; Natasa, Bulatovic Dale Peters; Vanderfeesten, M., Wallace, Julia, *PEER. D3.1 Guidelines for publishers and repository managers on deposit, assisted deposit and self-archiving*, 2009;
http://www.peerproject.eu/fileadmin/media/reports/D3_1_Guidelines_v8.3_20090528.Final.pdf
16. Knowledge Exchange, *Guidelines for the aggregation and exchange of usage data*,
http://wiki.surffoundation.nl/display/standards/KE+Usage+Statistics+Guidelines#KEUsageStatisticsGuidelines-GuidelinesfortheaggregationandexchangeofUsageData;
17. Jeffery, Keith; Lopatenko, Andrei; Asserson, Anne, *Comparative Study of Metadata for Scientific Information: the place of CERIF in CRISs and Scientific Repositories*, 2002, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.5689;
18. Gartner, Richard, *Intermediary schemas for complex XML applications: an example from research information management*, Journal of Digital Information, 12, n. 3 (2011), http://journals.tdl.org/jodi/article/view/2069/2086;
19. EuroCRIS – The European Organization for International Research Information, *CERIF 2008 – 1.2 Semantics*, EuroCRIS, November 2010;
http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release_1.2/CERIF2008_1.2_Semantics.pdf;
20. Rumsey, Sally; Shipsey, Frances; Fraser, Michael; Noble, Howard; Bide, Mark; Look, Hugh; Kahn, Deborah, *Scoping Study on Repository Version Identification (RIVER) - Final Report*, 2006,
http://www.jisc.ac.uk/uploaded_documents/RIVER%20Final%20Report.pdf;
21. Menzies, Kathleen; Birrell, Duncan; Dunsire, Gordon, *New Evidence on the Interoperability of Information Systems within UK Universities*, in Lecture Notes in Computer Science, 6273 (2010), p. 104-115, DOI: 10.1007/978-3-642-15464-5_12;
22. Digital Archiving Consultancy, *Towards a European e-Infrastructure for e-Science Digital Repositories: a report for European Commission*, e-SciDR, 2008, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/e-scidr.pdf;
23. EuroHORCs and the European Science Foundation, *Vision on a globally competitive European Research Area and road map for actions to help build it*, EUROHORCs, 2008;
24. Sutton, Caroline*, Sharing knowledge: EC-funded projects on scientific information in the digital age: Conclusions of a strategic workshop - Brussels, 14-15 February 2011 - Report*, Luxembourg, Publications Office of the European Union, 2011, doi:10.2777/63780 - ISBN 978-92-79-20451-7.

Endnotes

[1] Organisation for Economic Co-operation and Development (OECD), ***Frascati manual 2002: proposed standard practice for surveys on research and experimental development: the measurement of scientific and technological activities*****,** Paris, OECD - Organisation for economic co-operation and development, 2002.

[2] Research Excellence Framework 2014 (REF 2014), http://www.hefce.ac.uk/research/ref/; IRRA - Institutional Repositories and Research Assessment, http://irra.eprints.org/about.html; Leslie Carr, John MacColl, *IRRA (Institutional Repositories and Research Assessment): RAE Software for Institutional Repositories*, IRRA, 2006, http://irra.eprints.org/white/; *Open Access to research outputs: final report to RCUK*, LISU and SQW consulting, 2008, http://www.rcuk.ac.uk/documents/news/oareport.pdf; *Open Access to research outputs: annexes: final report to RCUK*, LISU and SQW consulting, 2008, http://www.rcuk.ac.uk/documents/news/oaannex.pdf.

[3] NARCIS - National Academic Research and Collaborations Information System, http://www.narcis.nl/; Elly Dijk, *NARCIS: linking CRISs and OARs in the Netherlands: A matter of standards and identifiers,* in *EuroCris Workshop on CRIS, CERIF and Institutional Repositories, CNR, Rome, 10-11 May 2010*, http://depot.knaw.nl/6365/.

[4] Confederation of Open Access Repositories (COAR). Working Group 2: Repository Interoperability, *The Case for Interoperability for Open Access Repositories. Version 1.0*, COAR, 2011, http://www.coar-repositories.org/files/COAR_Interoperability_Briefing.pdf.

[5] Maurits van der Graaf; Marjan Vernooy-Gerritsen (editor), *The European Repository Landscape 2008: Inventory of Digital Repositories for Research Output*, Amsterdam, Amsterdam University press, 2009, DOI 10.5117/9789089641908.

[6] *Ivi*, p. 19-21.

[7] ERA Expert Group 7 - EG 7: Rationales for ERA, *Developing World-class Research Infrastructures for the European Research Area (ERA)*, Luxembourg, Office for Official Publications of the European Communities, 2008, DOI 10.2777/96979; European Commission, *Work Programme 2012 - FP7 - Capacities: Part 1: Research infrastructures*, European Commission, 2011, European Commission, 2011, http://ec.europa.eu/research/infrastructures/pdf/wp2012_research_infrastructures.pdf#view=fit&pagemode=none [6]

[8] Keith G. Jeffery, Anne Asserson, *Institutional Repositories and Current Research Information Systems*, New Review of Information Networking, 14, n. 2 (2009), p. 71-83, doi:10.1080/13614570903359357 [7].
[9] Confederation of Open Access Repositories (COAR). Working Group 2: Repository Interoperability, *The Case for Interoperability op. cit*.
[10] Wendy White, *Institutional repositories: contributing to institutional knowledge management and the global research commons*, In *4th International Open Repositories Conference, Atlanta, Georgia, 18th – 21st May, 2009* [8], http://www.mendeley.com/research/institutional-repositories-contributing-to-institutional-knowledge-management-and-the-global-research-commons/; M. Vernooy-Gerritsen, G. Pronk, M. van der Graaf, *Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe*, Ariadne, n. 59 (April 2009), www.ariadne.ac.uk/issue59/vernooy-gerritsen-et-al/ .
[11] Organisation for Economic Co-operation and Development (OECD), ***Frascati manual 2002, op. cit.*;** Yoshiko Okubo, *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples*, in *OECD Science, Technology and Industry Working Papers*, Paris, OECD Publishing, 1997, doi: 10.1787/208277770603; Maurits van der Graaf; Marjan Vernooy-Gerritsen (editor), *The European Repository Landscape 2008: Inventory of Digital Repositories for Research Output*, *Op. cit.*, p. 100-110.
[12] ERA Expert Group 7 - EG 7: Rationales for ERA, *Developing World-class Research Infrastructures for the European Research Area (ERA)*, *OP. cit.*; European Commission, *Work Programme 2012 - FP7 - Capacities: Part 1: Research infrastructures*, *Op. cit.*
[13] A Current Research Information System (CRIS) records the R&D (Research and Development) activity either funded by or carried out by an organization, or within a thematic or subject area. Typically it covers projects, people (expertise), organizational structure, R&D outputs (products, patents, publications), R&D events and R&D facilities and equipment.
[14] Alma Swan, *Sharing knowledge: open access and preservation in Europe: Conclusions of a strategic workshop - Brussels, 25-26 November 2010 - Report*, Luxembourg, Publications Office of the European Union, 2011, doi: 10.2777/63410.
[15] Institutional Self-Archiving Mandate – Definition - ROARMAP (Registry of Open Access Repository Material Archiving Policies), http://roar.eprints.org/.
[16] S. Harnad, Open Access research, JeDEM 3 (1): 33-41, 2011
[17] Most of the Italian Phd Thesis indexed in Opengrey are not published, yet. Moreover, BNI (National Italian Bibliography) currently reports and describes all Italian Phd Thesis, also not published: in fact this document type is subjected to legal deposit at the National Library of Florence (in accordance with DPR 30.10.1997, n. 387, art. 4)
[18] International organization for standardization (ISO), *Space data and information transfer systems. Open archival information system: Reference model. Standard ISO 14721:2003,*Geneva, ISO, 2003.
[19]*David Giarretta, Henk Harmsen, Christian Keitel*, *Memorandum of Understanding to create* a *European Framework for Audit and Certification of Digital Repositories*, http://trusteddigitalrepository.eu/Site/Memorandum%20of%20Understanding.html.
[20] Mauro Guerrini, *Gli archivi istituzionali: Open access, valutazione della ricerca e diritto d'autore*, Milano, Editrice Bibliografica, 2010, p. 33-60; Jung-Ran Park, *Metadata Quality in Digital Repositories: A Survey of the Current State of the Art*, Cataloging & Classification Quarterly, 47, n. 3-4 (April 2009), p. 213 – 228; **Marieke Guy, Andy Powell, Michael Day,** *Improving the Quality of Metadata in Eprint Archives*, Ariadne, n. 38 (2004), http://www.ariadne.ac.uk/issue38/guy/.
[21] DINI Working Group Electronic Publishing, *DINI Certificate Document and Publication Services - 2010: version 3.0*, march 2011, http://nbn-resolving.de/urn:nbn:de:kobv:11-100182800.
A certificate that describes the technical, organizational, and legal aspects (including interoperability) that should be considered in setting up a scholarly repository service.
[22] Magchiel Bijsterbosch, Foudil Brétel, Natasa Bulatovic Dale Peters, Maurice Vanderfeesten, Julia Wallace, *PEER. D3.1 Guidelines for publishers and repository managers on deposit, assisted deposit and self-archiving*, 2009, http://www.peerproject.eu/fileadmin/media/reports/D3_1_Guidelines_v8.3_20090528.Final.pdf.
[23] OAI-PMH protocol limits interoperability to the unqualified Dublin Core schema, thus "flattening" research evaluation or increasing noise with an oversimplified metadata management process. Keith G. Jeffery, Anne Asserson, *Institutional Repositories and Current Research Information Systems*, *Op. cit.*; Open Archives Initiative – Object Reuse and Exchange (OAI-ORE) – Defines standards for aggregation of compound digital objects, http://www.openarchives.org/ore/.
[24] Knowledge Exchange, *Guidelines for the aggregation and exchange of usage data*, http://wiki.surffoundation.nl/display/standards/KE+Usage+Statistics+Guidelines#KEUsageStatisticsGuidelines-GuidelinesfortheaggregationandexchangeofUsageData
[25] Keith G. Jeffery, Andrei Lopatenko , Anne Asserson, *Comparative Study of Metadata for Scientific Information: the place of CERIF in CRISs and Scientific Repositories*, 2002, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.5689.
[26] In many metadata environments, particularly in that of the digital library, the problems of complex and highly flexible generic schemas are as acute as they are in that of CERIF - Common European Research Information Format. A tension arises particularly between flexibility and interoperability: the more potential approaches to encoding are offered by a standard, the more problematic is the transfer of metadata to different information systems and its interpretation and processing by them. Despite its great power as an encoding mechanism for the complex metadata needs of research environments, the CERIF model remains relatively underused in the area of research information management. Its flexibility and fragmented architecture in particular can produce significant problems for implementers and reduce its interoperability unless such key components as its semantic infrastructure are standardized between institutions. These problems were experienced by developer communities of such standards and were solved by some by using the architectural mapping features of SGML/XML. Without this facility in XML, the solution advocated here can replicate its best features but also add more powerful, non-syntactic features, such as semantic control.
The strategy has been tested thoroughly in several live research information management environments and found to be generally workable: the only problems experienced have proved to be those inherent in the metadata scheme on which the mapping to CERIF was based. The results have proved it to form a good compromise which allows the use of a key standard (with the consequent benefits of wider interoperability) in conjunction with a constrained, project-specific and more easily implemented

element set. The successful application of this methodology suggests that it may be beneficial in the wider area of digital library metadata in general, where several key metadata schemas are more easily implemented when constrained it this way.
Richard Gartner, *Intermediary schemas for complex XML applications: an example from research information management*, Journal of Digital Information, 12, n. 3 (2011), http://journals.tdl.org/jodi/article/view/2069/2086.
[27] Resource Description Framework (RDF) – A standard model for web-based data interchange, http://www.w3.org/RDF/.
[28] SKOS - Simple Knowledge Organization System is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web, http://www.w3.org/2004/02/skos/.
[29] The CERIF Semantics is one component of the CERIF 2008 – 1.2 Full Data Model (FDM). It aims at recommending a standardized formal semantics to be applied in the wider context of Current Research Information Systems (CRISs) with CERIF as the underlying data model to supply the relevant entities and their relationships. The semantic component in this version presents the current core semantics; that is, the types and roles considered relevant in a research context between the involved core entities. Compared to its preceding version, this release provides a major upgrade with respect to the quantity of relevant terms. EuroCRIS – The European Organization for International Research Information, *CERIF 2008 – 1.2 Semantics*, EuroCRIS, 2010.
[30] EuroCRIS – The European Organization for International Research Information, *CERIF 2008 – 1.2 Semantics*, *Op.cit.*
[31] VIAF – Virtual International Authority File, http://viaf.org/.
[32] Version Identification Framework Project, http://www2.lse.ac.uk/library/vif/index.html; VERSIONS (Versions of eprints. A user requirements study and investigation of the need for standards), http://www2.lse.ac.uk/library/versions/; The RIVER Scoping Study on Repository Version Identification - Sally Rumsey, Frances Shipsey, Michael Fraser, Howard Noble, Mark Bide, Hugh Look, Deborah Kahn, *Scoping Study on Repository Version Identification (RIVER) - Final Report*, 2006, http://www.jisc.ac.uk/uploaded_documents/RIVER%20Final%20Report.pdf.
[33] DataCite, http://www.datacite.org/.
[34] ORCID – Open Researcher and Contributor ID, http://orcid.org/.
[35] PersID – Project aimed at building a persistent identifier metaresolver infrastructure for digital publications and electronic resources, http://www.persid.org/.
[36] Kathleen Menzies, Duncan Birrell and Gordon Dunsire, *New Evidence on the Interoperability of Information Systems within UK Universities*, Lecture Notes in Computer Science, 6273 (2010), p. 104-115, DOI: 10.1007/978-3-642-15464-5_12.
[37] Keith G. Jeffery, Anne Asserson, *Institutional Repositories and Current Research Information Systems*, *Op. cit.*
[38] *Ibidem*.
An architecture for providing a complete research information environment at an institution is presented. The linking together, at an institution, of a "OA repository of articles (that is a repository of publications deposited institutionally for toll-free open access in parallel with a peer-reviewed publication), a CRIS (to provide contextual information), and an OA repository of research datasets and software provides that institution with an information resource suitable for all the end-users and roles. Furthermore, the formalized structure of the CRIS allows a reliable workflow to be engineered which, in turn, encourages deposit of research outputs by reducing the effort threshold by using intelligent prompts or suggestions based on the information already stored and any constraints on permissible values of attributes. However the requirements of the end-user extend beyond the individual research institution or funding organization. The institutional CERIF-CRIS system can be linked to others because they have a formal structure and, hence, can be interoperated reliably and in a scalable way. This, in turn, provides a network of access to institutional OA repositories or e-research repositories linked to each institutional CRIS via the CERIF-CRIS gateways, enhancing and controlling the access using the CERIF-CRIS information as formalized, structured, and contextual metadata which is more detailed than DC and suitable for intelligent (machine-understandable) interoperation.
[39] Digital Archiving Consultancy, *Towards a European e-Infrastructure for e-Science Digital Repositories: a report for European Commission*, 2008, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/e-scidr.pdf.
[40] Digital Library Federation (DLF), http://www.diglib.org/; DL.org Community – Digital Library Interoperability, Best Practices and Modelling Foundations, http://www.dlorg.eu/.
[41] EuroCRIS – The European Organization for International Research Information, http://www.eurocris.org/.
[42] COAR – Confederation of Open Access Repositories, http://coar-repositories.org.
[43] Knowledge Exchange is a co-operative effort that supports the use and development of Information and Communications Technologies (ICT) infrastructure for higher education and research.
[44] EUROHORCs, (European Heads of Research Councils), http://www.eurohorcs.org/E/Pages/home.aspx; EuroHORCs and the European Science Foundation, *Vision on a globally competitive European Research Area and road map for actions to help build it*, EUROHORCs, 2008;
[45] Caroline Sutton, *Sharing knowledge: EC-funded projects on scientific information in the digital age: Conclusions of a strategic workshop - Brussels, 14-15 February 2011 - Report*, Luxembourg, Publications Office of the European Union, 2011, doi:10.2777/63780.

Linking full-text grey literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012

Dominic J. Farace and Jerry Frantzen, GreyNet International, Netherlands;
Christiane Stock, INIST-CNRS, France
Laurents Sesink, DANS, Data Archiving and Networked Services, Netherlands
Debbie Rabina, Pratt Institute, School of Information and Library Science, United States

Abstract
This project seeks to circumvent the data vs. documents camp in the grey literature community by way of a middle ground provided through enhanced publications. Enhanced publications allow for a fuller understanding of the process in which data and information are used and applied in the generation of knowledge. The enhanced publication of grey literature precludes the idea of a random selection of data and information, and instead focuses on the human intervention in data-rich environments. The definition of an enhanced publication is borrowed from the DRIVER-II project, "a publication that is enhanced with three categories of information: research data, extra materials, and post-publication data". Enhanced publications combine textual resources i.e. documents intended to be read by human beings, which contain an interpretation or analysis of primary data. Enhanced publications inherently contribute to the review process of grey literature as well as the replication of research and improved visibility of research results in the scholarly communication chain.

Introduction

This enhanced publications project seeks to circumvent the data vs. documents camp in the grey literature community by way of a middle ground. Enhanced publications allow for a fuller understanding of the process in which data and information are used and applied in the generation of knowledge. The enhanced publication of grey literature precludes the idea of a random selection of data and information, and instead focuses on the human intervention in data-rich environments. The definition of an enhanced publication is borrowed from the DRIVER-II project, "a publication that is enhanced with three categories of information: 1) research data, 2) extra materials, and 3) post-publication data".[1] Enhanced publications combine textual resources i.e. documents intended to be read by human beings, containing an interpretation or analysis of primary data. Enhanced publications inherently contribute to the review process of grey literature as well as the replication of research and improved visibility of research results in the scholarly communication chain.

The goal of this project is threefold: to enhance GreyNet's existing collection of conference preprints by adding corresponding links to research data, to include commentaries i.e. post-publication data on GreyNet's existing conference preprints in metadata records, and to establish a workflow for future GreyNet enhanced publications based on the results of this project, where permanent access to full-texts and their data enriched components are made available via persistent identifiers accessible in trusted repositories.

Each of the four partnering organizations is brought together based on their expertise and tasks they will execute during the course of the project. GreyNet works together with INIST-CNRS to devise a questionnaire and carry-out a survey among its author base in the acquisition of research data linked to conference preprints. GreyNet will facilitate data entry in the DANS Easy Repository with link backs to corresponding metadata records in the OpenGrey Repository. And, GreyNet cooperates with Pratt Institute to establish basic criteria upon which commentaries by LIS students will be compiled and added to existing metadata records.

Not all grey literature is based on research data, and this holds for GreyNet's collection of conference preprints. While it is anticipated that GreyNet's contributing authors will be inclined to make their research data available, some data from previous years will not be retrievable. On the other hand, since student commentaries are related to academic credit, this portion of the project will harvest optimal results. And, the combined results of the project will contribute to a future workflow not only for GreyNet's enhanced publications but also for other grey literature communities.

A research project in progress

The international grey literature community is both diverse in scale and foci of research. The HEP (High Energy Physics) community[2] depicted at CERN in Geneva and the Karst communities such as the one at Florida State University[3] are two such examples. In fact, GreyNet's own community[4] of researchers in the field of library and information science is yet another example – one in which this study is focused.

This project carries with it the connotation of repurposing grey literature in that it not only has a retrospective element but seeks also to enrich already existing records in the collection. Through the years, GreyNet has been involved in providing retrospective access to its collections of conference preprints, which first started with bibliographic records in the SIGLE database, later their full-text records in the OpenSIGLE repository, and now in the newly launched OpenGrey Repository. The notion of enhancing these metadata records with related research data as well as post-publication data such as commentaries explains what is meant by 'repurposing grey literature'.

The Enhanced Publications Project (EPP) is being carried out in six phases, some of which overlap depending on the partnering organizations responsible for their implementation. The initial three phases comprise the first part of the project carried out in 2011 and the final three phases account for the second part of the project that will be carried out in 2012:

1. Project Proposal and Formation of the Team
2. Design of the Questionnaire and Author Survey
3. Criteria for commentaries and the selection of eligible conference preprints
4. Acquisition and Submission of research data
5. Data upload and cross-linking
6. Draft of enhanced publication guidelines and the design of a future workflow

Phase 1: Project Proposal and Formation of the Team

The project proposal was sparked by a paper written by Carroll el al[5] delivered at the Twelfth International Conference on Grey Literature entitled Scientific Data: Increasing Transparency and Reducing the Grey. In order to carry out the project, GreyNet would need to team-up with other partners and to this end, three organizations were contacted. INIST-CNRS who is the data provider for the OpenGrey Repository[6] in which GreyNet's publications are housed. The DANS Easy Repository[7] that would house the research data with crosslinks to the corresponding metadata records in the OpenGrey Repository. And, Pratt Institute's School of Information and Library Science[8] that would involve LIS graduate students in writing-up the commentaries of existing full-text papers in GreyNet's collection.

Phase 2: Design of the Questionnaire and Author Survey

The population of the survey was selected from among the 286 authors and co-authors in the International Conference Series on Grey Literature. It was decided that only first authors would receive the questionnaire, which narrowed the potential population of the survey to 162 authors of which only 95 were actually sent the online questionnaire. The reason the other 67 first authors were not included in the final survey population was due to a number of factors such as no current email address, retired, deceased, etc.

Number of (co)authors in the GL-Series	Number of first authors in the GL-Series	Number of EPP Survey Recipients	Number of EPP Survey Respondents
286	**162**	**95**	**50**

The 95 authors were sent a personalized email with a standardized text inviting them to participate in the survey by completing the online questionnaire. The survey was carried out using the freeware 'Survey Monkey' and the questionnaire contained 10 items, three of which were open-ended. Subheadings were also inserted in the questionnaire set off by quotation marks. These subheadings preceded each odd numbered question and were deemed relevant in achieving informed responses. The final results are based on the response of 50 of the 95 survey recipients, which amounts to roughly a 53% response rate.

Survey Results

While maintaining the anonymity of the individual respondent, it was possible to determine their geographic region (see Table 1, below). This was based on the response to item ten on the questionnaire in which the respondent was asked to enter his/her name and email address.

Table 1: Geographic Region of Respondents

	Percentage	Number
Asia	14,3%	6
Europe	45,3%	19
North America	33,3%	14
Non-Applicable	7,1%	3
	100,0%	42

While only 42 of the 50 survey respondents completed this item on the questionnaire, we were able to determine that authors from some 14 countries responded to the survey, and that the difference in the number of male and female respondents is negligible.

Table 2: Affiliation to the GL-Conference Series

	Percentage	Number
< 1 Year	42,9%	18
2 - 4 Years	16,7%	7
5 > Years	33,3%	14
Non-Applicable	7,1%	3
	100%	42

The respondents to item ten could also be classified into three groups based on their prior affiliation to the GL-Conference Series (see Table 2, above). Results show that about 43% were involved over the past year, about 17% were involved in the conference series 2 to 4 years ago, and some 33% of the respondents were involved as authors in the conference series 5 or more years ago. It was quite interesting to find that twice as many authors who contributed 5 or more years ago to the conference series chose to respond to the survey than authors distanced 2-4 years from the conference series.

"Data exchange is becoming the norm in open access communities"

Question 1: Does one or more of your conference papers in the GL-Series base its findings on empirical or statistical data?

	Percentage	Number
Yes	60%	30
No	40%	20
	100%	50

While it is generally known that not all of the conference papers in the GL-Series are based on statistical data, it still has yet to be determined what percentage of the collection is. However, the fact that 60% of the respondents' state that their work relies on empirical research data provides a clear indication of the relevance that this could have for the project.

Question 2: If so, would these data and/or datasets still be available in part or whole for archiving purposes?

	Percentage	Number
Yes	54,%	20
No	45,9%	17
	100%	37

Some 54% of the respondents to this question indicate that they still have research data available from their former studies. This is no doubt indicative of the value they place on their work.

"A data policy should be in place within research communities and organizations"

Question 3: Are you aware of any existing data archives or data initiatives in your country related to grey literature or other scientific publications?

	Percentage	Number
Yes	43,5%	20
No	56,5%	26
	100%	46

The response to this question shows that over 56% of the respondents are unaware of data initiatives related to grey literature in their own country. Unfortunately, we do not yet have cross tabulations by country. The incentive to further explore this finding takes on the implication that our project could contribute to an increased awareness of such initiatives within the grey literature community. It is also safe to assume that many if not most repositories are insufficiently robust to house and store statistical and other research data - this being the case with the OpenGrey repository. And, it is for this reason that GreyNet sought to partner with DANS in carrying out that phase of the project.

Question 4: If so, please provide the name(s) and corresponding URL(s) here?

	Percentage	Number
Specific	72,2%	13
General	11,1%	2
Non-Applicable	16,7%	3
	100%	18

Responses to this open-ended question were categorized into one of three clusters depending on how specific, general, or non-applicable the responses were. It can be noted that the number of respondents to this question was significantly less than to other items on the questionnaire. Nevertheless, if we look at the number of those who did provide specific names and/or URLs, we are able to compile a short list of archives housing research data on grey literature as show below.

ADP (SI), DANS (NL), IATUL (INT), ICPSR (US), IOP (INT)	ISS (IT), METIS (NL), Morphbank (US), NASA (US), NIH (US)	NOAA (US), NSF (US), NSIDC (US), NUSL (CZ), ORNL (US)	PLEIADI (IT), SIDR (FR), SYNABA (PL), TRAIL/CRL (US),

Of the 19 archives/portals named above, 8 are European, 9 are North American, and 2 are listed as international.

"Data counts as science output and should be recognized in references and citations"

Question 5: Would you be willing to submit data, datasets, or subsets to DANS (Data Archiving and Networked Services) that would in turn be linked to their existing metadata records in the OpenGrey Repository?

	Percentage	Number
Yes	48,9%	22
No	6,7%	3
Uncertain	44,4%	20
	100%	45

Almost half of the respondents appear willing to submit their data to the DANS Archive, while 44% express some uncertainty. The results of this question prove challenging for our project in that it will first be necessary to address the reasons for the respondents' hesitance, before embarking on a campaign to solicit the research data. Perhaps, by underscoring the advantages authors gain through increased referencing and citing of their work, as well as by providing them with ready guidelines for data entry, these authors/researchers would be more than willing to contribute their data to the project.

Question 6: If so, would you prefer that GreyNet entered your retrospective data and/or datasets in DANS, or would you prefer to do this directly?

	Percentage	Number
GreyNet	44,7%	17
Myself	18,4%	7
No Preference	36,9%	14
	100%	38

The 'no preference' response to this question can be interpreted not only as encouraging for acquiring retrospective data but also for the acquisition of ongoing research data that would be added to the DANS Repository and cross-linked to the corresponding metadata records in OpenGrey. While the initial scope of this project is geared to retrospective input, it is to be understood that empirical and statistical data underlying future conference preprints will be directly entered in the DANS Repository by authors themselves.

"Research data should be preserved and accessible in order to enhance scholarly communication"

Question 7: Do you agree that both the data producer as well as the data user stand to benefit by submitting data, datasets, or subsets for this Enhanced Publications Project?

	Percentage	Number
Yes	84,4%	38
No	0%	0
Uncertain	15,6%	7
	100%	45

Over 84% of the respondents agree that both the researcher/author as well as the potential data user would stand to gain from the enhancement of conference preprints with related data, datasets, and/or subsets. While it is long understood that researchers and authors are at the same time information users, and while it has been demonstrated that specific communities of researchers are more likely to first use the

sources/resources produced within their own community before searching beyond[9], we are now seeing a new development in which wider audiences (i.e. net users) now have open access to underlying research data, whereby this type of grey literature becomes more transparent and accessible worldwide.

Question 8: Do you think that guidelines for data entry should be available for future conference papers and other types of grey literature?

	Percentage	Number
Yes	91,1%	41
No	0%	0
Uncertain	8,9%	4
	100%	45

Over 90% of the respondents to this question are of the opinion that a set of guidelines should be made available for data entry. And, such guidelines will be addressed in the final phase of this Enhanced Publications Project.

"Data is disciplinary or subject based and this accounts for differences in formats used to acquire it"

Question 9: What kind of data and data formats have you used/are using in your research?

	Percentage	Number
Specific	44%	15
General	38%	13
Non-Applicable	18%	6
	100%	34

Responses to this open-ended question were categorized into one of three clusters depending on how specific, general, or non-applicable the responses were. If for example a respondent replied that they were 'no longer engaged in research' or 'not engaged in research at the moment', then such responses were categorized as non-applicable. If the respondents replied to this question in broad terms such as tables, charts, times series, etc., these responses were categorized as general. And, those respondents who actually named particular software or statistical packages such as SPSS, Mekov, Excel, Minitab, MS Access, etc. were identified as having provided a specific response.

Question 10: Please enter your name, email address, and any other comments or recommendations for this Enhanced Publications Project?

As indicated earlier in the paper, the response to this open-ended question allowed us to specify the geographic region of the author as well as his/her affiliation over time in the GL-Conference Series. Comments by four of the survey respondents are of particular interest in that they are assumed to be shared by other authors/researchers. These comments are recorded as follows:

> **"**I'm a firm believer that not all data is worth archiving.**"**
> **"**Will your system support the preservation and migration to new platforms?**"**
> **"**For many, it would need to be a local activity linked to our own sites.**"**
> **"**I share data with my colleagues and research teams, but I'm not sure if I would be willing to share them with anybody else at the moment?**"**

It is believed that the tenor of these comments have much to do with the uncertainty expressed in response to Question 5, and must likewise be kept in mind during the acquisition phase of this project.

Phase 3: Criteria for commentaries and the selection of eligible conference preprints
Graduate students from Pratt Institute's School of Information and Library Science were engaged in the selection of GreyNet's conference preprints currently accessible via the OpenGrey Repository. In their initial selection and review of the full-texts, they first determined whether the content would be of value in the research chain and whether the manuscripts were clearly written given the fact that English is not always the first language of the authors in the International Conference Series on Grey Literature.

Standardized Format
In order to facilitate their work, the students developed a standardized format used in drafting the commentaries. Each commentary comprises a brief summary, the strengths of the research, any noticeable limitations, and the takeaway *i.e.* what the student considered the most salient aspect in the study. A sample commentary is provided below.

http://hdl.handle.net/10068/697760 GL8_Anderson_et_al_2007_Commentary_by_Pratt_Institute.pdf

Harnessing NASA Goddard's Grey Literature: The Power of a Repository Framework

Summary:
In an organization like NASA, where researchers come and go depending on the project, it is especially important to have a central repository for the valuable scientific grey literature that is produced during the lifetime of any given mission. This paper details the steps that went into NASA Goddard Library's development of a Digital Asset System (DAS) capable of managing the wide range of items (multimedia objects as well as text documents) and multiple vocabularies used in its various projects over its long history. The authors developed their own extension of a Dublin Core Metadata scheme, and also created oral and video histories of many projects that had concluded.

Strengths:
- The description of the process the authors used to create and improve upon the DAS is clear and thorough.

Limitations:
- The paper might benefit from more in the way of concrete examples.

Takeaway:
This is a fascinating look at the development of a repository for a fabled government agency, the problems that arose along the way, and the solutions the authors devised. It is a valuable working document for anyone undertaking the creation of a digital repository for grey literature.

Reviewer Eloise Flood, Pratt SILS 2011

Currently 205 commentaries on GreyNet's conference preprints now are available in the OpenGrey Repository. This covers 79% of the total collection. More on this phase of the project was presented by a group of Pratt students during the GL13 Conference Poster Session.[10]

Project Continuation
The final three phases of this project will be undertaken in 2012 and involve the acquisition of research data and their subsequent upload and cross-linking between the DANS Repository and the OpenGrey Repository. This will allow for open access to research data linked to underlying full-texts, to extra materials such as PowerPoints, abstracts, and biographical notes already available via the metadata records, as well as the post-publication commentaries. A set of guidelines will be drafted and used in the future workflow of GreyNet's enhanced publications. In this last phase of the project, existing guidelines for the re-use, verification, and preservation of data[11] will be employed. Furthermore, in line with GreyNet's policy on information access and retrieval, neither the submission of the research data nor the format will be mandated.

In Close

Final conclusions will have to await completion of the enhanced publications project. However, it is worthwhile to summarize the results of the first half pertaining to the survey and post-publication data. While it was known from the start of the project that not all of the conference preprints in GreyNet's Collection were based on research data, it was encouraging that 50 of the 95 authors surveyed completed the questionnaire, 30 of whom stated that their findings were based on empirical or statistical data, and 20 of whom still maintain these data, datasets, and/or subsets, which are still available for archiving purposes. While it was surprising to find that a little over half the respondents were unaware of existing data archives related to grey literature in their own country, it was considered worthwhile to have accumulated a short list of subject-based grey literature repositories that do house research data. This list can now be further expanded and even linked to other subject-based resources.[12]

In the second part of the questionnaire, which focused not only on retrospective data but also ongoing and future research, we found that very few respondents were unwilling to submit their data for input in the DANS Repository. However, on the other hand, nearly half voiced uncertainty. In the next phase of our project, it will be important to address the uncertainties before proceeding with a campaign for data acquisition. Finally, the overwhelming majority of respondents agree that both the data producer and user would benefit from enhanced publications, and they look forward to guidelines for future data submission.

In regard to post-publication data, it was anticipated at the start of the project that the LIS students would produce a significant number of commentaries given this phase of the project was linked to course credit. However, it was beyond expectation that almost 80% of the existing conference preprints would have been completed in the first part of our enhanced publications project. This wealth of post-publication data now provides the opportunity to assess the impact that these commentaries have on GreyNet's collection of conference preprints.

Acknowledgements

Special thanks to the students at Pratt Institute's School of Information and Library Science, who under the direction of Prof. Dr. Debbie Rabina carried out the post-publication phase of this project by enhancing GreyNet's Collection of conference preprints with commentaries. Likewise, a word of thanks to Nathalie Henrot at INIST-CNRS, who was responsible for integrating over 200 commentaries in the OpenGrey Repository.

References

[1] http://www.driver-repository.eu Enhanced Publications: State-of-the-Art (PDF), Enhanced Publications: Object Models and Functionalities (PDF)

[2] The driving and evolving role of grey literature in High-Energy Physics / Anne Gentil-Beccot http://www.reference-global.com/doi/abs/10.1515/9783598441493.2.155

[3] Grey Literature in Karst Research: The Evolution of the Karst Information Portal, KIP / Todd Chavez http://www.reference-global.com/doi/abs/10.1515/9783598441493.2.181

[4] Grey Literature Network Service http://www.greynet.org

[5] Scientific Data: Increasing Transparency and Reducing the Grey / Bonnie Carroll, June Crowe, and J.R. Candlish http://hdl.handle.net/10068/700004

[6] http://www.opengrey.eu/ OpenGrey Repository – System for Information on Grey Literature in Europe

[7] https://easy.dans.knaw.nl/ui/home DANS Easy - Data Archive and Networked Services

[8] http://www.pratt.edu/academics/information_and_library_sciences/ Pratt Institute; School of Information and Library Science, SILS

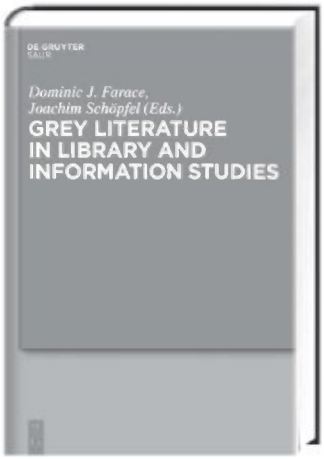
[9] See references 2 and 3 above.

[10] http://00215f8.netsolhost.com/images/GL13-PSS,_Edwards_et_al.pdf Shining a light on grey literature / Bethany Edwards, Eloise Flood, Thomas Keenan, and Ashley Rode

[11] http://www.dans.knaw.nl/en/content/categorieen/publicaties/dans-studies-digital-archiving-6 Selection of Research Data: Guidelines for appraising and selecting research data Heiko Tjalsma (DANS) and Jeroen Rombouts (3TU.Datacentrum), 2011.

[12] http://www.greynet.org/greysourceindex.html GreySource Index, A selection of classified web-based resources in grey literature

DE GRUYTER
SAUR



Dominic J. Farace and Joachim Schöpfel (Eds.)

GREY LITERATURE IN LIBRARY AND INFORMATION STUDIES

2010. vi, 282 pages
Hardcover RRP € 89.95 [D] / US$ 126.00. ISBN 978-3-598-11793-0
eBook RRP € 89.95 / US$ 126.00. ISBN 978-3-598-44149-3

The further rise of electronic publishing has come to change the scale and diversity of grey literature facing librarians and other information practitioners. This compiled work brings together research and authorship over the past decade dealing with both the supply and demand sides of grey literature. While this book is written with students and instructors of Colleges and Schools of Library and Information Science in mind, it likewise serves as a reader for information professionals working in any and all like knowledge-based communities.

CONTENTS

DE GRUYTER
SAUR

eBooks are currently only available for Libraries / Institutions
Prices in US$ apply to orders placed in North America only
Prices are subject to change. Prices do not include postage and handling

www.degruyter.com

Dependency on Regional Libraries for Grey Literature: Perceptions of Researchers in Engineering Sciences and Technology

N. Chowdappa, BMS College of Engineering, Bangalore, India
L. Usha Devi, Bangalore University, India
C.P. Ramasesh, University of Mysore, India
V.R. Shyamala, Government PU College, Mysore, India

Abstract

While highlighting the strength of Grey Literature collection in the engineering institutions recognized for research in the state of Karnataka, the study focuses on the extent of dependency of researchers and librarians on the regional libraries for grey resources. In the present study, response have been sought as to the frequency of access, discipline-wise use of grey collection, gender-wise awareness of Grey Literature available in the holdings of the regional libraries.

Further, the research study focuses on the perceptions of the faculty and research scholars as to the cooperation and assistance rendered by the library staff in getting Grey Literature. The study also covers the feed back of the researchers on notifying new arrivals of Grey Literature on the institution website, and the need for conducting orientation programmes for better access and use of Grey Literature available in the holdings of the regional libraries. The summary of findings depicts that an overwhelming majority of the researches opine that the libraries either individually or jointly have to notify new arrivals on the website and further there is felt-need for the conduct of orientation progammes.

Keywords: Grey Literature, Inter-Library Cooperation; Grey Literature, Resource Sharing; Grey Literature, User Studies.

The dependency on regional libraries is a necessity, especially in the era of information explosion. Libraries cannot be self sufficient in terms of its collection. In order to meet specific demands or needs of users, libraries need to depend on other libraries. Further, it is high time that libraries need to work in a network environment and share their information resources for mutual benefits. The present study aims at undertaking the need of regional library collection for the researchers working in the area of engineering sciences and technology. The objectives of the study are;

- To know the use pattern of GL available in the holdings of the regional libraries (RL)
- To identify the constraints in using regional libraries for GL collection
- To find out the need for digitization of grey resources and resource sharing venture.
- To analyse the training needs of researchers in order to effectively make use of facilities and services of the regional libraries.

Questionnaires were distributed among 65 librarians and 1270 researchers to obtain relevant data for the study. The total number of respondents is 1270, which account for 84.6%. Relevant opinions have been sought from the chief librarians working in the 65 research centres in engineering sciences and technology.

Table 1: Extent of Use of GL by Researchers

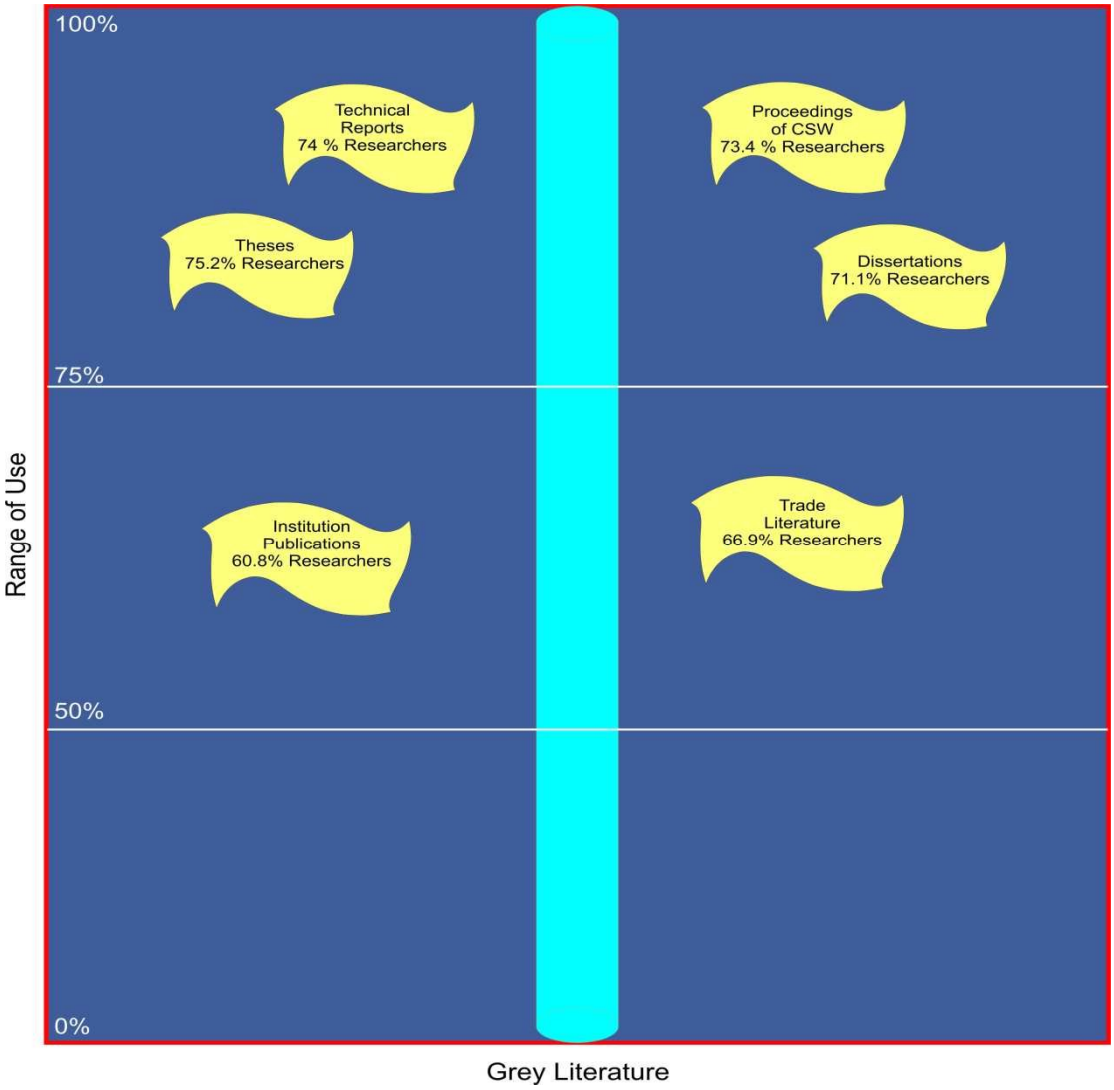


Table 1 shows the extent of access and use of various categories of GL in engineering research institutions. Technical reports and proceedings of conferences, seminars, and workshop (CSW) are being used to the highest extent of 75% to 100%. Even the sources, theses, and dissertations are also being accessed in the highest range of 75% to 100%. However, on the other hand, institution publications and trade literature are being accessed to a moderate extent of 50% to 75%. From this it is clear that GL is being used by researchers in the disciplines of engineering sciences and technology.

Table 2: Active Users of Regional Libraries

S. No.	Disciplines of Research	Percentage
01	Civil Engineering	32.5%
02	Mechanical Engineering	30.2%
03	Electrical and Electronics Engineering	39.2%
04	Electronics & Communication Engineering	35.4%
05	Computer Science and Engineering	29.4%
06	Chemical Engineering	54.2%
07	Textile and Silk Technology	21.1%
08	General Science	38.6%
09	Business Administration	55.6%
	AVERAGE NUMBER OF USERS	37.35%

Table 2 depicts dependency of active users of GL on the regional libraries. The users are shown discipline-wise and number of researchers are depicted in percentages. Among the active user of RL, chemical

engineering and business administration are comparatively more in number and the percentage of users are 54.2% and 55.6% respectively. Researchers from the discipline of textile and silk technology represent just 21.1% being the lowest among the active users of regional libraries. Above all, the average number of active users of RL covering all the disciplines stands at 37.35%. Therefore, it can be concluded that a large majority of the researchers does not depend on the RL frequently.

Table 3: Dependency of Librarians on RL for GL

S. No.	Types of GL	Extent of Dependency					Total
		100%	75%	50%	25%	Nil	
1	Theses	5 7.7%	13 20.0%	16 24.6%	21 32.3%	10 15.4%	65 100%
2	Dissertations	3 4.6%	15 23.1%	18 27.7%	22 33.8%	7 10.8%	65 100%
3	Institutional publications	3 4.6%	13 20.0%	15 23.1%	23 35.4%	11 16.9%	65 1005
4.	Trade literature	3 4.6%	11 16.9%	17 26.2%	19 29.2%	15 23.0%	65 100%
5.	Technical reports	5 7.7%	7 10.8%	24 36.9%	22 33.8%	7 10.8%	65 100%
6.	Proceedings of CSW	3 4.6%	12 18.5%	19 29.2%	23 35.4%	8 12.3%	65 100%

Table 3 shows the extent of dependency of librarians on GL of regional libraries. A large majority of librarians depend on RL for technical reports, proceedings of the CSW and theses volumes. The extent of dependency is low for trade literature and institutional publications. Librarians while extending information service and to meet the needs of their library users depend on RL for GL. The dependency is comparatively more for technical report literature and the proceedings of CSW.

Table 4: Researchers on Constraints in Using GL of Regional Libraries

S. No.	Category of Constraints	Research Supervisors	Research Scholars
01	Non up-to-date collection	21.5%	23.5%
02	Distance barrier & Library timings	52.2%	41.0%
03	Difficult to locate GL	19.1%	18.7%
04	Lack of assistance by staff	4.2%	4.6%

Table 4 projects various constraints in using regional libraries. Among the four major constraints, unsuitable library hours and distance barrier hinders the researchers in using the RL effectively. 52.2% of the research supervisors and 41.0% of the research scholars have indicated that distance and library timings are the constraints.

For some of the researchers, the main constraint is that the GL collections in the RL is not up-to-date. This segment of researchers account for 22.5%, a small percentage of researchers representing 18.9% opine that the GL in the regional libraries is scattered and difficult to trace or locate. Finally, very few research supervisors and research scholars have expressed that the library staff in the regional libraries is not cooperative. This group of researchers account for just 4.4%. From this it is clear that there is immediate felt need to extend the library hours, including the working hours on holidays. Secondly, the GL collections need to be updated in time.

Table 5: Response on the Extent of Up-to-date Collection

Category of GL	Extent of up-to-date collection					Total
	100%	75%	50%	25%	Nil	
Theses	168 13.2%	351 27.6%	452 35.6%	176 13.9%	123 9.7%	1270 100%
Dissertations	201 15.8%	346 27.2%	470 37.0%	188 14.8%	65 5.1%	1270 100%
Institutional Publications	246 19.4%	309 24.3%	334 26.3%	323 25.4%	58 4.6%	1270 100%
Trade Literature	123 9.7%	352 27.7%	379 29.8%	347 27.3%	69 5.4%	1270 100%
Technical Reports	222 17.5%	383 30.2%	421 33.1%	212 16.7%	32 2.5%	1270 100%
Proceedings of CSW.	242 19.1%	364 28.7%	450 35.4%	181 14.3%	33 2.6%	1270 100%

As regards extent of up-to-date collection of GL in regional libraries, researchers have expressed their opinion. Accordingly, institutional publications and trade literature are slightly out-dated in RL and on the other hand, theses, dissertations, technical report literature and proceedings of CSW are comparatively up-to-date in collection. Though the institutional publications are in house documents, many a time they are maintained in various departments rather than in libraries. Therefore, the need of up-dating the collection of GL is emphasized.

Table 6: Library Hours

Weekdays		Weekends	
Library Hours	No. of Libraries	Library Hours	No. of Libraries
8 hrs/ day	24	Doesn't work	36
12 hrs / day	22	4 hrs / day	14
15 hrs / day	19	8 hrs / day	10
Total	65	12 hrs / day	05

Table 6 furnishes working hours of the engineering institute libraries. 24 libraries representing 36.9% work for 8 hors during week days. On the other hand, 19 libraries representing 29.2% work for 15 hours during week days. Remaining 22 libraries work for 12 hours a day during week days.

As regard working hours during week days as well as weekends, the library hours are unsuitable for the researchers of other institutions to come and make use of the facilities. Therefore, there is felt need to keep open the libraries for longer hours both on weekdays and week ends.

Table 7: Extent of Cooperation Extended by Librarians

S. No.	Extent of Cooperation	Percentage
1	Highly cooperative	56.0%
2	Cooperative	39.5%
3	Slightly cooperative	4.0
4	Not at all cooperative	0.5%
	Total	100%

Table 7 shows the extent of cooperation and assistance rendered by the library staff for the researchers. According to the large majority of the researchers representing 56% feel that the library staff is highly cooperative. Another segment of researchers representing 39.5% opine that the library staff is cooperative and the performance of the librarians is satisfactory. Only a negligible percentage of 0.5% of the

researchers have expressed that the staff not at all cooperative. From this it is clear that an over whelming majority of the researchers are happy about the performance of the library staff and they need to be congratulated on their performance.

Table 8: Researchers on the Need of Orientation (OP) / Training

Sl. No.	Need for OP/Training	Research Supervisors		Research Scholars		Total	
		No.	%	No.	%	No.	%
1	Strongly agree	159	46.8	345	37.1	504	42.0
2	Agree	158	46.5	499	53.7	657	50.1
3	Agree to a little extent	18	5.2	79	8.5	97	6.8
4	Disagree	5	1.5	7	0.8	12	1.1

Table 8 depicts the training / orientation needs of the researchers. A large majority comprising of 92.1% opine that they need training to effectively make use of the facilities and services of the regional libraries. Orientation as to the availability and facilities extended by the regional libraries needs to be provided to the researchers who are the external members of the library. This is the basic necessity to avail facilities under resource sharing programmes.

Table 9: Research Scholars and Research Supervisors on Knowledge of GL of RL

S. No.	Response	Scholars		Supervisors	
		Male	Female	Male	Female
01	Positive	46.8%	32.3%	48.5%	35.6%
02	Negative	53.2%	67.7%	51.5%	64.4%
	Total	667 (100%)	263 (100%)	295 (100%)	45 (100%)

Table 9 shows gender-wise response of research supervisors and research scholars on the knowledge of GL collections available in the holdings of the regional libraries. Majority of the researchers are unaware of the collections of the regional libraries. Among the female research scholars 67.7% have no knowledge of the regional library collections. Among the research supervisors 64.4% do not know the collection of GL of regional libraries. Female researchers find it difficult to travel long distance to visit regional libraries where GL collections are strong. Secondly, the timings of a large majority of the regional libraries are not convenient for them. Therefore, it is recommended to keep open the libraries for longer hours and also digitize the useful GL and provide on line access. VTU has to initiate steps in this direction on top priority so as to share the resources available in the regional library holdings.

Table 10: Demand for Notifying New Arrivals of GL on the Website

Sl. No.	Category of GL	Positive Responses in Percentage		
		Research Supervisors	Research Scholars	Total
1	Theses	93.2%	93.5%	93.4%
2	Dissertations	92.1%	91.2%	91.6%
3	Inst. publications	86.5%	86.7%	86.6%
4	Trade literature	85.0 %	85.9%	85.5%
5	Technical reports	91.8%	92.8%	92.3%
6	Proceedings of CSW	93.8%	92.4%	93.1%
	Average percentage	90.4%	90.4%	90.4%

Table 10 furnishes the opinion of researchers on notifying the new arrivals of GL on the institution websites. An overwhelming majority of researchers representing 90.4% subscribes to the view that new arrivals of GL be brought to the attention. This presupposes that the researchers of the engineering institutions are highly in need of getting required current information on the web. Hence, it is high time that important information resources that are in great demand be digitized creating institutional repository and provide remote access for research.

Findings and Recommendations

- Active users of regional libraries (RL) stands at 37.35%.
- Theses, technical. reports and proceedings of CSW are used to a greater extent.
- Performance of library staff is appreciated at large.
- Large percentage of female researchers is unaware of the GL collections of the regional libraries.
- Immediate need of creating Institutional Repository is noticed.
- Lower dependency of librarians on RL is depicted.
- Orientation Programme has to be conducted for researchers on facilities and services of regional libraries.
- Notify the new arrivals of GL on the institution web sites.
- Libraries have to be kept open on weekends to facilitate research
- Steps need to be initiated to up-date GL collections in the libraries.

Bibliography

1. Farace, Dominic J (2004). Grey literature: Grey matters in the world of networked information. *Publishing Research Quarterly,* 20 (1):3-91.
2. Gelfand, J. M. (2000). Grey literature poses new challenges for research libraries. *Collection Management,* 24 (1-2):137-147.
3. Gokhale, P.A. (1999). Information system for Grey literature in India: A model. *Library Science with Slant to Documentation and Information Science,* 36 (4): 229-236.
4. Laloo, Bikika T (2002). *Information needs, information seeking behavior and users*. New Delhi : Ess Ess Publications.
5. Lee, Lafleur and Rupp, Nathan (2005). Making grey literature available through institutional repositories. *The Grey Journal*, 1 (2): 67-72.
6. Manorama, Tripathi and Jeevan, V.K.J (2007): Grey literature archiving in open universities: A model for India. *The Grey Journal*, 3 (2) : 100-106.
7. Needham, Paul A, et.al (2002). *The MAGIC Project : Managing Access to Grey literature Collection,* Final Report. London ; The British Library.
8. Ramadevi, Tella (2006). *Access and use of grey literature in social sciences :Problems and prospects.* New Delhi: Ess Ess Publications.
9. Ranger, Sara L (2005). Grey literature in special libraries : Access and use. "Work on Grey in Progress". Proceedings of the Sixth International Conference on Grey Literature, New York, Dec. 6-7, 2004: (http://opensingle.inst.fr).preprint. *Publishing Research Quarterly*, 12 (1) : 53-63.
10. Chowdappa, N (2010). Organisation and use patterns of grey literature by faculty and researchers in research centres affiliated to Visvesvsraya Technological University: A study. Mysore, University of Mysore (Doctoral Thesis).

From Unprocessed Unknowns to Digital Diffusion: Uncovering Grey Literature at the Center for the History of Psychology

Jodi Kearns and Cathy Faye
Center for the History of Psychology, The University of Akron, United States

Lynn Willis
PsycEXTRA, The American Psychological Association, United States

Abstract
The Center for the History of Psychology (CHP) at The University of Akron is home to a special collection of grey literature that documents the history of psychology and related human sciences, such as philosophy, biology, and anthropology. The collection comprises 6,000 paper-and-pencil tests; 15,000 organizational newsletters, white papers, conference programs and proceedings, and membership directories; and 8,000 US military technical reports. These data, covering a broad variety of contemporary and historical topics, are an untapped but immensely useful resource for historical research, as well as contemporary meta-analytic studies and general literature reviews. The collection grows daily through donations from psychologists and organizations, and nearly 2,000 paper-and-pencil tests, for example, have been added to the grey literature collection since the original abstract was submitted.

The Problems and Goal

The central goal of the CHP was to find a way to make a massive collection of unknown materials available to scholars and researchers from a variety of fields. Though digitization and distribution were an obvious solution, the CHP lacked personnel and financial resources necessary for such a massive project. A more traditional approach of finding grants to fund a project for a year was dismissed in favor of initiating and sustaining a relationship with a partner that had the capability to diffuse the content broadly by publishing rich "discovery tools" (Willet, 2009). The existing basic donor-repository relationship between the CHP and the American Psychological Association (APA) was developed into a working partnership where both institutions found benefit: the CHP hired a certified librarian specializing in digital projects to manage the digitization initiatives, and APA received a mass store followed by an ongoing trickle of high quality images of documents to populate research sources for professionals and scholars of psychology and related human sciences. Such a symbiotic relationship permitted the CHP librarian to uncover, process, digitize, and provide access to its largely unknown grey literature documents that may have otherwise spent another several years in unprocessed collections of boxes.

A quick scan across literature discussing collection, description, and access to grey literature shows some common themes: that there is a lot of grey literature, it's valuable research material, and it takes a lot of time to adequately describe in an institution's system of retrieval (see Aina, 2000; Sulouff et al, 2005; Ranger, 2005). The grey literature collection at the CHP is plentiful and considered by patrons and donors tremendously valuable to the historical record, and, by the librarians and archivists, daunting by way of indexing, organization, and access. Further, the staff of the CHP recognizes that the scholarly appeal of the grey literature collection may be limited to a particular audience of researchers in the field of psychology, populous though this audience may be. The CHP itself is the largest collection of its kind in the world, and the only collection of its kind in the United States. The grey literature comprises only a small portion of the collections that also include manuscripts, artifacts, photographs, books, and film. With a global, though specialized, audience requesting access to dusty documents still in the boxes sent by the donor, the staff of the CHP recognized urgency in generating quick and complete availability and dissemination of this one-of-a-kind collection for a specialized, international patronage. This urgency was addressed in the partnership between the CHP and APA.

The Collections

The CHP's grey literature collection holds national and international military technical reports; organizational newsletters and circulars; conference proceedings, programs, and announcements; advertisements of new books and lectures; and other grey literature documents, but this collection was

mostly unknown until the CHP/APA partnership. Presented with 85 boxes from the Saul Sells collection, from donor and CHP Board Member Dr. Ludy Benjamin, the Digital Projects Manager with a team of students unboxed and organized what turned out to be approximately 20,000 Air Force, Navy, and Army technical reports and studies ranging from the 1910s to the 1990s covering topics in survival, auditory and visual perception, medical experimentation, nutrition, statistical processes, and psychological testing. The Sells collection also includes newsletters, circulars, conference programs and proceedings, directories, and advertisements from assorted national and international organizations and psychological societies, special interest groups, and government offices. Documents in this grey literature collection were published from the 1880s until the present. Pulling also from processed and unprocessed collections from additional donors, the digitization team eventually queued 15,000 grey literature documents for digitization. Not every document that was uncovered qualified for digitization because the Digital Projects Manager was legally required (to the best of her knowledge and capability) to adhere to both Title 17 of the US Code and a selection policy written by APA. Documents were selected for digitization based on contracts made between APA and Content Owners regarding access through PsycNET and continued rights to the original work.

The Importance of the Collection to the Research Community

Though most contemporary psychologists rely on peer-reviewed, published literature, grey literature has tremendous potential as a useful resource for psychologists, including the more than 100,000 members of the American Psychological Association. Grey literature is being successfully used in a variety of disciplines to study a vast array of topics such as famine in Africa (von Braun, Teklu, & Webb, 1999), antibiotic use in low-income countries (Radyowijati & Haak, 2003), and the use of medical simulations in teaching (Issenberg, McGaghie, Petrusa, Gordon, & Scalese, 2005). Searches of the literature suggest that grey literature is rarely used among psychologists, though this perhaps reflects the simple fact that it is difficult to locate and obtain. Prior to the implementation of the PsycEXTRA database, there was no unified database for grey literature in the field of psychology. Scholars seeking such literature would have to search miscellaneous databases for conference proceedings, theses, and technical reports and then contact authors to obtain the text (Faragher, Cass, & Cooper, 2005).

Scholars have suggested that grey literature is particularly useful for disciplines such as psychology that make use of meta-analysis, a statistical tool for combining the results of several research studies on a given topic (Conn, Valentine, Cooper, Harris, & Rantz, 2003). Meta-analyses and general literature reviews usually involve combing the published literature for studies on a given topic, and then reviewing and/or statistically combining the results from those published studies to arrive at a tentative conclusion regarding the state of research on that particular topic. Such studies are prevalent and useful in the field of psychology and often serve as the starting point for a new research study on a given topic. However, it has been suggested that because these review studies rely primarily on published findings, they suffer from a publication bias, where studies lacking positive results are less likely to be published. This problem is thought to lead to an overestimation of the magnitude of a relationship or scientific finding (Thornton & Lee, 2000). Because grey literature represents the unpublished scientific record, it may help to ameliorate this bias and make such reviews more comprehensive and more representative of the field. The CHP's holdings, and specifically its large collection of technical reports, are a particularly useful and previously unknown source of data on everything from workplace satisfaction to the physiological and psychological effects of prolonged isolation. The grey literature collection at the CHP is also a tremendous resource for historians of psychology and related sciences. The test collections document more than 100 years of changing practices in the measurement of aptitudes, personality, abilities, and preferences. The collection of military documents details the long and sometimes complicated relationship between the military and social scientists. Conference proceedings, reports, and agendas all document the ways in which particular aspects of these sciences grew and changed over time. Access to these documents would therefore be indispensable in helping historians trace the development of the field in a variety of ways.

Document Selection

APA staff visited the CHP to triage the collection for digitization. First priority was assigned to documents published by organizations with which PsycEXTRA already had an ongoing agreement. PsycEXTRA partners with nonprofit organizations, educational and research institutions, and for-profit companies to archive, index, and distribute their grey literature to the academic community. Some organizations included in this effort are Psychonomic Society, Society for the Psychological Study of Ethnic Minority Issues, US Air Force, Midwestern Psychological Association, Society for Industrial and Organizational Psychology, Southwestern Psychological Association, Society for the Teaching of Psychology, Rocky Mountain Psychological Association, National Institute of Mental Health, Southeastern Psychological Association, Society for General Psychology, Federal Aviation Administration, US Navy, Western Psychological Association, Stockholm University, and National Institute on Drug Abuse.

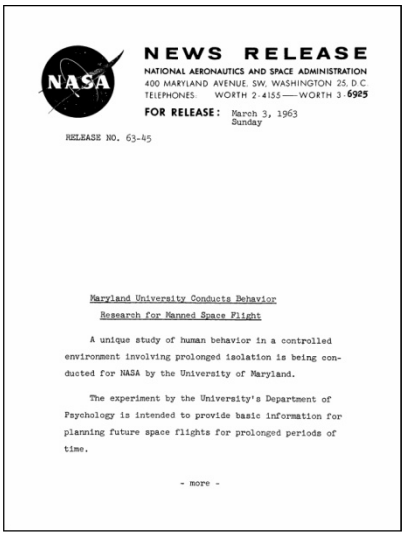
Second priority went to public domain documents that filled a gap in PsycEXTRA holdings. Broad assumptions about current holdings were used to choose material rather than matching each document against the database, and as a result, some duplication occurred.

Finally, documents published by organizations with strong ties to APA were chosen. The promise of archival documents scanned at no charge to these organizations was, in many instances, persuasive reason to partner with PsycEXTRA.

Below is shown a cross-section of organizations and societies whose collections were digitized with images from the CHP digital collections. [Aside: These highlights are going to appear to favor grey literature from the American Psychological Association, research departments of the United States military, and research laboratories of the United States government. The collection is largely composed of documents from these organizations, true, but the underrepresented organizations and societies may have granted permission under contract for us to include their grey literature documents in our digitization project, and subsequently in PsycEXTRA, but we have not obtained permission to necessarily include images in this document.]

Some Content Owners with Images from Mass Digitization Project

NASA



NEWS RELEASE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
400 MARYLAND AVENUE, SW, WASHINGTON 25, D.C.
TELEPHONES: WORTH 2-4155 — WORTH 3-6925

FOR RELEASE: March 3, 1963 Sunday

RELEASE NO. 63-45

Maryland University Conducts Behavior Research for Manned Space Flight

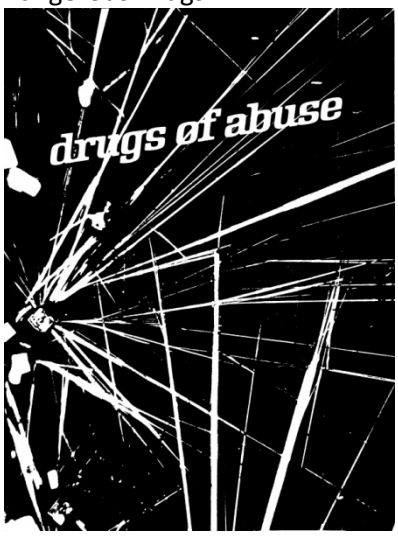
A unique study of human behavior in a controlled environment involving prolonged isolation is being conducted for NASA by the University of Maryland.

The experiment by the University's Department of Psychology is intended to provide basic information for planning future space flights for prolonged periods of time.

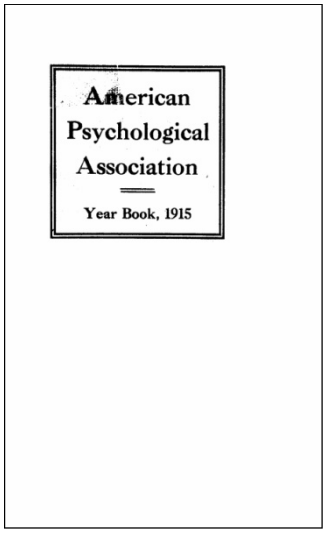
- more -

Southwestern Psychological Association

Bureau of Narcotics and Dangerous Drugs

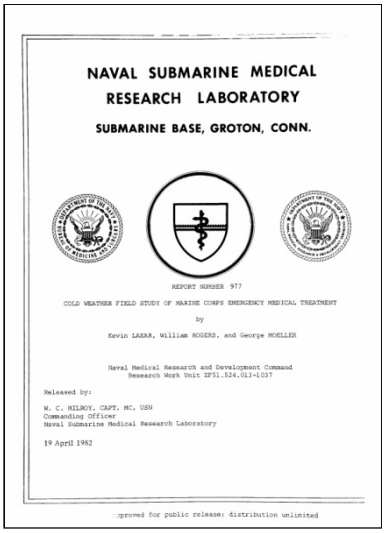


American Psychological Association

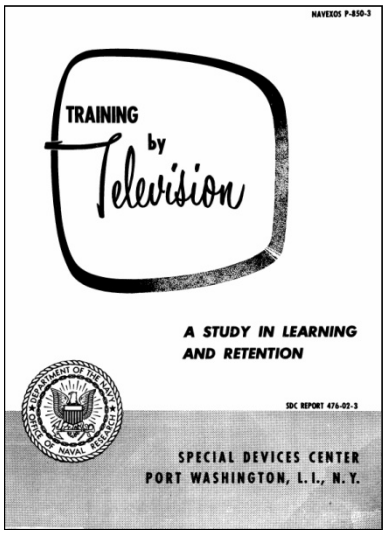


Rocky Mountain Psychological Association

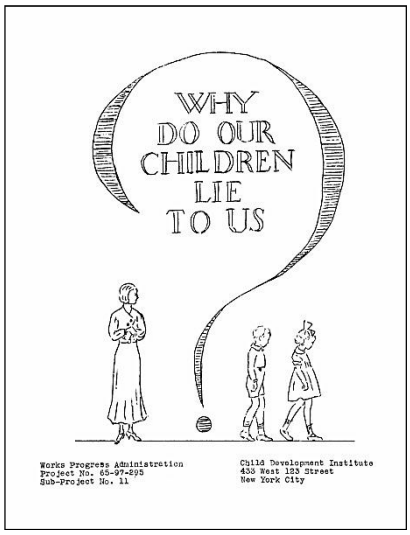
Naval Submarine Medical Research Laboratory



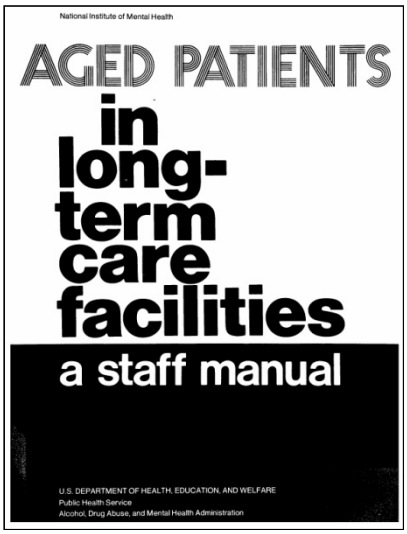
Special Devices Center



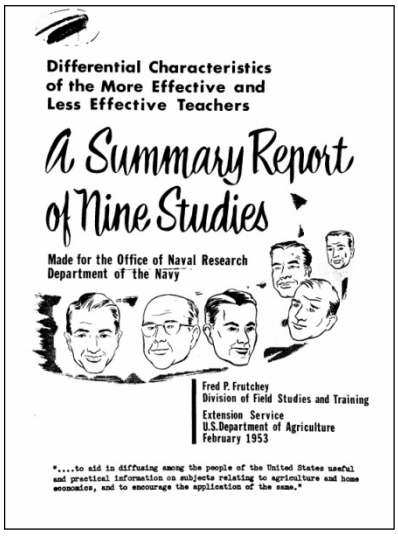
Works Progress Administration



National Institute of Mental Health

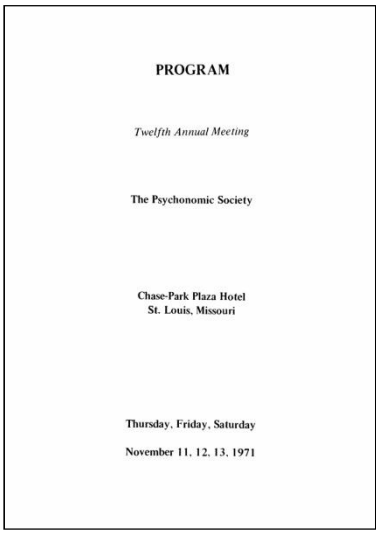


US Department of Agriculture



US Department of Labor Children's Bureau

Psychonomic Society



The Mass Digitization Project

University of Akron and American Psychological Association legal teams prepared a contract to outline the relationship and responsibilities of both partners. For example, it was important to the CHP and to the University that, even while APA holds the right to distribute digitized documents to database subscribers, students, staff, and faculty had a right to unrestricted access to CHP documents in both physical and digital forms (Kaufman, 2007) and it was essential to both partners that careful regard was given to the rights of active copyright owners (Radice, 2009).

Digitization was conducted onsite at the Center for the History of Psychology (then, the Archives of the History of American Psychology) using two Kirtas book scanners that allowed rapid digitization each with a robotic arm for turning pages and two Cannon EOS SLR cameras mounted above the document for grabbing images. APA rented Kirtas scanners and hired the technicians to run them. In phase one of the project, Kirtas technicians captured and manipulated images of document pages and delivered them to APA server. During phase two of the project, the Kirtas machines were replaced with a large-surface flatbed scanner for continued digital reproductions, and images continued to be uploaded by the CHP digitization team to APA server.

During the first phase, each document received a unique barcode by which it could be identified in the Kirtas system. At a later phase, the PsycEXTRA team would rename these files within its own system, while maintaining the original barcode as part of the metadata so the original documents could be easily located when necessary.

Three concerns arose in regards to working with hired technicians and rented machinery. Because the grey literature digitization project ran concurrent to a rare book digitization project, the same scanners captured the materials. First, much of the grey literature in the CHP collections is not in book form, so book scanners may not have always offered the most efficient digitization method. Second, common archival practices prohibit the use of automatic feeds for making copies of historical documents, and because of this, it was usually impossible to rely on the automatic features and robotic arms to turn fragile pages of document artifacts. Third, the technicians were experts in using their equipment for rapid mass digitization, though not necessarily experts in careful archival practices, document handling, document organization, and appraisal of document condition and sustainability through the process (Teper, 2011).

Physical documents were organized four times as a result of the many necessary phases from discovery to repository. Those phases are: 1- initial organization of documents by institution, military branch, or government office in order to categorize unknown materials; 2- secondary alphabetical organization of publishing or sponsoring agencies within government or otherwise broad categories, like the Air Training Command, Air University, Human Resources Directorate, and School of Aviation Medicine (few of many laboratories within the Air Force publications); 3- tertiary organization of documents within categorized groups, usually by report numbers, years, volumes, or issues; (In this order, documents were assigned a barcode and queued for digitization); 4- final reorganization of scanned documents and placement in vertical files for on-site viewing.

The Costs of Mass Digitization

Costs were split between the grey literature project and a concurrent book scanning project. This drove down overall project costs due to economies of scale and the guarantee that machines would never be idle while material was being prepared for scanning.

APA contracted with Kirtas for onsite digitization. We rented high-end planetary scanners based on the needs of the book scanning project. Full-time technicians were hired and trained by Kirtas to operate the machines. High resolution image files were sent via FTP to a Kirtas facility where they were made into pdfs, optimized, and OCR'ed. Kirtas sent finished files to PsycEXTRA in Washington, DC, for use in the database.

Scanning began in March of the first year with two scanners and two technicians. In October of that year, we stepped down to one Kirtas scanner as the bulk of the work on the concurrent book project was complete. In June of the following year, we concluded our relationship with Kirtas and moved to a flatbed scanner procured by APA and operated by CHP staff.

Costs fell under three general categories: scanning machine rental, per-page scanning charges, and personnel. Rental costs were fixed per machine and steady throughout the project. The per-page scanning charge was more expensive for the grey literature than for books. The grey literature contained fewer

images per file on average, increasing set-up and processing times. Because the automated page turner often could not be used, images were generated more slowly.
APA contracted the hiring and management of onsite staff to CHP. In addition to the project manager and scanning operators, students at the University of Akron were hired to help with retrieval and restacking material. Additional personnel were assigned to the PsycEXTRA database to accession and index the new material. Technical staff at University of Akron and APA were also called upon to deal with problems that crept up at the intersection of the various file handling systems.

The Resulting Benefit for APA: PsycEXTRA
PsycEXTRA staff match PDF files to content publishing organizations and investigate files for copyright restrictions. Those documents for which we have an agreement with the content owner, or which are determined to be in the public domain, are accessioned into PsycEXTRA production system. Other documents are retained in a dark archive while staff reach out to the organizations for permission to reproduce the work in PsycEXTRA.
PsycEXTRA staff then create bibliographic records. These records, like all PsycEXTRA records, include abstracts, are indexed with controlled vocabulary, and classified using the same subject system as that used for PsycINFO. Each record for a CHP-sourced document includes the sentence: “This historic document is included through collaboration with The University of Akron, The Archives of the History of American Psychology, University Libraries.”
Currently almost 15,000 documents from CHP have been released into PsycEXTRA. Created in 2004, PsycEXTRA has about 230,000 records covering historical and current grey literature on subjects relevant to the behavioral science community. Inclusion in PsycEXTRA makes these documents available to individual and institutional subscribers throughout the world.

The Resulting Collections for the CHP
The CHP/APA scanning project represents an attempt to build a partnership between subject-specific special collections and the disciplines with which they are most closely aligned. This partnership has proven to be a very successful one, resulting in the digitization and distribution of 20,000 documents from the CHP Grey Literature collection. In addition to being available through PsycEXTRA, the documents are available to researchers onsite at the CHP and digitally to students, faculty, administration, and staff of the University of Akron, and up to 10% at any time as part of the CHP web-presence. Further, physical arrangement of documents on site at the CHP is now organized in a system of vertical files, currently using 60 sets of filing cabinets with 10 more on deck to accommodate the daily growth of the physical collection. Access to both physical and digital versions of documents remains a priority for the CHP, because, to the archivist and to the historian, “the physical manifestations in which [writings] originally were expressed” (Radice, 2009) are often as important to the historical record as the contents of those documents.
The rapid mass digitization of the grey literature forced the digitization team to worry later about descriptive and bibliographic metadata in order to maintain a constant availability of documents in the scan queue while the equipment was still under the rental agreement. As a result, the CHP staff and researchers are now faced with a massive backlog of documents to index in its onsite system and collections—a recognizable dilemma in rapid and mass digitization projects (Gueguen, 2009), even without addressing legacy data for digital objects. Not all documents digitized are yet available in PsycEXTRA nor did all documents uncovered in the process qualify for digitization. CHP staff and students have created to date more than 50,000 records in local databases, thereby opening access to this impressive collection of grey literature. During the metadata creation phase, the PsycEXTRA team added the following byline to the abstract of each document, thereby enabling easy identification of the CHP materials in PsycEXTRA for the CHP community: “This historic document is included through collaboration with The University of Akron, The Archives of the History of American Psychology, University Libraries.”

Sustaining Relationships
At least three distinct relationships have been sustained as a result of the partnership: 1- the Center for the History of Psychology and the American Psychological Association; 2- the donor and research community who supply and access previously unprocessed documents in CHP collections; 3- the broader research community and APA.

The CHP and APA continue to hold a professional relationship to digitize grey literature collections and to make the documents accessible via PsycEXTRA and other APA-published databases. Currently, the CHP Digitization Center is focused on selection and digitization of documents in the CHP Test Center: paper-and-pencil tests designed to measure intelligence, aptitude, mental and emotional capacity, occupational preferences, academic achievement, and perception. Years of publication of this historical test collection ranges from the 1890s to the present, though not all documents qualify for digitization under APA projects. Careful selection based on public domain and copyright registration renewal results in a stream of documents scanned for PsycTESTS , a new addition to the PsycINFO suite of databases. CHP Digital Projects do not, at this point, digitize materials under Section 108 of Title 17 of the United States Code, that is, to preserve content simply for the sake of generating an archival facsimile (U.S. Copyright Office, 2009) as part of the digital projects partnership with APA. PsycEXTRA is an example of APA's continuing commitment to provide a full range of information products to the psychology community. Archiving information in PsycEXTRA guarantees that these documents will have an electronic home for today's and tomorrow's researchers. This partnership has drawn attention to CHP collections and growing access to archival documents, which inevitably has increased visibility in the research community and strengthened relationships with donors and researchers across the globe. Separate from the CHP researchers can access these documents that have been integrated into the databases of the American Psychological Association to enrich access to historical documents in their collections.

Concluding Thoughts

The grey literature collection at the Center for the History of Psychology was the catalyst for a functional and continuing relationship with the American Psychological Association that produced a research database to diffuse otherwise unknown historical documents. The Center employs digitization experts and supporting students in order to continue document discovery and diffusion on the digital landscape that have increased visibility of collections. The benefits of the collaboration continue to place the right documents into the hands of historians, psychologists, and scholars around the globe.

References

Aina, L. O. (2000). Grey literature and library and information studies: a global perspective, *International* ***Journal on Grey Literature, 1*****(4), 179-182.**

Conn, V., Valentine, J., Cooper, H., & Rantz, M. (2003). Grey literature in meta-analyses. *Nursing Research*, 52, 256-261.

Faragher, E. B., Cass, M., Cooper, C. L. (2005). The relationship between job satisfaction and health: a meta-analysis, *Occup Environ Med, 62*, 105-112.

Gueguen, G. and Hanlon, A. M. (2009). A collaborative workflow for the digitization of Unique Materials, *Journal of Academic Librarianship, 35*(5), 468-474, doi:10.1016/j.acalib.2009.06.001.

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee, G. D., and Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review, *Med Teach, 27*(1), 10-28.

Kaufman, P. B. and Ubois, J. (2007). Good terms: Improving commercial-noncommercial partnerships for mass digitization, *D-Lib Magazine*, *13*(11/12).

Kirtas Technologies Automatic Book Digitization, http://www.kirtas.com/bookscanners.php

Radice, A. (2009). Digital Debates, *Conference Proceedings from Web Wise, Institute of Museum and Library Services*: Washington, DC.

Radyowijati, A. & Haak, H. (2003). Improving antibiotic use in low-income countries: an overview of evidence on determinants, *Social Science & Medicine, 57*(4), August 2003, 733-744, 10.1016/S0277-9536(02)00422-7.

Ranger, S. L. (2005). Grey literature in special libraries: Access and use, *Publishing Research Quarterly, Spring*, 53-63.

Sulouff, P., Bell, S., Briden, J., Frontz, S. and Marshall, A. (2005). Learning about grey literature by interviewing subject librarians: A study at the University of Rochester, *C&RL News, July/August*, 510-515.

Teper, J. H. and Shaw, E. F. (2011). Planning for preservation during mass digitization projects, *portal: Libraries and the Academy, 11*(2), 717-739.

Thornton, A., and Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53*(2), 207-216. doi: 10.1016/S0895-4356(99)00161-4

U.S. Copyright Office (2009). *Circular 21: Reproduction of copyrighted works by educators and librarians*, U.S. Copyright Office, http://www.copyright.gov/circs/circ21.pdf.

von Braun, J., Teklu, T., & Webb, P. (1999). *Famine in Africa: Causes, responses, and prevention.* Baltimore, MD: Johns Hopkins University Press.

Willett, P. (2009). Mass digitization and its impact on interlending and document supply, *Interlending & Document Supply, 37*(3), 143 – 148.

Grey Literature Between Tradition and Innovation: Is There a Continuum?

Gabriella Pardelli, Sara Goggi, and Manuela Sassi
Istituto di Linguistica Computazionale, "Antonio Zampolli"
Consiglio Nazionale delle Ricerche (CNR), Italy

Abstract
This study wants to explore new ways of social media communication for Grey Literature. In particular it describes the role of social media in relation with traditional channels and how social media applications can be used for Grey.

Keywords: Grey Literature, Communication networks, Knowledge networking, Knowledge exchange.

1. Introduction

From the Vth century, when monasteries started to be built in Western Europe, the transcription of ancient literary works was made in one of the abbey rooms called 'scriptorium'. This method for the replication of knowledge aimed at retrieval and conservation rather than at divulgation: there was a selection of the works to bequeath and preference was given to Latin Classics, religious works and some heathen works as well.

This Medieval way of knowledge was preservative rather than popular and was addressed to a niche of a few users: its main merit was to pass on knowledge from age to age and transmit it to our time Information Society (IS). Information Society is a computer-based society which spreads knowledge, meets on social networks, establishes realities such as Wikipedia and Facebook thus reducing the world into a global village. Information currently means knowledge and becomes a new form of wealth, since productivity and welfare are influenced by the management and the transmission of information. The Medieval "scriptorium" is somehow rebuilt on the web with a new transcription system called "digitalization": knowledge is not only created but also disseminated for then being shared. The computer-based culture offers people the opportunity for a cultural growth while technology creates new professions daily and enriches the traditional crafts with new contents.

Nowadays on the web there is a rapid growth of new ways of learning dictated by the massive information exchange and the shared knowledge: "The web is immense, free and available by mouse-click. It contains hundreds of billions of words of text and can be used for all manner of language research" (Kilgarriff and Grefenstette, 2003).

Internet has become a universal repository where users, by means of web-based interfaces, can extract information. But not only: the web makes available to users what is called "social networking", that is a tool of interactive approach, a collective voice and a new way of communicating. The interactive nature of this tool makes information expand and creates an added value which enlarges knowledge.

Discussion and shared interest for knowledge is the adhesive of social network members; and social networks are a sort of modern literary saloons where the meetings have a "free" and spontaneous nature, participants have a common socio-cultural background and the participation itself is the intrinsic purpose. This said, the importance of social networks lies just in the richness of relations – human, cultural, scientific, economic - which can be developed.

Is the web – alongside with its new ways of social media communication - the modern form of medieval "scriptorium"? How do the social media applications relate to Grey Literature and how can they be employed for research on Grey?

Is there a continuum? Answering to these questions will be the scope of this article.

2. Why Social Media, now?

In January 2010, *The Economist* published a special report on social networking titled "A world of connections": on page 14 we read: "This democratization of technology is driving the socialisation of the web and fundamentally changing the way that people interact with one another, as well as with businesses and governments".

The use of technology becomes democratic, that is available and easily accessible to most people regardless of age, culture, country and economic condition. This is the first time in history that a free communication system increasingly becomes a flow of information through space and time. This process does not duplicate knowledge as it happened in the Medieval 'scriptorium' nor store it but is a simple transmission of news and events happened all over the world, a stream of thought which relates all those who are connected.

Nowadays the concept of *Social Networking Sites (SNS)* does not imply the presence of a final user considered as a target because everyone has the same needs and the social network becomes a virtual society, a tribe where the community can find reasons to join. Social networking therefore stands for a tool of interactive approach and means a collective place where people aggregate and communicate: information runs on the web thus enlarging and making knowledge shareable to most citizenship. This is the richness of today's society, where the rapid increase and the ever stronger need of human relationships run alongside. Google – first brand in this society's world - is its new paradigm: the immaterial.

2.1 Why Social Media for Grey , now?

Over the last decade both Grey Literature and traditional literature underwent the transformation of the communication channels of scientific information: the consequence has been a drastic change of the editorial look which became increasingly headed towards electronic publication on the web. For a long time books, journals and grey material have been shipped to users by ordinary transport links while nowadays information runs on the web and delivery means open access for various types of documents such as publications of digital nature from the origins and publications which have been digitally transformed subsequently.

This material has been stored little by little on ad-hoc academic, industrial and governmental web sites and afterwards, at a international level, common topics have been further assembled on powerful infrastructural web sites (in our field good examples are repositories such as OpenSIGLE and the ACL Anthology). The aim is to concentrate knowledge for then disclose, disseminate and share it with citizens: obviously the purpose is no longer to conserve and preserve knowledge for transmission to a niche of few users as it used to be in the Middle Ages.

These days the interactive nature of social networking allows a very close – and inconceivable till a few years ago - relationship between producers and users of knowledge: the rapid quickening of information delivery causes the gathering of politicians, academics, investors for defining politics, generating knowledge and producing wellness. By means of the social networking sites system (SNSs) – the new added value to production and expansion of knowledge – Grey Literature is involved as well, but how? Groups of Grey Literature scholars with common interests and goals discuss, ponder and compare with new methodologies and new type of information for trying to retrieve/define new paradigms. In this modern XVIIIth century-like literary saloon, culture is created and spread around with the purpose of establishing clusterings of contents and objectives. For sure a new Grey generation is born and the answer to the question " why social media for Grey, now" is very simple: "It is just moving with the times".

3. Grey Literature between tradition and innovation: is there a continuum?

> "Les concepts font l'objet d'une description systématique qui recens different types d'informations don't les plus fréquentes sont: le terme utilize pour designer le concept, la definition de concept via celle du terme utlilisé, le domaine et le sous- domains d'employ du terme, les éventuelles variantes (orthographiques ou synonymes de type géographique, par example) et les equivalents dans d'autres langue."(Kister, Jacquey, Gaiffe,2000:27)

The importance of terms – seen as anchors for information retrieval and pointers of time, science and technology - will be discussed in the next paragraph.

3. 1 Social Networking and words for Grey

The linguistic tools developed within the "DylanLab Lab for Computational Models of the Dynamics of Language and Cognition" of the Institute of Computational Linguistics allowed a textual analysis of the abstracts belonging to the 'Social Networking' session of this *Thirteenth International Conference on Grey Literature*.

Table 1. Single Terms

ID	single terms	relevance
1	knowledge	100.0%
2	networking	95.1%
3	content	76.0%
4	library	75.5%
5	web	71.3%
6	database	71.3%
7	literature	71.2%
8	user	70.8%
9	information	68.8%
10	tool	68.6%
11	communication	67.0%
12	access	60.4%
13	community	58.5%
14	environment	51.9%
15	research	51.4%
16	sharing	50.7%
17	site	50.6%
18	health	49.1%
19	representation	45.4%
20	technology	42.9%

Table 2. Relevant Multiple Terms

ID	relevant multiple terms (general purpose relevance)	relevance
1	grey literature	100.0%
2	ability of user	60.4%
3	new searchable database	60.4%
4	interest group	52.4%
5	presence blending information product	42.4%
6	community of grey literature	40.2%
7	discoverability of grey literature	39.2%
8	kind of grey literature	38.6%
9	interactive communication	38.1%
10	social networking	38.1%
11	incremental approach	38.1%
12	mode of knowledge sharing\creation	38.1%
13	range of social networking	38.1%
14	concept from large document	38.1%
15	area of general medicine	38.1%
16	many more specialized example	38.1%
17	communication with much variation	38.1%
18	tool for volatile knowledge	38.1%
19	online research interest group	38.1%
20	realm of human interactivity	38.1%

Table 3. Domain-Specific Multiple Terms

ID	domain-specific multiple terms (text-dependent relevance)	relevance
1	grey literature	100.0%
2	social networking	100.0%
3	ability of user	100.0%
4	new searchable database	100.0%
5	interactive communication	100.0%
6	incremental approach	100.0%
7	interest group	100.0%

Although informative contents are usually independent from the technological wrapping, the analysis of the words of the three tables suggests that there should be a connection between inside contents and outside technology: as a matter of fact 12 out of 20 terms of Table 1 refer to the Information and Communication Technology (ICT) domain (*knowledge, networking, web, database, user, information, tool, access, communication, sharing, site, technology).* The first seven terms have a relevance between 100% and 68,8%.

As far as the relevant multiple terms of Table 2 are concerned, 10 of them are referring to topics proper to Computer Science domain: *ability of user, new searchable database, interest group, presence blending information product, interactive communication, social networking, mode of knowledge sharing/creation, range of social networking, tool for volatile knowledge, online research interest group* and, at last, *realm of human interactivity,* just in reference to social networking. The relevance of the first six terms is between 60,4% and 38,1%.

Finally in Table 3, the domain-specific multiple terms become seven and only the first one, *Grey Literature*, does not refer straight to Computer Science while the other six words can be divided in two subsets: 1) *social networking, interactive communication, incremental approach, interest group*; 2) *ability of user* and *new searchable database.* The four terms of the first set are the perfect summary of the second paragraph of this article while the two terms of the second set are directly connected with the concept of repository and the important ability to consult the stored information for extracting knowledge.

In this scenario the role of information extraction technology becomes increasingly relevant: the priority is to clean up the scene, to sort out linguistic phenomena such as synonimity and homonymy by using domain ontologies. Simply retrieving digital information from academic, industrial and governmental documentation is not enough: nowadays intelligent information retrieval is needed and required.

"Nowadays the need of retrieving the great amount of digital knowledge available on the web is ever more important: the vast majority of this knowledge is conveyed by means of textual material stored in scientific documentary repositories and digital archives. This documentary word preserve inside the wealth of far-off terms belonging to the past which have often fallen out of use as well as a more recent terminological production derived from the feverish need of coining new terms, a very common trend in certain branches" (Sassi et alii, 2009).

Scientific dissemination in the field of Grey has therefore to be closely connected to the possibility of being easily and quickly retrieved from the web infrastructures.

To conclude, it might be asserted that finding continuity between the old traditional way of producing and sharing Grey Literature and the current technological methods is a difficult task; continuity is rather to be found at the level of contents because Grey Literature will always be Grey Literature, over time and technology.

References

BANKS M. (2010). Blog Posts and Tweets: The Next Frontier for Grey Literature. In Dominic J. Farace, and Joachim Schöpfel (Eds.), *Grey Literature in Library and Information Studies.* De Gruyter. 217-226.

BUHL HANS U. (2011). From Revolution to Participation: Social Media and the Democratic Decision-Making Process. *Business & Information Systems Engineering*, Volume 2, Number 4. Online First™, 16 June 2011. DOI 10.1007/s12599-011-0166-4. Gabler Verlag 2011.
This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Buhl HU (2011) Von der Revolution zur Partizipation: Social Media und der demokratische Willensbildungsprozess. WIRTSCHAFTSINFORMATIK. doi: 10.1007/s11576-011-0279-x.

CHIH-LU LIN, HUNG-YU KAO (2010). Blog popularity Minining. Using Social Interconnection, Analysis. *IEEE Internet Computing*, July/August. 41-49

GUPTA S. THAKUR N. (2010). Semantic Query Optimisation with Ontology Simulation. *Semantic Query Optimisation with Ontology Simulation*, 1(4), 1 - 10.

KILGARRIFF A., GREFENSTETTE G. (2003). Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics,* Volume 29, 3, Association for Computational Linguistics, 333-347.

Kister L., Jacquey E., Gaiffe B. (2008). Réperage de la reference à partir du thesaurus, de la terminologie et de la sémantique lexical. In Terminologia analisi testuale e documentazione nella città digitale. AIDA informazioni 1-2/2008. 25-36.

MAXWELL HARPER F., XIN LI S., CHEN Y., KONSTAN JOSEPH A. (2007). Social Comparisons to Motivate Contributions to an Online Community. Lecture Notes in Computer Science, 2007, Volume 4744, Persuasive Technology, Pages 148-159.

PIEVATOLO MARIA C. (2010). Back to the future: authors, publishers and ideas in a copy-friendly environment. In *Open Scholarly Communities on the Web: Legal, Economic and Social Framework*, Oxford (United Kingdom),21-22 April 2008.(Unpublished) [Conference Paper].

ROSEN D., BARNETT GEORGE A., KIM JANG H. (2011). Social networks and online environments: when science and practice co-evolve. Social Network Analysis and Mining, 2011, Volume 1, Number 1, Pages 27-42.

SASSI M., PARDELLI G., BIAGIONI S., CARLESI C., GOGGI S. A Digital Archive of Research Papers in Computer Science. In*: LREC'10 - Seventh International Conference on Language Resources and Evaluation* (Valletta, Malta, 17-23 May 2010). In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, Mike Rosner, Daniel Tapias (eds.), Proceedings. Paris, European Language Resources Association ELRA, 2010. Pages 1245 - 1248.

SILLANPÄÄ N., ÄLLI S., ÖVERMARK T. (2010). Easy-to-Use Social Network Service, *Lecture Notes in Computer Science*, 2010, Volume 6179, Computers Helping People with Special Needs, Pages 544-549.

SCOTT J. (2011). Social network analysis: developments, advances, and prospects. *Social Network Analysis and Mining,* 2011, Volume 1, Number 1, Pages 21-26.

SØRENSEN ANNE T., SHKLOVSKITHE I., (2011) Hugging Team: The Role of Technology in Business Networking Practices. In Susanne Bødker, Niels Olof Bouvin, Volker Wulf, Luigina Ciolfi and Wayne Lutters (eds.), ECSCW 2011, *Proceedings of the 12th European Conference on Computer Supported Cooperative Work*, 24-28 September 2011, Aarhus Denmark. Springer-Verlag, London . 333-352.

SØRENSEN LENE T., SKOUBY KNUD E. (2009). Requirements on Next Generation Social Networking—A User's Perspective. *Wireless Personal Communications* ,2009, Volume 51, Number 4, Pages 811-823.

THE ECONOMIST, (January 30th 2010) .A World of Connections: A Special Report on Social Networking".

WODZICKI K., SCHWÄMMLEIN E., CRESS U. (2009). How to Get Proper Profiles? A Psychological Perspective on Social Networking Sites. *Lecture Notes in Computer Science*, 2009, Volume 5794, Learning in the Synergy of Multiple Disciplines, Pages 338-343.

OpenSIGLE Repository, System for Information on Grey Literature in Europe: http://www.greynet.org/opensiglerepository.html

ACL Anthology, A Digital Archive of Research Papers in Computational Linguistics: http://aclweb.org/anthology-new/index.html

A linked data Vocabulary of the Types of Grey Literature: Version 1.0

Petra Pejšová, Tereza Simandlová, and Jindřich Mynarz
National Technical Library, Prague, Czech Republic

Abstract
The aim of this article is to introduce the Vocabulary of the Types of Grey Literature, the development of which was initiated at the National Technical Library in Prague, Czech Republic. The vocabulary is created in a way compliant with the principles of linked data. The 1st version of the vocabulary is available from the project website http://code.google.com/p/grey-literature-typology

Keywords: grey literature, controlled vocabulary, typology, RDF, SKOS, linked data

Vocabulary overview

The Vocabulary of the Types of Grey Literature (henceforth vocabulary) is a controlled vocabulary in RDF (Resource Description Framework) data format expressed as a SKOS (Simple Knowledge Organization System) concept scheme. The vocabulary is based on an analysis of six existing grey literature typologies[1], which was presented at the Twelfth International Conference on Grey Literature (GL12) in 2010. *"The aim of the analysis was to create, define, and implement a current credible grey literature document typology, in order to open discussions in the grey literature community, leading to a means of collecting GL from reputable events and producers rather than relying on social networking tools or Wiki contributions. While the later types of sources can assist researchers, scientists, and teachers with their information-seeking pursuits, documents of this nature need to be evaluated on a regular basis."* (Pejsova, 2011)

The vocabulary can be seen as a formalization of the outputs of the systematic examination done during this analysis. This means that the typology is built upon the already available typologies for grey literature, creating an intersection of their features that were found to be important, while excluding duplicates and merging the same types with distinct labels.

The vocabulary has a loose structure with hierarchical relations. *"Each type is provided with a definition and most of them are exemplified with a prototypical example of a document for which it can be used. By design, the typology is focused on the description of types. Other documents' attributes, such as content or format annotations, are excluded from the vocabulary."* (Grey-literature-typology, 2010)

Moreover, the vocabulary is published as linked data. Linked data is a publication model for exposing structured data on the Web in a way that uses links between datasets to create a network of interlinked data. Each of its types is identified by URI and the vocabulary is interlinked and mapped to other datasets.

Working group

The Working Group for Grey Literature Typology[2], established to support the further development of the vocabulary, consists of a set of experts from various domains, ranging from grey literature and knowledge organization systems, to semantic web technologies.

The core part of working group is formed by the National Technical Library (NTK) project team, which is the informal body coordinating the vocabulary's development. The members of the working group are responsible for building the vocabulary and its updates, maintaining the project's website, incorporating changes, determining the work time schedule, publishing new versions of the typology, providing support in the case of implementation of the vocabulary, and presenting the project and the vocabulary.

NTK project team's members are (all from the National Technical Library, Czech Republic):

- Petra Pejšová
- Jindřich Mynarz
- Tereza Simandlová

Other members of the working group are responsible for providing feedback on the vocabulary's development via comments or by filing issues, addressing the lists of proposed changes on the call by the project owners, presenting the vocabulary and supporting the vocabulary's implementations in the systems of grey literature. The members of the extended working group are:

- Marcus Vaska – University of Calgary, Canada
- Vojtěch Svátek – University of Economics, Czech Republic
- Anne Gams Steine Asserson – University of Bergen, Norway
- Claudia Marzi – CNR Institute for Computational Linguistics, Italy

- June Crowe – Information International Associates, Inc., U.S.
- Keith G Jeffery – STFC Rutherford Appleton Laboratory, United Kingdom
- Marta Dušková – Slovak Centre of Scientific and Technical Information, Slovakia
- Joachim Schöpfel – Charles de Gaulle University of Lille 3, France
- Dominic J. Farace – GreyNet, Netherlands
- Daniela Luzi – IRPPS-CNR, Italy

Development

The typology is developed collaboratively as a Google Code project. As we mentioned above, its draft version, based on analysis of the six existing grey literature typologies and the result, was presented at GL12 in 2010. The draft version of the vocabulary was published in January 2011 and being reviewed by the 13-member international working group until July 6th, 2011. The members of the working group primarily analyzed the structure of collections and document types definitions and commented directly to http://code.google.com/p/grey-literature-typology/issues/list .

NTK's team gathered 42 comments and suggestions that were accepted and incorporated into the first version of vocabulary. Changes mainly redefined some collections, when some document types were merged and others excluded. The major changes concerned the following collections: monographs, informative materials, datasets, and research plans. The first version of the vocabulary is available for use from November 2011 on the project website http://code.google.com/p/grey-literature-typology. Hierarchical structure of the current version is consisting of 16 main types, which we called *collections*, and 55 subtypes in these collections. Draft and first version of the vocabulary is available for the purposes of comparison in appendix 1 and 2.

The development cycle of the vocabulary was planed to be bi-annual, but it proved to be unrealistic, and it was re-assessed to release a new iteration of the vocabulary once a year. All versions will be issued via the Google Code repository and will be persistently linkable.

Example:

```
gltype:conference-proceedings a skos:Concept ;
       skos:inScheme gltype: ;
       skos:prefLabel "Conference proceedings"@en, "Sborník z konference"@cs ;
       skos:broader gltype:conference-material, gltype:proceedings ;
       skos:definition "Conference proceedings are the collection of academic papers that
       are published in the context of an academic conference. They are usually
       distributed as printed books or electronic version either before the conference
       opens or after the conference has closed. Proceedings contain the contributions
       made by researchers at the conference."@en ;
       skos:editorialNote "Taken from Wikipedia, the Free Encyclopedia:
       http://en.wikipedia.org/wiki/proceedings, retrieved May 23, 2011"@en  .
```

Usage example:

```
<> dc:type gltype:conference-proceedings .
```

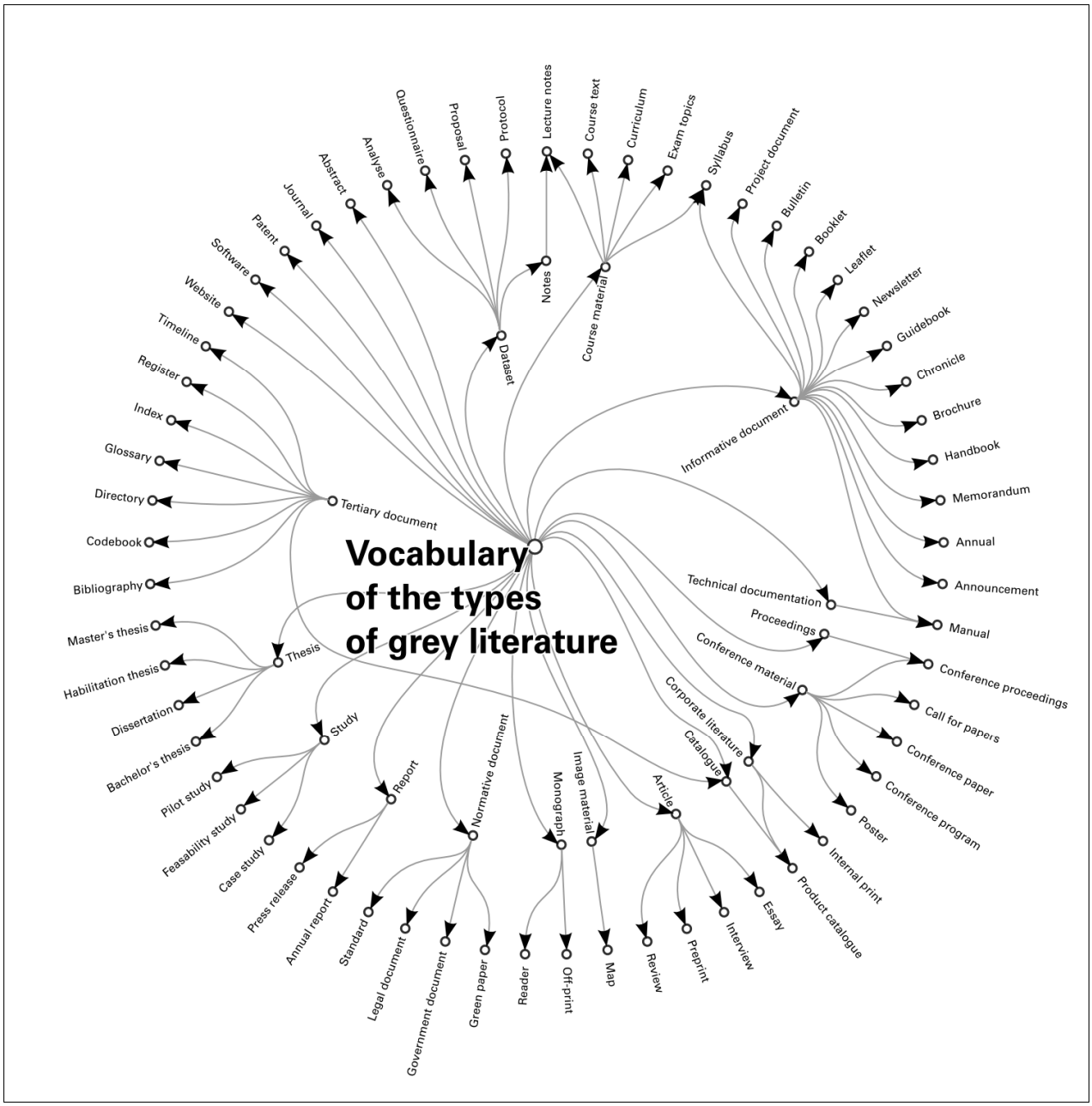
Benefits

The vocabulary is the first functional general model that can directly implemented in grey literature repositories. Unlike the other analyzed typologies, the vocabulary is available in a machine readable format. Each grey literature type has a definition and has a globally unique identifier, and as such can be unambiguously a persistently referenced. Another benefit of the vocabulary is explicit mapping to other vocabularies, e.g., Dublin Core Metadata Initiative Types, or DBPedia.

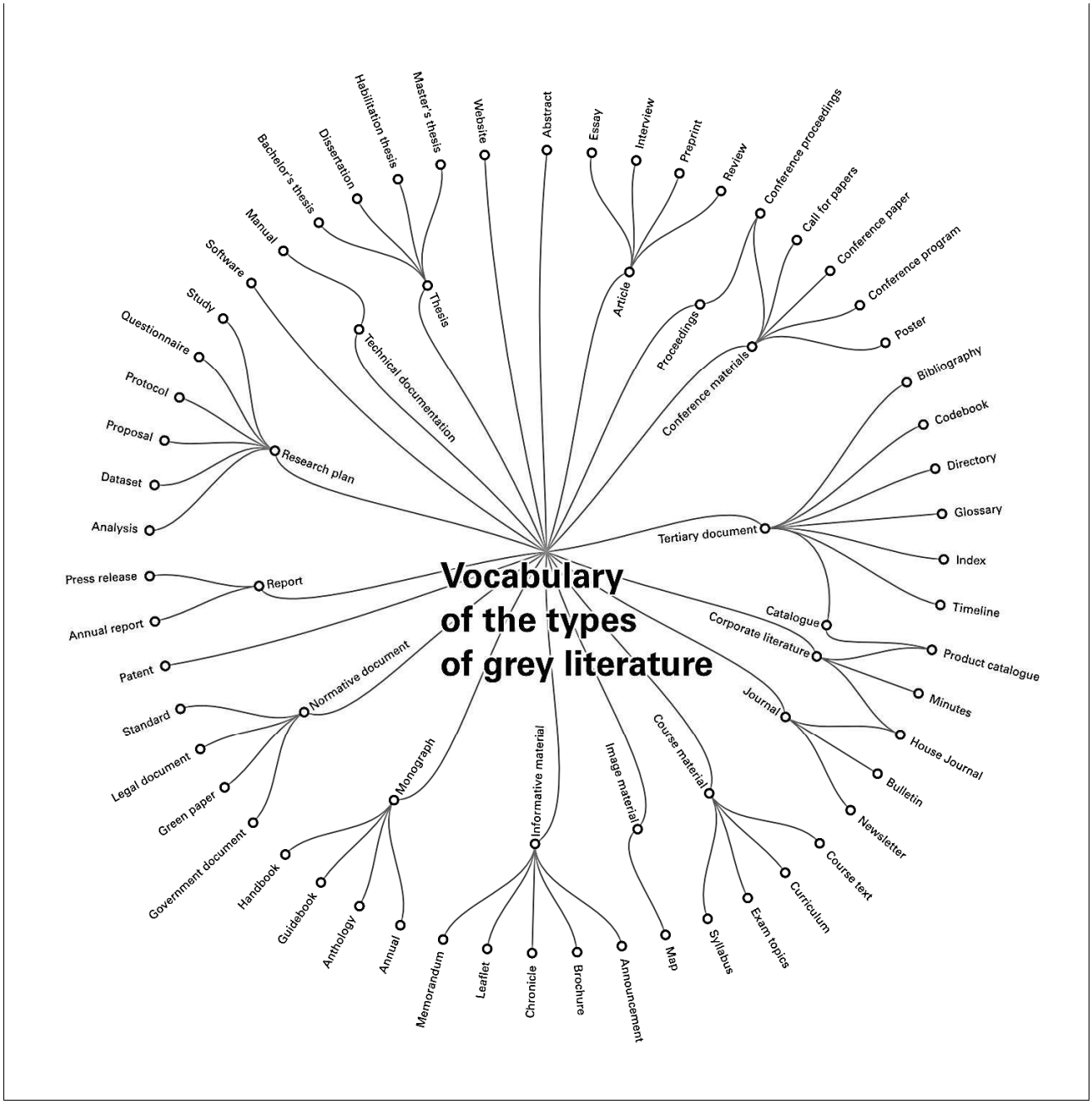
Future

Working group expects that the vocabulary will find a specific implementation. The first interest in implementation was expressed by the Slovak Centre of Scientific and Technical Information for their national system for collecting grey literature. Effort of the working group has not been stopped, the first version is prepared for further comments and suggestions. These will be included in updated versions. Now, only 10 % of document types are provided with an example. We plan to accompany all of the grey literature document types with examples in 2012. And last but not least, the vocabulary is available only in Czech and English. We are looking for experts for translation to other languages.

Apendix 1: Draft version of the Vocabulary of the Types of Grey Literature



Apendix 2: The version 1.0 of the Vocabulary of the Types of Grey Literature



References

Grey-literature-typology (2010). Vocabulary of the Types of Grey Literature. Retrieved November 20, 2010 from <http://code.google.com/p/grey-literature-typology/>.

PEJŠOVÁ, Petra; VASKA, Marcus. An Analysis of Current Grey Literature Document Typology. In *Twelfth International Conference on Grey Literature : Transparency in Grey Literature, 6-7 December 2010.* Amsterdam : TextRelease, 2011. ISBN 978-90-77484-16-6. ISSN 1386-2316.

[1] GreyNet, (the Grey Literature Network Service), OpenSIGLE, (the System for Information on Grey Literature in Europe), and the Registry of Open Access Repositories (ROAR), as well as focusing on national schemata in the Czech Republic, namely ASEP (Register of Publication Activity of the AS CR), NRGL (National Repository of Grey Literature), and RIV (Information Register of R & D Results)

[2] http://code.google.com/p/grey-literature-typology/wiki/Project_members

Pre-Conference Announcement

Fourteenth International Conference on Grey Literature

Tracking Innovation through Grey Literature

Rome, Italy 29-30 November 2012

National Research Council – CNR

Central Library “Guglielmo Marconi”, IRPPS Rome, ISTI Pisa

GL14 Program and Conference Bureau

TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, The Netherlands
www.textrelease.com conference@textrelease.com
Tel/Fax +31-20-331.2420

List of Participating Organizations

American Psychological Association, APA	United States
Bangalore University	India
BMS College of Engineering	India
Center for the History of Psychology	United States
Centre National de Recherché Scientifique, CNRS	France
Centre of Information Technologies and Systems of Executive State Authorities	Russia
Consiglio Nazionale delle Ricerche, Biblioteca Centrale	Italy
Data Archiving and Networked Services, DANS	Netherlands
Defense Technical Information Center, DTIC	United States
EBSCO Publishing	United States
European Organization for Nuclear Research, CERN	Switzerland
Federal Library and Information Center Committee, FLICC	United States
Federal Library Information Network, FEDLINK	United States
Florida State University, FSU	United States
Free Pint Inc.	United Kingdom
Government P.U. College	India
Grey Literature Network Service, GreyNet	Netherlands
Health Information Network Calgary, HINC	Canada
Information International Associates, Inc.	United States
Institut de l'Information Scientifique et Technique, INIST	France
Institute for Geosciences and Earth Resources, IGG	Italy
Institute of Information Science and Technologies, ISTI-CNR	Italy
Institute of Research on Population and Social Policies, IRPPS	Italy
International Centre of Informatics and Electronics	Russia
International Council for Scientific & Technical Information, ICSTI	France
Istituto di Linguistica Computazionale, ILC	Italy
Japan Science and Technology Agency, JST	Japan
Korea Institute of Science & Technology Information, KISTI	Korea
Kugler Publications	Netherlands
Law Library of Congress, LLC	United States
Lexis-Nexis	United States
Library of Congress, LC	United States
Library of Congress; Asian Division	United States
Library of Congress; Geography and Map Division	United States
Library of Zeeland	Netherlands
Ministry for Education and Science	Russia
National Research Council, CNR	Italy
National Technical Information Service, NTIS	United States
National Technical Library, NTK	Czech Republic
New York Academy of Medicine, NYAM	United States
Office for Information Systems; CNR	Italy
Office of Scientific and Technical Information, OSTI	United States
Open Source Center, OSIS	United States
Peter Lougheed Knowledge Centre	Canada
Pratt Institute	United States
PricewaterhouseCoopers, PwC	Netherlands
PsycEXTRA	United States
Senate Library	Italy
Slovak Centre of Scientific and Technical Information, CVTISR	Slovak Republic
TextRelease	Netherlands
U.S. Department of Energy, DOE	United States
Université Charles de Gaulle Lille 1	France
Université Charles de Gaulle Lille 3	France
University Corporation for Atmospheric Research, UCAR-NCAR Library	United States
University of Akron	United States
University of Calgary	Canada
University of California, Irvine Libraries, UCI	United States
University of Ljubljana, UNI-LJ	Slovenia
University of Mysore	India
University of South Florida, USF	United States

Reception Welcome

Asian Reading Room — AREA STUDIES ASIAN DIVISION

Good evening ladies and gentlemen,

Welcome to the Library of Congress. My name is Anchi Hoh, Head of Collections Services in the Asian Division. I am very honored to be asked by the GL13 Conference organizers to extend a warm welcome to our distinguished guests this evening.

The Asian Division is one of the four area studies divisions in the Library of Congress. The other three divisions are: the African and Middle Eastern, European, and the Hispanic Divisions. The Asian Division houses a collection of over 3 million items in a variety of formats written in vernacular languages pertaining to Asian studies. We have a total of 27 specialists and technical staff members who provide reference and readers services to the Congress, the scholarly community, and the general public.

In addition to materials published in the traditional formats, the Asian Division also collects grey literature documenting a wide variety of subjects including the growth of civil societies, local folklife in minority languages, and environmental change. The area studies divisions' grey literature collections are expected to grow more rapidly as the Library of Congress continues its endeavors in designing and implementing large-scale and collaborative initiatives to collect research data sets, audio/visual ephemera, digital official foreign government information, as well as other types of the grey literature.

These collaborative initiatives would not be successful without a broadly shared vision and concerted effort. The impressive list of sponsors of this Conference proves just that. The GL13 conference sponsorship includes prestigious institutions from around the globe. We are deeply grateful for your generous support.

The first day of the Conference has been very exciting, informative, and successful. It is absolutely wonderful to be part of a Conference in which interested professionals from the international community gather together, explore, and seek solutions to various issues and challenges concerning grey literature. I am sure you are looking forward, as much as I am, to more exciting lectures and discussions tomorrow.

Thank you very much. Again, welcome to the Library and enjoy your evening and the rest of the Conference.

Dr. Anchi Hoh
The Asian Division
Library of Congress, LJ 150
101 Independence Ave., SE
Washington, DC 20540-4810
Tel: (202) 707-5673
Fax: (202) 707-1724
Email: adia@loc.gov
http://www.loc.gov/rr/asian/

Site Tour

Geography & Map Reading Room

GEOGRAPHY & MAP DIVISION

The Geography and Map Division (G&M) has custody of the largest and most comprehensive cartographic collection in the world with collections numbering over 5.5 million maps, 80,000 atlases, 6,000 reference works, over 500 globes and globe gores, 3,000 raised relief models, and a large number of cartographic materials in other formats, including over 19,000 CDs/DVDs. The online Map Collections represents only a small fraction that has been converted to digital form. These images were created from maps and atlases and, in general, are restricted to items that are in public domain, meaning those which are not covered by copyright.

The Collections

The Geography and Map Division of the Library of Congress provides cartographic and geographic information for all parts of the world to the Congress, Federal agencies, state and local governments, the scholarly community, and to the general public. It is the largest and most comprehensive cartographic collection in the world, numbering over 5.2 million maps, including 80,000 atlases, 6,000 reference works, numerous globes and three-dimensional plastic relief models, and a large number of cartographic materials in other formats, including electronic.

Among the earliest original maps in the collections are three manuscript portolan atlases and 19 portolan charts from the fourteenth through seventeenth centuries drawn on vellum by Italian, Portuguese, and Spanish cartographers. The excellent collection of atlases dates from a 1482 printed edition of Claudius Ptolemy's Geography and includes representative volumes of all significant publishers of atlases for the last five centuries. The atlases cover individual continents, countries, cities, and other geographic regions, as well as the world ranging in scope from general to topical.

Of particular interest to genealogists and local historians is a large collection of U.S. county and state maps and atlases published in the nineteenth and early twentieth centuries. Atlases published during the past four or five decades and covering national, regional, state, and provincial resources form another noteworthy reference group.

The division has an excellent collection of manuscript and printed maps of colonial America, the Revolutionary War, the War of 1812, the Civil War, and the wars of the twentieth century. Supplementing these historical records are photo-reproductions of manuscript maps from various American and European archives. The Hummel and Warner collection include rare manuscript and printed maps and atlases of China, Japan, and Korea from the seventeenth century.

About 55 percent of the maps are individual sheets of large- and medium-scale map series and nautical and aeronautical charts published during the nineteenth and twentieth centuries. Official topographic, geologic, soil, mineral, and resource maps and nautical and aeronautical charts are available for most countries of the world.

The collection of single maps embraces more than two million general and special subject maps of the world and its various political entities, divisions, and subdivisions, with maps of the Americas and countries of the Western Hemisphere predominating. North America, the United States, each of the 50 states, and the largest cities are especially well represented.

Among the numerous county maps and city and town plans are some 700,000 large-scale Sanborn fire insurance maps, in bound and loose sheet volumes. The Sanborn Map Company was the dominant American publisher of fire insurance maps and atlases for over 100 years. Founded in 1867, the firm has issued and periodically updated detailed plans of 12,000 American cities and towns. Some areas are represented by as many as eight different editions. This collection constitutes an unrivaled cartographic and historic record of America's urban settlement and growth over more than a century.

Mr. Ralph Ehrenberg
Chief, Geography and Map Division
101 Independence Avenue, SE
Madison Bldg., Room LM B01
Washington, DC 20540-4650
USA

T. 1-202-707.1992
F. 1-202-707.8531
E. rehr@loc.gov
U. http://www.loc.gov/rr/geogmap

Author Information

Blaaij, Cees de
Cees de Blaaij studied Social and Economic History at the University of Nijmegen and Library Science at the University of Amsterdam. He worked for Ernst & Young, management consultants, and the Institute for Information Law (University Amsterdam) as information professional. At the moment he is working as an academic librarian and coordinator digital services for the Public and Academic Library of Zealand, Netherlands. He took part in several GL conferences. He published several articles on issues concerning copyright in the digital environment and accessibility of grey literature on the Internet. cdbl@zebi.nl

Choi, Heeyoon
Hee-Yoon Choi is the director of Knowledge Information Center at Korea Institute of Science and Technology Information(KISTI), sponsored by the Korean government. Before coming to KISTI, she worked for POSCO Research Institute(POSRI) as the Director of Knowledge Asset Center. She is involved in various other projects such as serving as the Editor-in-Chief of a quarterly academic journal called "Information Management Research", the vice president of Korea Knowledge Innovation Association and the Public Relations Chief of WLIC 2006 Seoul. She is also the vice president of ICSTI(the International Council for Scientific and Technical Information) which is the unique forum for interaction between S&T information organizations all over the world. Her particular areas of interest are knowledge ecology, knowledge management, scholarly communications and collection development. She holds masters and PhD in Library & Information Science from Yonsei University in Korea, and post-master degree from Dominican University in United States. Email: hychoi@kisti.re.kr

Chowdappa, N.
Narayanappa Chowdappa obtained a Post-Graduate Degree in Geo-science and Master's Degree in Library and Information Science from Bangalore University, Bangalore with a distinction. Obtained Doctorate degree for his Thesis on "Organization and Use Patterns of Grey Literature in Engineering Research Institutions" from the University of Mysore, Mysore in Library and Information Science, serving as Chief Librarian at BMS College of Engineering, Bangalore for the last 26 years. Special interest includes promoting the use of grey literature among faculty and researchers in engineering discipline. Responsible for organization of 22 professional and extension programs for Teachers and Librarians in Science and Technology. He served as resource person in library and information science, and Academic Counselor for Indira Gandhi National Open University, New Delhi and presently holds the position of President, AKELPA - All Karnataka Engineering College Library Professionals Association, Bangalore. Areas of professional interest are: Scholarly Communications, Facilitating Research Programs and Reference Service. Email: ncbmsce@yahoo.co.in

Cocciolo, Anthony
Anthony Cocciolo (Ed.D., Columbia University) has extensive experience in digital repositories, educational technology, and GIS, and teaches courses on Social Media, Digital Archives and Digital Libraries. His research interests are in the uses of emerging information and communication technologies (ICTs) to enhance libraries and education. Prior coming to Pratt, he was the Head of Technology for the EdLab and the Gottesman Libraries at Teachers College, Columbia University. He completed his undergraduate degree in Computer Science from the University of California, Riverside. Email: acocciol@pratt.edu

Davis, Marilyn J.
Lynn Davis is an information scientist and account executive for the U.S. Department of Energy's (DOE) Office of Scientific and Technical Information (OSTI) in Oak Ridge, Tennessee. She serves DOE, other government agencies, and other organizations by developing solutions to provide access to DOE scientific and technical information meeting their specific needs. Examples of her successful solutions and collaborations include managing an historic legacy digitization project for the U.S. DOE Geothermal Technologies Program (www.osti.gov/geothermal); developing new media services for the DOE National Science Bowl to lay and build upon a foundation for fostering engagement among students, teacher, parents, community, academic sponsors, and the media around an annual nationwide academic competition that tests students' knowledge in all areas of science (www.osti.gov/nsbjournal); and leveraging the abundance of scientist/researcher contact information in OSTI databases to facilitate peer-to-peer communication, collaboration, and sharing between and among scientists (www.Science-Forums.net). Ms. Davis has a Bachelor of Science from James Madison University and a Master of Science in Information Sciences from the University of Tennessee. DavisL@osti.gov

De Biagi, Luisa
Luisa De Biagi received her degree in Literature and Philosophy at 'La Sapienza' Univ of Rome with a Specialization in 'Archivist-Palaeographer' (Vatican School of Palaeography, Diplomatics and Archivistics at the Vatican Secret Archive) as well as a Specialization Degree in Archivistics, Palaeography and Diplomatics (Archivio di Stato, Rome). Degree of the Vatican School of Library Sciences. De Biagi further holds a Master in 'Business Publishing' (LUISS Management – Rome). She was a member of the SIGLE Working Group and Italian Grey Literature Data-Base Working Group from 2002 and a Member of the CNR Working Group for Cedefop-Refernet Project (Consortium for Professional Education and Training coordinated by ISFOL). Member of the Commettee for Legal Deposit Acquisition at CNR Central Library, and European Association of Health Information and Libraries (EAHIL). Since 2010 is Responsible for the Italian National Referring Centre of Grey Literature at CNR Central Library 'G. Marconi' (Opensigle/Opengrey Network Project) and for the Library Functional Units 'Education and Training' and 'Cultural Activities Management', organizing didactics laboratories for students, professional training courses and teaching in professional trainings for librarians, students and users. Email: luisa.debiagi@cnr.it

Di Cesare, Rosa
Rosa Di Cesare is responsible for the library at the Institute for research on populations and social policies of the National Research Council (CNR). She worked previously at the Central library of CNR where she became involved in research activities in the field of Grey literature (GL) as member of the Technical Committee for the SIGLE database. Her studies have focused on the use of GL in scientific publications and recently on the emerging models of scholarly communication (OA and IR). Email: r.dicesare@irpps.cnr.it

Farace, Dominic
Dominic J. Farace is Director of TextRelease, an Amsterdam based information bureau specializing in grey literature and networked information. He is a native Louisianan and holds two degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked for the past twenty-seven years. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (Swidoc/KNAW), he founded GreyNet, Grey Literature Network Service, in 1993 and has since been responsible for the international GL-Conference Series. In this capacity, he serves as Program and Conference Director as well as managing editor of the conference proceedings. Since 2004, he is a Guest Lecturer on Grey Literature in the Masters Program at the University of Amsterdam; Instructor of Grey Literature via UNO Distance Education, and Editor of TGJ, The Grey Journal.
Email: dominic.farace@textrelease.com

Faye, Cathy
Cathy Faye is the Assistant Director of the Center for the History of Psychology at The University of Akron. She received a Master's degree in psychology from York University in Toronto Canada in 2005 with a specialty in the history of psychology and related sciences. She is currently pursuing a doctoral degree in the same field. She has authored and coauthored articles on the history of psychology as well as articles exploring research topics in contemporary psychology.
Email: cfaye@uakron.edu

Author Information

Frantzen, Jerry
Jerry Frantzen graduated in 1999 from the College of Amsterdam in Library and Information Science. Frantzen is the technical editor of The Grey Journal (TGJ). And, since 1996, he is affiliated with GreyNet, Grey Literature Network Service, as a freelance technical consultant. Email: info@greynet.org

Gelfand, Julia
Julia Gelfand has been engaged with Grey Literature for nearly 20 years and has written and spoken extensively about the role of grey literature as it aligns with contemporary publishing and its role in library collections. She has explored the new fringe of grey literature in virtual worlds, evidence-based practices and now in social media and is interested in how it will be tracked, preserved and archived. She is the Applied Sciences & Engineering Librarian at the University of California, Irvine where she has been a member of the staff for over 30 years and is actively involved in collection development, scholarly communication, and e-science activities at national and international levels. Email: jgelfand@uci.edu

Južnič, Primož
Primož Južnič is an associate professor at the Department of Library and Information Science and Book Studies at Faculty of Arts, University of Ljubljana (Slovenia). His main area of research and interest is bibliometrics, collection management and LIS education. He teaches the following courses: Bibliometrics, Special libraries, and Collection Management. Before starting his university career, he was a heading different special and academic libraries and information/computer centres. He was also working at the European Commission, for three years, as the seconded informatics expert.
Email: primoz.juznic@ff.uni-lj.si

Kearns, Jodi
Jodi Kearns is the Digital Projects Manager at the Center for the History of Psychology at The University of Akron. She earned her PhD in Information Science from the University of North Texas. She also serves as Assistant Professor at Kent State University School of Library and Information Science. Jodi has authored and coauthored articles on the nature of library information, particularly regarding Shannon's Information Theory. Her 2008 book Doing Things with Information, with Brian O'Connor and Richard Anderson, encapsulates recent explorations in these areas. Email: jkearns@uakron.edu

Le Bescond, Isabelle
Isabelle Le Bescond is librarian at the Lille 1 University Central Library, Science and Technology, since 2005. She is currently responsible for developing the digital library IRIS. She has served in different University Libraries (Paris, Strasbourg) since 1994. She studied German language and literature. Email: isabelle.le-bescond@univ-lille1.fr

Lee, Seon-Hee
Seon-Hee Lee is Senior Researcher at the Knowledge Information Center of the Korea Institute of Science and Technology Information. She has a master's degree in Library and Information Studies (MLIS) from the University of California, Los Angeles (1996) and a Master of Art degree in Philosophy from Ewha Women's University (1988). Her research interests include: collection development, grey literature, e-journals, Open Access Journals, and collaborative digital reference services. Email: wisdom@kisti.re.kr

Lin, Anthony
Tony Lin is the Research Librarian for Business at the University of California, Irvine. He holds a MIS from the University of Michigan-Ann Arbor, School of Information. At the University of California, Irvine, he utilizes and actively explores new ways to use emerging library technologies to effectively deliver reference services to both on-campus and distance education users. In addition to his work with emerging technologies, he also works closely with small business owners and entrepreneurs to form business plans, perform competitive analysis, and market research for local and regional businesses. His many interests include Geographic Information Systems, US Government Information, and institutional data repositories. ablin@uci.edu

Lin, Yongtao
Yongtao Lin has been a Health Information Network librarian, working at the Tom Baker Cancer Knowledge Centre since 2008. She provides library services to support health care professionals in their evidence-based practice. The library is also part of a provincial patient-centered education strategy that supports cancer patients and their families. She was a librarian with hospital libraries in rural Nova Scotia for a few years before working with the University of Calgary. Her prior experience being an instructor has led her to integrate education into various aspects of library programs. Yongtao is interested in the impact of grey literature in health care and a strong believer of evidence-based practice. Email: yolin@ucalgary.ca

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

Marzi, Claudia
Claudia Marzi Position: Researcher (CNR Tecnologo) PhD student in Acquisitional and Computational Linguistics, University of Pavia. Laurea degree in Modern Languages (English) at Pisa University in 1998, with the dissertation "The power of words: language creativity in Edgar Allan Poe's narrative". Since 2001 working at the Institute for Computational Linguistics "A. Zampolli" (ILC), National Research Council (CNR), in Pisa. She contributed to many national (funded by Italian Ministry of Research) and international (funded by European Commission) projects. Programme Coordinator of the European Science Foundation Research Networking Programme "The European Network on Word Structure. Cross-disciplinary approaches to understanding word structure in the languages of Europe". Member of board at Institute for Computational Linguistics. Main areas of interest: Document and knowledge management; Second language acquisition; Child language; Computer modelling of the Mental Lexicon.
Email: claudia.marzi@ilc.cnr.it

Mynarz, Jindřich
Jindřich Mynarz has got a bachelor's degree in Library and Information Science at the Institute of Information Studies and Librarianship, Charles University in Prague, and he continues with New media studies accredited at the same university. He works at the Development of electronic services department at the National Technical Library in Prague, Czech Republic. The main focus of his work is on library data and their transformation to more web-compatible data models and their exposing in conformance with linked data principles. Other areas of interest include knowledge organisation systems, particularly in the ways in which they can be used in linked data setting, and in user interfaces that can be built to access library data. Email: jindrich.mynarz@techlib.cz

Pardelli, Gabriella
Gabriella Pardelli was born at Pisa, graduated in Arts in 1980 at the Pisa University, submitting a thesis on the History of Science. Since 1984, researcher at the National Research Council, Institute of Computational Linguistics "Antonio Zampolli" ILC, in Pisa. Head of the Library of the ILC Institute since 1990, responsible for the Archives of the ILC Institute since 2005. Her interests and activity range from studies in grey literature and terminology, with particular regard to the Computational Linguistics and its related disciplines, to the creation of documentary resources for digital libraries in the Humanities. She has participated in many national and international projects including the recent projects:- BIBLOS: Historical, Philosophical and Philological Digital Library of the Italian National Research Council, (funded by CNR); - For digital edition of manuscripts of Ferdinand de Saussure (Research Programs of Relevant National Interest, PRIN - funded by the Ministry of Education, University and Research, MIUR).
Email: gabriella.pardelli@ilc.cnr.it

Pavlov, Leonid P.

Author Information

Leonid P. Pavlov 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

Pejšová, Petra
Petra Pejšová studied information science and librarianship at Charles University. She works as an information specialist in the State technical Library, Czech Republic. Actually she is leading a project Digital Library for Grey Literature – Functional model and pilot.
Email: petra.pejsova@techlib.cz

Prost, Hélène
Hélène Prost is responsible for studies at the Institute of Scientific and Technical Information (INIST-CNRS). The different studies concern the evaluation of collections, document delivery, usage analysis, grey literature and open access to information. Expertise in statistical tools and knowledge in library information science allowed her to participate in various research projects and writing of several publications.
Email: helene.prost@inist.fr

Puccinelli, Roberto
Roberto Puccinelli is currently head of Section I at CNR's "Information System Office" and he's been working for CNR since 2001. He has previously worked in the private sector as system and network engineer. As adjunct professor, he has held courses for the First University of Rome "La Sapienza" ("Operating Systems II") and for the Third University of Rome ("Programming and Computing Laboratory"). He graduated in Electronic Engineering at the University of Rome "La Sapienza" and holds a master cum laude in Enterprise Engineering from the University of Rome "Tor Vergata". In the past he has worked in several research projects in the field of Grid technologies both at the national and international level (executive manager of Work Package 11 within the DataGrid project – V Framework Programme, et al.). He's currently involved in the design and development of CNR's information system. In particular, he coordinates projects for the development of application systems and is responsible for the design and implementation of CNR's data warehouse. He is also responsible for CNR's Local Registration Authority management. He's currently involved in projects regarding the design and development of research product open archives and persistent identifier registers/resolvers. He is author of several articles in the fields of Grid technologies, Autonomic Computing, Software Engineering, Open Archives and Persistent Identifiers. Email: roberto.puccinelli@cnr.it

Rabina, Debbie
Debbie Rabina (Ph.D., Rutgers University) teaches Information Policy, International Information Resources and Academic Libraries and Scholarly Communication. Her research is informed by the belief that societies that exercise a policy of open dissemination of and access to information are those that allow its citizenry to achieve personal and societal goals. The aim of her research is to examine questions of how to achieve information policies and practices that will allow information to serve as an engine of renewal and growth. Email: drabina@pratt.edu

Ramasesh, C.P.
C.P. Ramasesh obtained Master's Degree in Philosophy from the University of Mysore in 1976 with specialization in Vedanta Philosophy. He also obtained a Master's Degree (1978) and Doctorate Degree (1989) in Library and Information Science from the University of Mysore. Ramasesh is the recipient of Dr. S.R. Ranganathan Memorial Gold Medal for securing first rank in the Master's Degree. He served as professional librarian for three decades and is responsible for the organization of several extension programmes. Worked as the Deputy Director of Centre for Information Science and Technology (CIST), Mysore and coordinated for the implementation of Quality Procedures under ISO 9001 Standards of Quality Management System. Compiled Quality Manual, delivered special lectures and served as Quality Auditor/ Performance Auditor at CIST. Presently serving as University Librarian of the University of Mysore and supervising research programmes in the field of 1) Grey Literature 2) Institutional Repositories of Heritage Collection and 2) Use Pattern of Online Journals. Email: cpramasesh@gmail.com

Ranjit, Kathryn M.E.
Kathryn M.E. Ranjit is a librarian at the Peter Lougheed Knowledge Centre, Health Information Network. She received her MLIS and BMSc from the University of Western Ontario. She enjoys the challenges of providing quality information service to health care professionals, patients and families and seeks to implement evidence-based decision making in her practice. Her current areas of research include web development to facilitate the online access to health information and the use of social media for online reference. She is also an advocate for the impact of health literacy on health care.
Email: kathryn.ranjit@ucalgary.ca

Ricci, Marta
Marta Ricci has an undergraduate degree in Humanities and a Master degree in Library Science from the University of Rome "Tor Vergata" with a thesis on bibliometric tools and citation analysis. She had an internship experience in the library of the Italian Prime Minister's office (Chigi's Library), where she was responsible for the Inventory of part of the library collections. Currently she is collaborating with the library of the Institute for Research on Population and Social Policies of the Italian National Research Council (CNR), in the field of GL.
Email: biblio.irpps@irpps.cnr.it

Ruggieri, Roberta
Roberta Ruggieri is librarian at the Senate of the Republic where she is responsible for the supervision of a digitalization project on Senate parliamentary print documents for the I to X Legislature. Her activity in managing digitalization project also includes document addition and classification in the electronic Senate catalogue. From 2004 she has been collaborating with the Institute for research on populations and social policies of CNR in research activities related to the field of grey literature and Institutional repositories.

Saccone, Massimiliano
Massimiliano Saccone graduated in Letters and specialized in Library sciences at the "Sapienza" University of Rome. He works as librarian at the Central Library of CNR. He is the responsible for Legal deposit and Open Access activities. He has participated actively in several national and international projects on Information and Knowledge Management (Digital Preservation Europe – DPE, Italian Network of National Bibliography Number – NBN, etc.). His main interests are in scholarly communications, open access, digital preservation, metadata quality control, information system interoperability.

Sassi, Manuela
Manuela Sassi 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
Joachim Schöpfel is head of the department of information and library sciences at the Charles de Gaulle University of Lille 3 and researcher at the GERiiCO laboratory. He is interested in scientific information, academic publishing, open repositories, grey literature and usage statistics. He is member of GreyNet and EuroCRIS.
Email: joachim.schopfel@univ-lille3.fr

Sesink, Laurents
Laurents Sesink (1965) studied history at the University of Utrecht and historical information science at the University of Leiden. Laurents worked during 1995-2002 at the Netherlands Institute for Scientific Information Services on different large scale Research & Development projects as senior specialist digitisation services, technical scientific programmer and coordinator of a software development group. During the period 2003-2006 he was worked as a senior policy advisor at the Dutch Academy of Sciences within the subject of scientific and administrative information. Laurents is since januari 2006 employed as an information scientist at DANS and his focus is on fundamental issues regarding accessibility to digital scientific data and how theoretical

Author Information

solutions can be practically implemented in the R&D projects at DANS. Email: laurents.sesink@dans.knaw.nl

Shyamala, V.R.
V.R. Shyamala obtained her Master's Degree in Zoology from the University of Mysore. She has been trained in IT applications at the Centre for Information Science and Technology, University of Mysore. She has served as faculty for almost two decades. Presently, working as the Principal at the Government P.U.College, M.G.Koppal, Mysore. Shyamala has managed library services while she was serving as faculty at various colleges. She has inclination to organize and manage library resources on modern lines. She has delivered special lectures on e-resources and application of ICT to educational institutions. Her areas of special interests are: Reproductive Biology, Philosophy, Professional Ethics and Library Collection Development.
Email: shyamalasesh@gmail.com

Stock, Christiane
Christiane Stock is the head of the Monographs and Grey Literature service at INIST-CNRS and manager of OpenGrey (European grey literature repository) and THESA (doctoral theses prepared in the French Grandes Ecoles). Her experience with grey literature began as a member of the Technical Committee for the SIGLE database from 1993 to 2005, and continued with the set up of the national agency for ISRN (International Standard Report Number). She then focused on making grey literature available in open access repositories with mémSIC (master's theses in information sciences) and LARA (French reports). She was a member of the AFNOR expert group who prepared the recommended metadata scheme for French electronic theses (TEF). Email: christiane.stock@inist.fr

Trufelli, Luciana
Luciana Trufelli works as technical at the Central Library of CNR. She is responsible for Studies Office. She promotes and collaborates on many national and international projects on Information and Knowledge Management. She currently works in activities under open access, metadata quality control, and information system interoperability. She also has extensive experience in the areas of statistics and performance indicators of the R&D. Her main interests are in quality systems and institutional communication.

Usha Devi, L.
L. Usha Devi obtained her Bachelor's Degree in Science and Master's Degree in Library and Information Science from the University Of Mysore, Mysore, She has been serving as Assistant University Librarian at Bangalore University, Jnana Bharathi, Bangalore for the last 28 years. Ushadevi has also secured Master of Philosophy Degree in Library and Information Science. She has contributed 12 research papers at various National and International seminars and conferences .She is a life member of professional associations/ organisations: Indian Library Association, Indian Association for Special Libraries and Information Centres, Indian Association for Teachers in Library and Information Science and Karnataka Library Association. Her areas of research interest are Library User Study and Organisation and Use of Grey Literature.

Vaska, Marcus
Marcus Vaska is librarian at the Health Information Network Calgary (HINC), Holy Cross Site. He was a former librarian for the Physician Learning Program (PLP), a collaborative initiative between the Universities of Alberta and Calgary, funded via an Alberta Medical Association (AMA) trilateral agreement, where he was responsible for assisting physicians in their research, and addressing their perceived and unperceived learning needs. Prior to that position, he was a librarian at the University of Calgary's Health Sciences Library. Marcus' current interests focus on educational techniques aimed at creating greater awareness and thereby bringing grey literature to the forefront in the medical community. Email: mmvaska@ucalgary.ca

Vowell, Lance
Lance Vowell is currently employed by Information International Associates and assigned to the U.S. Department of Energy's (DOE) Office of Scientific and Technical Information (OSTI) contract as the Contract/Program Manager. In this capacity, Mr. Vowell is responsible for all phases of IT support ranging from requirements gathering, deployment and product maintenance. Mr. Vowell has been instrumental in the development of core application framework that has markedly reduced application development time and maintenance costs at OSTI. Vowell has served as a technical resource and/or Principal Investigator on six Small Business Innovation Research (SBIR) grants for the DOE on topics including "Contextual Spell Checking in the Digital Library Environment", "Mobile Device Applications in the Digital Library Environment" and "Interactive Peer to Peer Scientific Communication in the Digital Library Environment"; two Small Business Technology Transfer (STTR) grants for DOE and the U.S. Department of Defense; and is currently serving as a technical advisor on an DOE Experimental Program to Stimulate Competitive Research (EPSCoR) grant, "Exploiting the Use of Social Networking to Facilitate Collaboration in the Scientific Community". Email: VowellL@osti.gov

Willis, Lynn
Lynn Willis is the Content Development Manager for PsycINFO, a suite of databases produced by the American Psychological Association. Her interest in gray literature dates back to when she was chasing stories on government technology as a science writer. Currently she sits on the Board of Directors of NFAIS. Email: lwillis@apa.org

GreyNet Members less 20%

The Grey Circuit
From Social Networking to Wealth Creation

Library of Congress, Washington D.C. USA 5-6 December 2011

Publication Order Form

THIRTEENTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

Forthcoming February 2012	*No. of Copies*	x	*Amount in Euros*	Subtotal
GL13 CONFERENCE PROCEEDINGS – Printed Edition ISBN 978-90-77484-17-3 ISSN 1386-2316 *Postage and Handling* ***excluded***[*)]		x	99.00 =	€
GL13 CONFERENCE PROCEEDINGS - CD-Rom Edition ISBN 978-90-77484-17-3 ISSN 1386-2316 *Postage and Handling* ***included***		x	99.00 =	€
GL13 Conference Proceedings – Online Edition ISBN 978-90-77484-17-3 ISSN 2211-7199 ***Password Protected Access***		x	99.00 =	€
	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 with an invoice marked paid.

☐ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to "GL13 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: CREDIT CARD TRANSACTIONS CAN BE AUTHORIZED BY PHONE, FAX, OR POSTAL SERVICES. EMAIL IS NOT AUTHORIZED.

TextRelease
www.textrelease.com

GL13 Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
T/F +31-(0) 20-331.2420 Email: info@textrelease.com

Index to Authors