Twelfth International Conference on Grey Literature

National Technical Library, Prague, Czech Republic, 6-7 December 2010

GL 12
pantone red 032c

Transparency in Grey Literature

Grey Tech Approaches to High Tech Issues

Conference Proceedings

ISSN 1386-2316

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

Host and Sponsors

Centrum Vedecko-Technických Informácií SR

British Library

The Law Library of Congress

New York Academy of Medicine 1847

cnrs inist

IIa Information International Associates

FLICC FEDLINK

Enabling Science and Innovation
ICSTI
International Council for Scientific and Technical Information

GreyNet
Grey Literature Network Service

EBSCO
INFORMATION SERVICES

GL12 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

GL12 Conference Proceedings
Twelfth International Conference on Grey Literature : Transparency in Grey Literature, Grey Tech Approaches to High Tech Issues, 6-7 December 2010 in the National Technical Library, Prague, Czech Republic / ed. by D.J. Farace and J. Frantzen ; GreyNet, Grey Literature Network Service. – Amsterdam : TextRelease, February 2011. – 142 p. ; 30 cm. – Author Index. – (GL-conference series, ISSN 1386-2316 ; No. 12)
The British Library, FLICC-FEDLINK, INIST-CNRS, NTK, and NYAM are corporate authors and associate members of GreyNet. These conference proceedings contain the full text of sixteen papers presented during the two days of plenary and parallel 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, and sponsored advertisements.

ISBN: 978-90-77484-16-6

Foreword

Transparency in Grey Literature

Grey Tech Approaches to High Tech Issues

At the onset of twenty-ten, Grey Literature emerged into the public arena after more than a quarter century in the corridors of libraries and in workplaces and meeting rooms of information practitioners and professionals. Grey Literature is now a topic of news in the world media. Coverage in magazines and newspapers *e.g.* Nature, New Scientist, The Economist, the Guardian, etc. carrying articles on the IPCC use/misuse of grey literature is current and in-depth. For those following these news threads, much of the publicity is less than complimentary. And, the grey literature community has not been hesitant in their response via blogs, listservs, distribution lists, etcetera.

GL12 provided the international grey literature community with an opportunity to bundle its efforts in order to address issues that stand at the core of grey literature and which have come under fire in the public media. One thing is certain, now that grey literature has entered the mainstream press, it will not simply disappear. It is now up to the corporate authors and publishers of grey literature as well as organizations processing and distributing it in both print and electronic formats to address misconceptions and unknowns about this field of information science. The Twelfth International Conference on Grey Literature offered a global forum for stakeholders in government, academics, business and industry to come together on issues dealing with transparency in grey literature and the almost seamless processes of research, authorship, publication, indexing, as well as uses and applications to which it is exposed in knowledge based communities. Many of these processes are the same faced by commercial publishing, where only the differences lie in *grey tech approaches to high tech issues*.

Dr. Dominic J. Farace Amsterdam,
Grey Literature Network Service February 2011

GL12 Conference Sponsors

NTK, Czech Republic
National Technical Library

BL, United Kingdom
The British Library

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

FLICC/FEDLINK, USA
Library of Congress

EBSCO, USA
EBSCO Information Services

GL12 Conference Sponsors

ICSTI, France
International Council for Scientific and Technical Information|

IIA, USA
Information International Associates, Inc.

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

LCC, USA
The Law Library of Congress

NYAM, USA
The New York Academy of Medicine

GL12 Program Committee

Martin Svoboda and Petra Pejsova *(Co-Chair)*
National Technical Library, NTK
Czech Republic

Elizabeth Newbold
BL, British Library
United Kingdom

Daniela Luzi
CNR, National Research Council
Italy

Dominic J. Farace
GreyNet, Grey Literature Network Service
Netherlands

Bonnie Carroll
IIA, Information International Associates, Inc.
USA

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

Roberta Shaffer
LLC, Law Library of Congress
USA

Janice Kaplan
NYAM, The New York Academy of Medicine
USA

Joachim Schöpfel
University of Lille 3
France

Table of Contents

Program

Advertisements

Appendices

SESSION ONE – REDEFINING GREY LITERATURE

Chair, *Christiane Stock, Institute for Scientific and Technical Information, INIST/CNRS, France*

SESSION TWO – NEW STAKEHOLDERS IN GREY LITERATURE

Chair, *Elizabeth Newbold, The British Library, United Kingdom*

SESSION THREE – STANDARDIZATION IN GREY LITERATURE

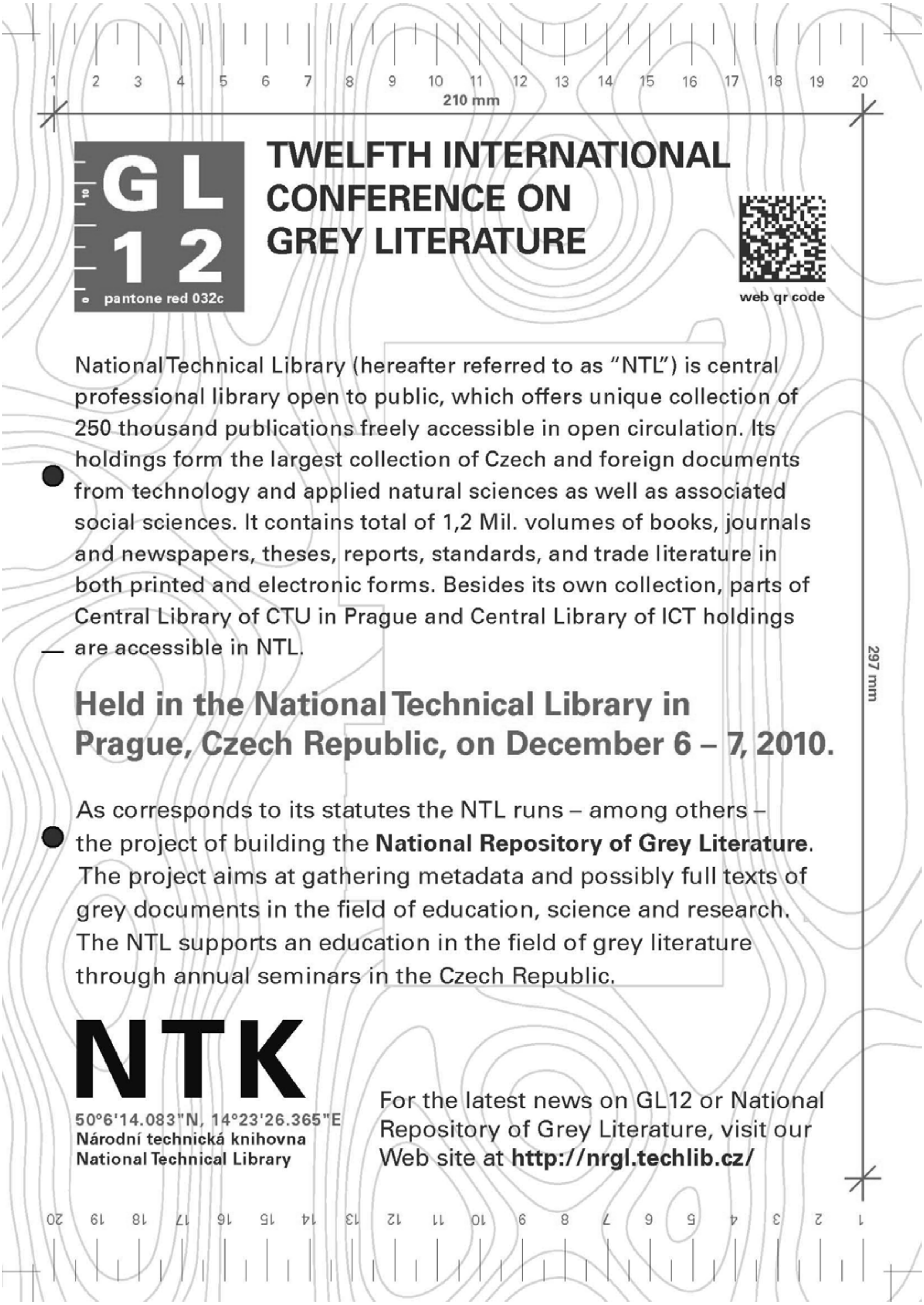
Chair, *Janie Kaplan, New York Academy of Medicine, United States*

SESSION FOUR – NEW FRONTIERS IN GREY LITERATURE

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

PARALLEL SESSION

Chair, *Petra Pejšová, National Technical Library, Czech Republic*


210 mm
297 mm
GL 12
pantone red 032c
web qr code

TWELFTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

National Technical Library (hereafter referred to as "NTL") is central professional library open to public, which offers unique collection of 250 thousand publications freely accessible in open circulation. Its holdings form the largest collection of Czech and foreign documents from technology and applied natural sciences as well as associated social sciences. It contains total of 1,2 Mil. volumes of books, journals and newspapers, theses, reports, standards, and trade literature in both printed and electronic forms. Besides its own collection, parts of Central Library of CTU in Prague and Central Library of ICT holdings are accessible in NTL.

Held in the National Technical Library in Prague, Czech Republic, on December 6 – 7, 2010.

As corresponds to its statutes the NTL runs – among others – the project of building the **National Repository of Grey Literature**. The project aims at gathering metadata and possibly full texts of grey documents in the field of education, science and research. The NTL supports an education in the field of grey literature through annual seminars in the Czech Republic.

NTK

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

For the latest news on GL12 or National Repository of Grey Literature, visit our Web site at **http://nrgl.techlib.cz/**

Towards a Prague Definition of Grey Literature

Joachim Schöpfel
Charles de Gaulle University Lille 3

Abstract
The most common definition of grey literature, the so-called 'Luxembourg definition', was discussed and approved during the 3rd International Conference on Grey Literature in 1997. In 2004, at the 6th International Conference on Grey literature in New York City, a postscript was added. The main characteristic of this definition is its economic perspective on grey literature, based on business, publishing and distribution models of the disappearing Gutenberg galaxy. With the changing research environment and new channels of scientific communication, it becomes clear that grey literature needs a new conceptual framework.

Research method: Our project applies a two-step-methodology: (1) A state of the art of terminology and definitions of the last two decades, based on contributions to the GL conference series (1993-2008) and on original articles published in *The Grey Journal* (2005-2010). (2) An exploratory survey with a sample of scientists, publishing and LIS professionals to assess attitudes towards of the New York definition and to gather elements for a new definition.

Results: Based on the state of the art and the survey data, we make 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." The attributes and challenges are discussed.

1. Introduction

The concept of grey literature is historical. Some decades ago the term grey literature did not exist as a category although what is considered grey today was among the extant literature. When Butterworths published the first edition of Charles P. Auger's landmark work on grey literature in 1975, paradoxically neither the summary nor the index mentioned this term. The book was just about reports literature (Auger, 1975).

Despite the absence of a label, Auger described the nature of this "vast body of documents" in a way that would later characterize grey literature, referring to its "continuing increasing quantity", the "difficulty it presents to the librarian", its ambiguity between temporary character and durability, and its growing impact on scientific production. He also pointed out the "number of advantages over other means of dissemination, including greater speed, greater flexibility and the opportunity to go into considerable detail if necessary". For Auger, reports were a "half-published" communication medium with a "complex interrelationship (to) scientific journals".

The description sounds familiar. "Semi-published literature" is a connotation of grey literature (Keenan, 1996). But it reminds, too, that one can speak about reports without a generic concept. Auger promoted the term of "grey literature" only in the 2nd edition of his book (Auger, 1989). Since then, the meaning of "GL" remained a challenge to scientists and librarians. Does "GL" make sense? Is it necessary? Is it (still) helpful for the study and processing of scientific literature? Or using a variation on the famous quote from Dorothy L. Sayers, will it "run away (…) like cows if you look (it) in the face hard enough"?

There are several definitions of grey literature, the most common being the so-called "Luxembourg definition," which was discussed and approved during the Third International Conference on Grey Literature in 1997: "[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." In 2004, at the 6th conference in New York, a postscript was added for purposes of clarification "...not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body" (see Schöpfel & Farace, 2010).

The Luxemburg definition accentuates the supply side of grey literature, e.g., its production and publication both in print and electronic formats. It calls attention to the question of dissemination, the difficulty to identify and access documents described as ephemeral, non-conventional or underground.

Material that "may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents" (U.S. Interagency Gray Literature Working Group): this concept meets Mackenzie Owen's observation that "grey does not imply any qualification (but) is merely a characterization of the distribution mode" (1997).

Now, Internet transforms the whole value chain of publishing. The Web offers new tools and channels for producing, disseminating and assessing scientific literature. Author and reader, producer and consumer change their information behaviour. We definitely left the Gutenberg era. So what about the definition of grey literature? Is it still empirically sound?

Our study returns to the roots of grey literature and provides insight in past definitions and present opinions. Based on a critical discussion of this evidence, a new definition ("Prague definition") is suggested that may stimulate future research and theoretical work on this "vast body of documents".

2. Methodology

The study applies a two-step-methodology and combines a review of literature (state of the art) with an empirical survey.

2.1. State of the art: content analysis of GL corpus

The state of the art focuses on conceptual studies and definitions of the last decades, e.g., contributions to the GL conference series (1993-2008) and original articles published in *The Grey Journal* (2005-2010).

The corpus consists of 32 documents selected from 219 GL conference communications published on the *OpenSIGLE* website[i] (sampling = 15%), through a content analysis of titles, abstracts and full texts (Fig. 1).

Conference	**GL1**	**GL2**	**GL3**	**GL4**	**GL5**	**GL6**	**GL7**	**GL8**	**GL9**	**GL10**
Year	*1993*	*1995*	*1997*	*1999*	*2003*	*2004*	*2005*	*2006*	*2007*	*2008*
Total nb	27	21	28	26	18	24	27	16	17	15
Selection	4	4	3	6	5	4	0	3	1	2

Figure 1: Corpus of GL communications

The selection criterion was substantial debate on (and not only recall of) definitions and concepts of grey literature.

Some of these communications were also published in *The Grey Journal* (TGJ). For this reason and to avoid double entries, the selection of TGJ articles was limited to original contributions. The selection criterion ("substantial debate") was the same as for the GL conferences.

Between 2005 and 2010, *The Grey Journal* published 101 articles referenced in the online RefDoc database[ii]. From these articles, we selected three original articles (not published in GL proceedings) with substantial debate on grey literature (sampling= 3%) and added them to our GL corpus (Fig. 2).

Volume	**Vol 1**	**Vol 2**	**Vol 3**	**Vol 4**	**Vol 5**
Year	*2005*	*2006*	*2007*	*2008*	*2009*
Selection	1	2	0	0	0

Figure 2: Corpus of TGJ articles

Taken together, the corpus for the state of the art is composed of 35 documents published between 1993 and 2008, corresponding to 11% of the papers in GL conference series and TGJ.

The content of each communication or article was indexed with main topics of GL definitions (production, dissemination etc.) and traditional functions of scientific publishing (registration, preservation etc.).

2.2. Empirical evidence: online survey

A survey on grey literature adds qualitative and exploratory data to this state of the art, especially attitudes towards the New York definition and elements for a new definition.

The survey was carried out in October 2010. The questionnaire was made available online[iii]. The survey population included 1390 information specialists and scientists from GreyNet's distribution list. Promotion was also done on Twitter and through the social networks LinkedIn, Viadeo and Facebook.

The questionnaire contains eight questions on functions, elements of the current definition, statements and prognostics on grey literature (see annexe B). Only one part of the results is analysed and discussed here.

3. Results

3.1. Content analysis of papers on grey literature

"Grey literature is difficult to define" (Wood & Smith, 1993). Studies on grey literature often begin by trying to help understand grey literature, review literature and sometimes even suggest a new definition. Our corpus contains at least four contributions that provide deeper insight in terminology and conceptualisation of grey literature (Di Cesare & Sala, 1995; McDermott, 1995; Gokhale, 1997; Nahotko, 2007).

In the aftermath of the 1997 conference, most authors cite the Luxemburg definition as reference although it was never meant to be a final definition but rather to instigate and promote research. More recent studies add the New York postscript while the earlier US Interagency Working Group definition appears to be more or less elapsed.

3.1.1. Essential attributes of the definition of grey literature

Which are the main features mentioned in the sample corpus? Two-thirds of the studies insist on **dissemination** as the central characteristic of grey literature, e.g. the unconventional or unusual mode of distribution through non-commercial channels (see Figure 3).

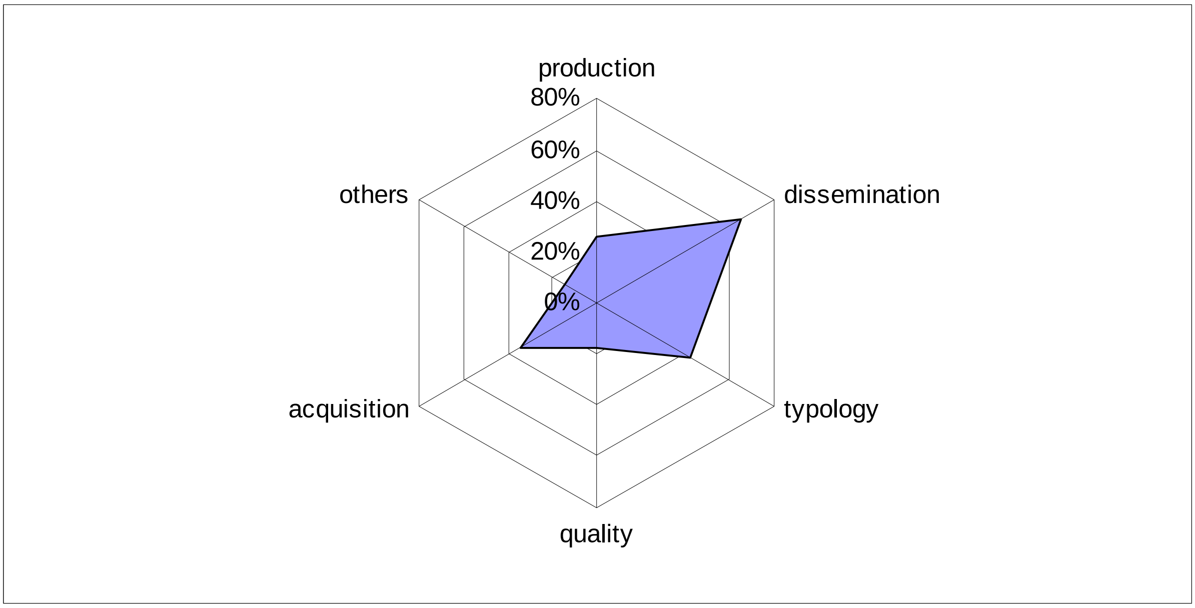


Figure 3: Main topics of GL definitions in corpus

These authors link grey literature and the information market. For instance, Owen (1997) defines grey literature "loosely (…) as information distributed directly by its creator". Gelfand (1999) underlines its "alternative way of distribution", and Boekhorst et al. (2004) stress the "dichotomy grey vs. commercial" as a "cognitive tool" for understanding this kind of scientific literature.

Sometimes, another attribute is added: the fact that grey documents are most often disseminated in limited (small) numbers (Aceti et al., 1999; Nahotko, 2007).

Closely related to this economic definition are papers that focus on the supply side (**production**). For instance, de Blaaij (2003) considers grey literature as "information (largely) produced in the public domain and financed with public money". Ten years earlier, Chillag (1993) distinguished between publications and documents: "In theory, and generally speaking, the former are not grey literature at all". Following Chillag, reports become "white" when collected and sold; he considered documents with different versions, working papers, documents that do not pass through any registry system as "black hole material". At the same time, Cotter & Carroll (1993) stated that grey literature is "not published by established (commercial) publishers", anticipating the Luxemburg and New York definition.

About 40% studies adopt a **typological approach**. In such a definition the operative issue is which type of document belongs to grey literature, which doesn't? Librarians mostly agree that theses and dissertations, conference proceedings, reports and working papers are grey. But what about patents and preprints, blogs, datasets, and tweets? Grey literature "embraces such things as non-conventional literature, archival material, fugitive material, non-book material and unpublished documents" (Kufa, 1993). Luzi (1995), Luzi et al. (2003) and Ranger (2004) worked on new forms of scientific information, such as electronic conferences, protocols, websites or digital datasets. Stock & Schöpfel (2008) evaluated the presence of more traditional items – theses, reports, working papers etc. – in open archives. Sulouff et al. (2005) provide a cross-disciplinary comparison of different types of grey literature associated with academic departments and disciplines, derived from survey data.

The specific mode and problem of **acquisition** have been used as a conceptual feature of GL by one third of the studies. McDermott (1995): "You know you have grey literature when you can't place a standing order for it". Nahotko (2007) expresses the prevailing opinion: "They are difficult to acquire in libraries". Does Internet change the situation? Following Natarajan (2006), it doesn't: "GL, also known as the grey or hidden web, the information that is not searchable or accessible through conventional search engines or subject directories".

Fewer authors raise the question of **quality**, and they do so in a controversial way. Erwin (2006) observes, "quality (of grey literature) continues to be suspect even among researchers" and is not surprised that "because of the range of quality in grey literature (…) grey literature continues to be absent from most formal academic collection development policies". In contrast, Wessels (1997) argues that "much grey literature is published by prestigious organizations whose names are a guarantee for quality" and puts forward its uniqueness[iv].

In the margins of GL definitions, we find some interesting observations that may be helpful for future research:

Intellectual property: Following de Blaiij (1999), grey literature may improve "the sharing of information in the public domain" because its legal status is different from commercial publishing. Cornish (1999) asks for "some easily recognised system internationally (…) to make it clear to users (…) what the owner of copyright in grey literature is willing to permit without seeking permission". Pavlov (2003) describes "(how) to introduce the scientific results documented in grey literature into the legal space of intellectual property and to monitor the processes of their commercialization and rights transfer".

Current Research Information Systems: Jeffery & Asserson (2006) suggest a definition of grey literature as intelligent and hyperactive "grey objects" in a CRIS environment – "they get a life" through metadata and associated document management software.

Open source: Crowe & Davidson (2008) place grey literature into the intersection "of open source and intelligence" and describe the way how information that is openly available and is lawfully obtained may become classified and become a source of intelligence.

3.1.2. Functions of grey literature

Authors like Nahotko (2007) provide a rich and detailed description of different kinds of grey literature. But why does grey literature exist?

Only a small number of studies deal with the question of which sort of needs GL does or should satisfy. We indexed the corpus following Oldenburg's historical description of a scientific journal's main functions. The result is not really surprising (Figure 4).

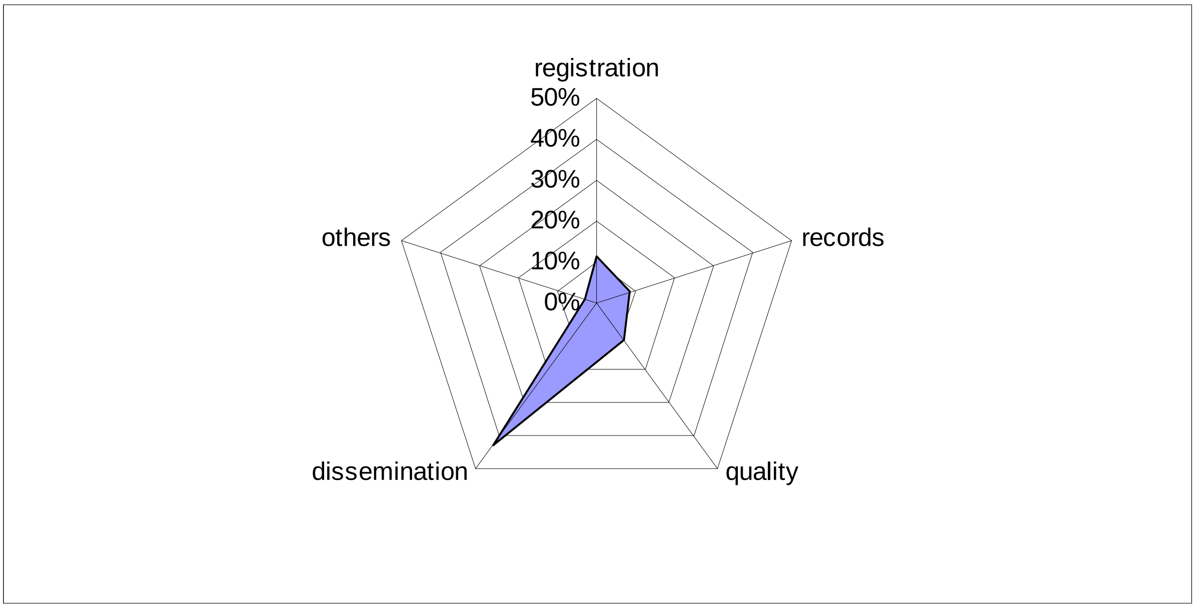


Figure 4: Main functions of GL in corpus

Most of the papers addressing functional aspects of GL highlight its role for **dissemination** of information, research results, etc. Obviously, other traditional functions of scientific publishing are considered as less important, e.g., registration to establish ownership and priority and to clearly date-stamp the work, archiving to provide a permanent record of the work, or certification to have the quality of the research acknowledged by others, through peer review etc.

Kufa (1993) identified an interesting specific role for grey literature in African countries insofar as it may be more relevant to local needs and conditions[v]. Following Kufa, non-commercial publishing may be useful to communicate and access locally produced materials that are not published elsewhere.

3.2. Survey results

The questionnaire was completed by 108 information specialists (70%), scientists and/or scholars (24%), students and others (6%) corresponding to a total average response rate of 7,8%. 97 individuals completed the whole questionnaire.

3.2.1. Attitudes towards the current definition

What do experts think about the current definition of grey literature, e.g., the New York definition? Is it still useful or not? Should it be revised? The questionnaire suggested some current opinions expressed in papers or conferences. The answer may seem paradoxical (Figure 5).

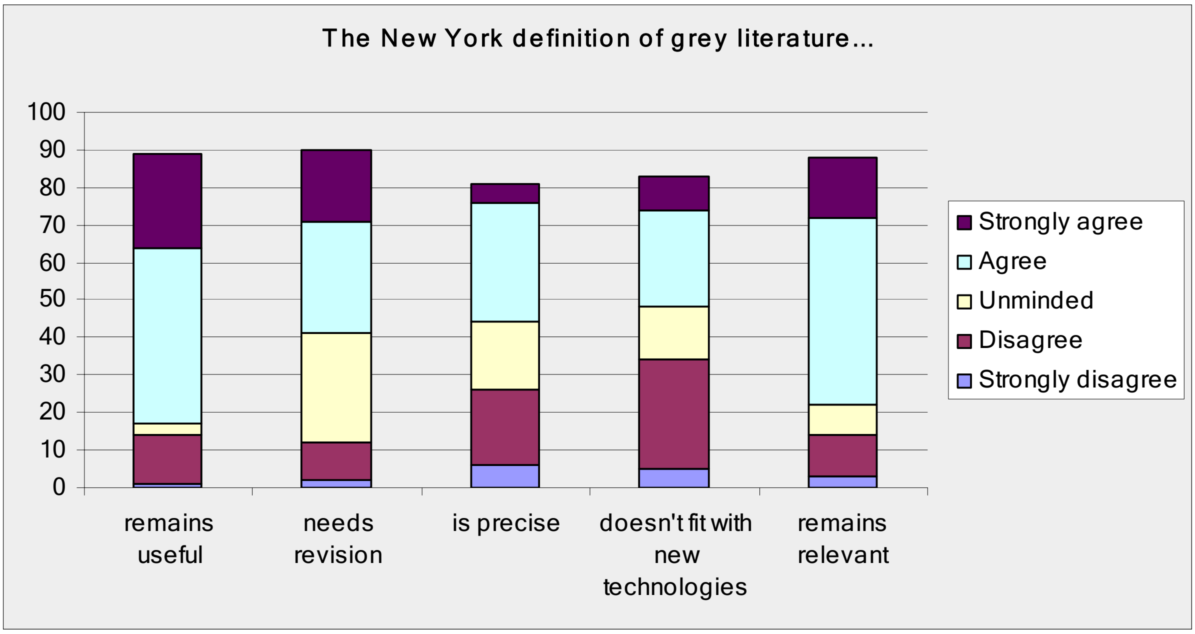


Figure 5: Opinions on the current NY definition

The overall sentiment on the New York definition seems positive. More than 60% responses say that it remains useful and relevant. However, only one third of the people believe that the current definition is precise and fits with new technologies, and only 11% think that it doesn't need revision. Taken together, the common attitude looks like "revision yes, but not abandonment of the current definition" – e.g., continuity and change but not desertion.

3.2.2. Elements for a revised definition

Asked for their opinion on the essential attributes found in papers and described above, people answered in an astonishing but consistent way (Figure 6).

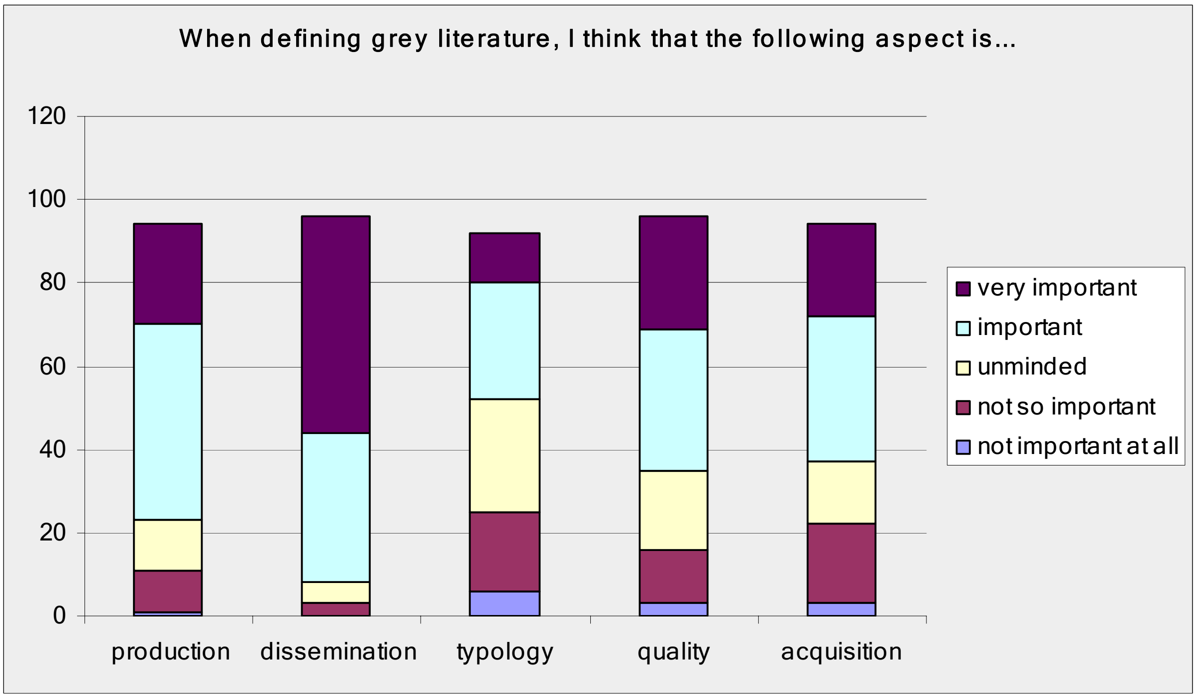


Figure 6: Opinions on essential attributes of a GL definition

Obviously, all suggested items are evaluated as more or less important, and there is no attribute considered as not important. Again, this seems consistent with the overall sentiment on the New York definition – "continuity, not rupture".

Still, a comparison between the items reveals significant differences (Figure 7).

	important	not important
Dissemination	81%	3%
Production	66%	10%
Quality	56%	15%
Acquisition	53%	20%
Typology	37%	23%

Figure 7: Ranking of opinions on essential attributes of grey literature

The key attribute of grey literature is, at least in this sample, without any doubt the specific mode of dissemination, followed by production. Still important but on a lower level are features related to quality issues and (difficulties of) acquisition.

The real surprise is that the typological approach seems no longer an appropriate way of defining grey literature: only 37% rated "typology" as important or very important for the future definition of GL, while 23% think the contrary.

Figure 8 reveals the difference with the content analysis described above in the state of the art (see Figure 3).

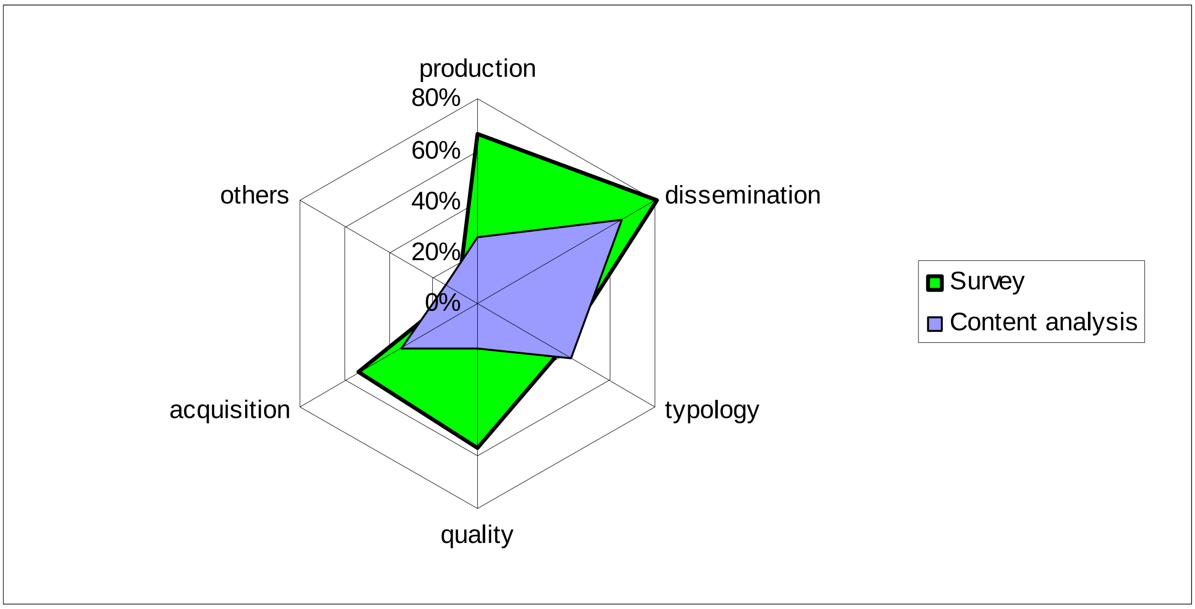


Figure 8: Main topics of GL definitions in corpus (content analysis) and survey

Figure 8 shows clearly that the economic approach to grey literature still prevails, e.g., professionals, scientists and scholars highlight dissemination and production as essential attributes when defining grey literature. But the results reveal, too, that for more than the half of respondents quality and acquisition issues should be taken into consideration for a revision.

3.2.3. Perceived functions

The survey data are not fundamentally different from the content analysis. Again, nearly 80% emphasize "dissemination" as the main function of grey literature (Figure 9).

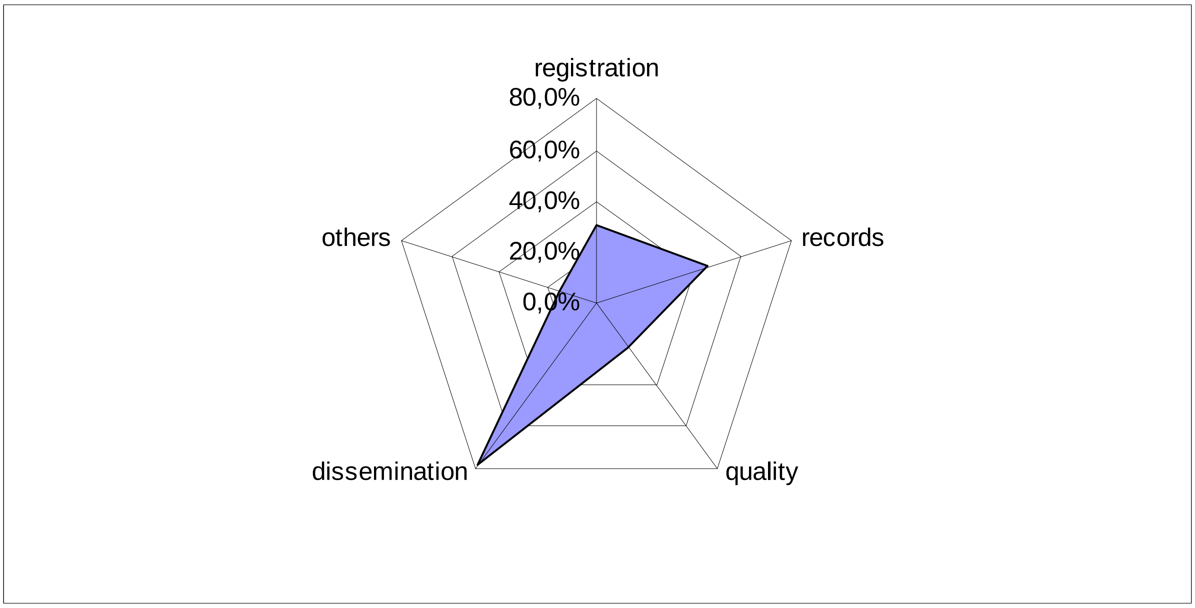


Figure 9: Main functions of GL in survey

The importance of this function is confirmed by open comments like the following:

> *To provide information regarding specific activities to funding agencies.*
> *Provide data/information for research - Research projects of the organisation.*
> *To create awareness and open access to material that has previously been inaccessible.*
> *Be aware prior to article or book publication new trends, new discoveries etc.*
> *To present the results of work conducted and document it in written form.*

Yet, other functions of academic publishing are pointed out, in particular "archiving" to provide a permanent record of the work (46%) and to a lesser extent, registration to establish ownership and priority (31%).

In comparison, the acknowledgment of quality (certification) seems less important (22%) even if grey literature may play a specific role, such as serving for validation at the end of a research project.

3.2.4. A world without grey literature?

"Can you imagine a world without grey literature?" To this question, most respondents (87%) answered with "No". Why? Because grey literature is everywhere, because it is important for research, because it contains valuable information, because it is "non-mainstream" and part of freedom of information production and dispersal, because of the evolution of technology and communication, but also because grey literature corresponds to specific needs. Some citations:

- The frontier of research is grey. The historical record is white.
- Processes of publishing are becoming more and more unofficial.
- Information is the prerequisite for the development. Grey literature is the first information.
- Because grey literature is the most rapid way to disseminate information.
- It is an established method for the production and dissemination of information.
- If a standard, commercial publication can be compared to a "main course" then grey literature accounts for the appetizer, dessert and setting...

Again, these answers enumerate the main attributes and functions of grey literature. One respondent replied by asking:

- Can you imagine a world with only white literature?

For this expert, a world with only one form of information and one mode of communication – the commercial one – is neither imaginable nor desirable. The potential of grey literature for information, culture and freedom is mentioned by another respondent:

- To make this world a better place we need at least ten more Wikileaks and addition to that quality control.

I am not sure if Wikileaks is a good example for grey literature but it is worthy of note that this last argument relates free and not-for-profit information to quality, e.g., introduces a condition we already mentioned above.

4. Discussion

A good definition should specify the necessary and sufficient conditions for an item being part of grey literature. It should offer an if-and-only-if condition for when an item satisfies the definition.

Survey and state of the art indicate that the current definition of grey literature is considered as useful and relevant. It seems globally accepted as a kind of benchmark, reporting actual usage and meaning within the community of grey literature.

So why at the same time our study reveals, too, a general conviction that this definition needs revision and that there are some problems waiting for solution?

The reason is that the definition suffers from some more or less obvious fallacies. We shall describe these problems briefly and then discuss some aspects that might be valuable as essential attributes for a revised definition of grey literature.

4.1. Fallacies of the New York definition

The comparison of papers and survey results points out four problems with the current definition.

The New York definition is over-broad. It does not allow for a clear distinction between grey literature and other "objects", and it applies to items that are not part of the extension of the definition. Too many papers based on the common definition fail to differentiate between grey and other literature and/or include "stuff" that may be grey but clearly is NOT literature.[vi]

The New York definition lacks essential attributes of grey literature. The current definition highlights the specific economic aspect of grey literature: the mode of dissemination and production. It fails, however, to introduce other attributes that, since Auger's study in 1975, have been essential to the understanding of grey literature, such as the character of document, intellectual property, quality, collection, preservation, etc.

The New York definition bears a risk of circularity. This means that it partially assumes a prior understanding of what grey literature is. For instance, how is it possible to define the meaning of "where publishing is not the primary activity of the producing body"? Simply spoken, the New York definition allows for interpretations such as "grey is that which is not white (nor black)"[vii]. This circularity may be useful for community cohesion and acceptance but not for furthering understanding.

The New York definition also bears a risk of obscurity. It allows for interpretations using ambiguous terms such as fugitive, ephemeral etc. and/or for definitions by negative or excluding concepts. An example from the survey:

> Grey literature is the ABSENCE of key elements, such as occurrence in mainstream literature. Grey literature is hard to find, hard to classify, hard to collect, etc. It occurs in fringe areas (small journals, isolated repositories, unusual formats).

To resume: obviously, the New York definition was (and remains) helpful for an economical understanding of production and dissemination of grey literature but is of limited utility in light of new technology and usage because it doesn't distinguish between online published grey literature and other, non-commercial material.

Based on results from survey and content analysis, we shall try in the following section to identify some "hot topics" that may or should be added as essential attributes to the definition.

4.2. Grey literature as a medium in context: intellectual property

Since the introduction of the concept of grey literature, the question of typology has troubled the grey community. The emergence of the Internet complicated the problem even more.

Strictly applying the New York definition which is centered on production and dissemination, most of the Web material could be described as grey – published in limited number (or unique), badly controlled, non-commercial (or not-for-profit) dissemination. Now, is this realistic, especially given that in some ways the Internet makes material more widely accessible than material printed by publishing houses?

Are digital objects grey, such as datasets, emails, technical copies, wikis, blogs, tweets or even Websites[viii], etc., or are they even literature?

Some respondents to our survey argued in favour of a large, all-embracing concept:

> "Defining grey literature as the 'opposite' to commercial publishing is wrong. (That is, if commercial publishing means publishing for a profit). It is wrong because many non-profit organisations publish on exactly the same

> basis as commercial publishers (e.g. American Chemical Society) - and their publications are not grey. To my mind, grey literature is all scholarly work that is published without a formal peer-review (or equivalent) process outside the traditional journal and book channels. Therefore, grey literature includes data outputs, analytical tools, software as well as the more usual 'classic' pre-print or working paper outputs."

It is doubtful whether data outputs, analytical tools or software should or can be regarded as "scholarly work" in terms of literature. Perhaps we should consider the question from another perspective. In the post-Gutenberg era, the French interdisciplinary network RTP-DOC elaborated a 3D-theory describing digital documents as sign (content), form (structure) and medium (context) (Pédauque, 2003). According to the RTP-DOC approach, the former print-related criteria and attributes can't be transposed to digital objects.

To define grey literature (only) as a specific content or structure (e.g., type of document) seems impossible and will fail. Obviously it is a value on the third dimension, medium, and should be defined in its context, linked to attitudes, understanding and awareness, social norms, and in relation with users, usage and procedures.

So first, the concept of grey literature should be limited to the specific meaning of *literature*, not as a content or structure/type, but derived from its social or legal nature: A digital object is grey literature if and only if it is an item protected by intellectual property rights. In other words, grey literature implies authorship and a character of works of the mind.

The author's rights (or copyrights) may be held by the producing body and/or by the author(s) but in no case by a commercial vendor. Recently, Kansa et al. (2010) raised the question of licensing. This may be an adequate answer for grey items disseminated on the Web by their producers and/or authors but is in no case a sufficient condition to define grey literature.

In contrast, a functional definition may not be beneficial. Our results indicate that grey literature does not fulfill functions dissimilar from serials or books sold by vendors but rather focuses on a more limited variety especially when it comes to dissemination and preservation. As an attribute, this is not enough to differentiate grey items from others.

4.3. From typology to quality

Some authors tend to define grey literature through a list of different types of documents. Such an extensional or denotative definition of grey literature is condemned, for at least two reasons: (a) there are too many different categories of grey literature, especially in digital format, and (b) a definition can't keep up with new types and formats; and the same type of document can be grey as well as white, at different moments of its life-cycle and depending on the context.

For instance, are PhD theses grey? Are they always grey? Are Master theses grey? What about undergraduate dissertations? What is the difference? Some reports are sold through conventional sales and distribution channels, and some grey items are reports. What makes some reports turn grey?

Our explanation is quality and scientific and/or technical value. Alongside with dissemination and intellectual property, it helps to distinguish grey from other "dark grey" items without interest for research and scholarship. In fact, while dissemination and production are attributes essential to fix the line between commercial (white) and grey items, intellectual property and quality are useful on the "darker side" of grey literature, for instance to make a clear distinction between undergraduate work and Master or PhD theses, or between a working paper and a manuscript in a preliminary draft version.

Quality means quality assurance and control and intrinsic scientific value. A recent synthesis on reports states that "it is (…) a misconception that grey literature is not peer reviewed (…) reports often proceed through several levels of internal and agency review before final printing" (Seymour, 2010). The 2005 initiative GLISC[ix] or "Nancy group" was expected to establish a label of quality to reports. Other grey items such as theses, dissertations or communications undergo procedures of labelling and/or validation that guarantee a minimum quality level. Our own surveys on grey literature in France and Belgium reveal up to 60% items with some kind of quality control.

Quality becomes a central question of grey literature. In our survey, more than 50% of the respondents consider quality as essential when defining grey literature. Farace (2010) compares peer-review procedures of grey literature to serials. In terms of quality management, some procedures are rather designed to detect non-quality (for instance, assessment and examination of PhD theses) while others are supposed to prevent from non-conforming or non-acceptable quality levels (like the GLISC guidelines).

All the same, quality is not an absolute, stand-alone value but linked to standards, usage or specific purpose, and it is based upon a customer's evaluation and/or experience. This leads us to the next and last essential attribute of grey literature.

4.4. Grey literature without library?

Who is the customer of grey literature? Net users most often are not aware of special categories of items and dissemination channels. Their problem is access, not production, distribution or collection. Unpublished or semi-published documents mean little to them. They rarely bother with grey or white documents yet they don't like "black hole literature", material they cannot get.

In fact, the real customer of grey literature, since the invention of the concept, is and always has been the librarian, the LIS professional who tries to collect it. This is a striking point when going through the papers written on grey literature in the last two decades. "Searching and accessing the grey literature (…) have always been a great challenge *for librarians and documentalists*" (Natarajan, 2006)[x]. This is a problem not for scientists, students or other patrons but for LIS professionals – this is the point.

In other words, would the problem with grey literature persist if librarians would stop trying to collect it? Would a working paper be grey if nobody cares for it? A preprint sent by its author to a colleague, is it grey? Without a doubt, this seems a bad question. A better question would be: would a library consider this specific item as worthwhile for collection, because of its quality, scientific value, uniqueness or heritage character? Alternatively, does a document become labelled as grey because it is produced or validated by an institution?

This means that the definition of grey literature is intimately conditioned by the fact that it is an object of collection and acquisition. A document becomes grey not only because it is a work of the mind and not sold by a vendor but insofar someone – an institution, a library, an information service, a professional – shows interest to get it. It becomes grey because it purports to have (or is imbued by the librarian or scientific or preservation community to have) quality, scientific value, uniqueness or heritage character and while it is being collected so it can be accessed or referenced in the future the "grey" label acknowledges that the document is unique with respect to one or a series of other issues, such as dissemination, production and so on.

This helps us to clarify two aspects. First, "collection" does not mean "storage". It stands for selection and acquisition policy, is dynamic, usage-oriented and linked to cultural goals or community needs. A "difficult-to-get" item becomes grey when it is considered useful (or thought to be useful in the future) for a scientist, a research team, a laboratory, an institution or a community.

Second, grey literature is not hidden or classified but open source material. Items that can't be collected are not grey. Again, the crucial point is not hiding or classifying but collection and making available for the future.

Today, 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. This may explain the growing interest of the grey literature community for this special type of open archives[xi]. A new definition should or could mention this.

"Can you imagine a world without grey literature?" This is one question of the survey (see above). Most experts cannot. Now, can you imagine grey literature without LIS professionals? The challenge of grey literature is a challenge of (inter)mediation. There is no stand-alone grey literature. Grey lies in the eye of the beholder. It is a medium.

5. Conclusion

The current definition of grey literature – the New York definition – remains helpful and should not be replaced but adapted to the changing environment.

As we showed above, the typological approach doesn't provide an exhaustive and explicit list of items. The economic approach of the New York definition, on the other hand, is intensional and specifies the necessary condition for a document being part of the grey literature. But the same definition is not sufficient in the context of Internet publishing, and we need to designate more essential attributes to clearly differentiate grey from other items.

Our proposal is to add four attributes to the New York definition:

1. *Document character of grey literature (concept of RTP-DOC).*
2. *Legal nature of works of the mind, e.g., protection by intellectual property.*
3. *A minimum quality level (peer review, label, validation).*
4. *The link to (inter)mediation, e.g. the interest of grey items for collection (and not for the end-user).*

Our proposal for a new definition ("Prague definition") of grey literature is as follows:

> "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."

Grey literature includes all kind of quality or seminal documentary material a library would like to collect but can't easily because of non-conventional distribution channels. It is not (only) a question of production and dissemination but (also) of quality and collection. Without (inter)mediation by libraries, no grey literature. It is a case for LIS professionals, a challenge that brings together the communities of grey literature and special collections.

A couple of years ago, the main problem with grey literature appeared to be economics. Simpson (1995) observed, "peripheral materials, including grey literature, expand unabated. Libraries having difficulty collecting traditional materials have little hope of acquiring the periphery."

Today, due to the overwhelming success of Web publishing and access to documents the focus has shifted to quality, intellectual property and (inter)mediation. Without a revision that includes the mentioned attributes, the current definition risks being increasingly unable to differentiate grey from other documents.

Our proposal for a revised "Prague definition" brings together the former economic approach with new attributes. The next step should be to check this definition against common usage in libraries and different types of grey and other documents. Once done, the value of the definition can be evaluated on the basis of the answers to the following two questions:

1. Does this new definition include all kind of documents usually considered by LIS professionals as grey literature, including today's difficult-to-process and hard-to-collect items?
2. Does it lead to further differentiation or better understanding of how grey literature may be distinguished from other forms of literature?

Doing so, special attention should be paid to three challenges facing us at the initiation of the 21st century.

(1) The development of institutional repositories by publishing organizations as a complementary and sometimes concurrent service to tradition library holdings; and the place and processing of grey literature in theses archives.

(2) The tendency of disintermediation in the traditional value chain of scientific and technical information. The "risk" of grey literature is not web-based technology but the somehow fading role of libraries and information professionals as intermediaries between authors, publishing bodies and the end-user. And tell the reader why this is important other than job preservation.

(3) The so-called *Fourth Paradigm*[xii], e.g. data-intensive science and the access to datasets that together generate a trend to transform and/or marginalise literature (documents).

With reference to grey literature, replies to our survey stated "(…) it is important for knowledge" and "it is a question of freedom" or "non-mainstream publishing". The future will show if our concept of grey literature remains "ephemeral" and if it contributes to better understanding and processing of this special part of scientific and technical information.

Acknowledgement. My sincere thanks to all the colleagues who have contributed to the state of the art and the survey, especially to Dominic Farace, Leonid P. Pavlov, Deni J. Seymour and Hélène Prost for helpful advice and proofreading.

6. References

H. M. Artus (2003). `Old WWWine in New Bottles? Developments in electronic information and communication: structural change and functional inertia'. In D. Farace & J. Frantzen (eds.), *GL5 Conference Proceedings. Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information.* TextRelease.

C. P. Auger (ed.) (1975). *Use of Reports Literature*. Butterworths, London.

C. P. Auger (ed.) (1989). *Information Sources in Grey Literature.* Bowker Saur, London, second edn.

D. Farace & J. Schöpfel (eds.) (2010). *Grey Literature in Library and Information Studies*. De Gruyter Saur, Berlin.

T. Hey, et al. (eds.) (2009). *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Microsoft Corporation.

E. Kansa, et al. (2010). `Googling the Grey: Open Data, Web Services, and Semantics'. *Archaeologies* **6**(2):301-326.

S. Keenan (1996). *Concise Dictionary of Library and Information Science*. Bowker Saur, London.

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

J. Mackenzie Owen (1997). `Expanding the Horizon of Grey Literature'. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information, 13-14 November 1997*.

R. T. Pédauque (2003). `Document: Form, Sign and Medium, As Reformulated for Electronic Documents'. Tech. rep., CNRS RTP-DOC.

J. Schöpfel (2008). `Grey literature on bilingualism in Belgium'. *Cahiers de la Documentation / Bladen voor Dokumentatie* **62**(2):22-30.

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.

D. Seymour (2010). `Sanctioned Inequity and Accessibility Issues in the Grey Literature in the United States'. *Archaeologies* **6**(2):233-269.

D. N. Wood & A. W. Smith (1993). `SIGLE: A Model for International Co-operation'. *Interlending & Document Supply* **21**(1):18-22.

Annexe A - Corpus

A. Aceti, et al. (1999). 'A profile of GL producers in the field of safety and health at workplaces in Italy: results of a sample survey'. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature, 4-5 October 1999*.

M. A. Banks & C. de Blaaij (2006). 'Implications of Copyright Evolution for the Future of Scholarly Communication and Grey Literature'. In *Eighth International Conference on Grey Literature: Harnessing the Power of Grey, 4-5 December 2006*.

C. de Blaaij (1999). 'Intellectual property of the move'. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature, 4-5 October 1999*.

C. de Blaaij (2003). 'Grey literature from invisibility to visibility: The quest for grey content in the domain of the invisible web'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information, 4-5 December 2003*.

A. K. Boekhorst, et al. (2004). 'Grey Literature Survey 2004 - A research project tracking developments in the field of grey literature'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, 6-7 December 2004*.

J. P. Chillag (1993). 'From Weimar to Maastricht and beyond : half a century with grey literature'. In *First International Conference on Grey Literature: Weinberg Report 2000, 13-15 December 1993*.

G. P. Cornish (1999). 'Copyright: Black and White or just making you see Red? '. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature, 4-5 October 1999*.

G. A. Cotter & B. C. Carroll (1993). 'A new generation of Grey Literature: the impact of advanced information technologies'. In *First International Conference on Grey Literature: Weinberg Report 2000, 13-15 December 1993*.

J. Crowe & T. S. Davidson (2008). 'The 'Grey' Intersection of Open Source information and Intelligence'. In *Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society, 8-9 December 2008*.

R. Di Cesare & C. Sala (1995). 'The use of grey literature in the agricultural economics field : a quantitative analysis'. In *Second International Conference on Grey Literature: Grey Exploitations in the 21st Century, 2-3 November 1995*.

P. A. Erwin (2006). 'The Messy World of Grey Literature in Cyber Security'. In *Eighth International Conference on Grey Literature: Harnessing the Power of Grey, 4-5 December 2006*.

J. Gelfand (1993). 'Academic libraries and collection development implications for grey literature'. In *First International Conference on Grey Literature: Weinberg Report 2000, 13-15 December 1993*.

J. Gelfand (1999). 'Grey Literature: Plausible Applications for Distance Education Practices'. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature, 4-5 October 1999*.

P. Gokhale (1997). 'Grey Literature Varieties - Definitional Problems'. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information, 13-14 November 1997*.

H. von Hofe (2004). 'Towards a Genealogy of Grey Literature via Newton's Journals'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, 6-7 December 2004*.

K. G. Jeffrey (1999). 'An Architecture for Grey Literature in a R&D Context'. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature, 4-5 October 1999*.

K. G. Jeffery & A. Asserson (2006). 'Hyperactive Grey Objects'. In *Eighth International Conference on Grey Literature: Harnessing the Power of Grey, 4-5 December 2006*.

J. C. Kufa (1993). 'Grey literature in academic research in Developing Countries : Problems and Prospects'. In *First International Conference on Grey Literature: Weinberg Report 2000, 13-15 December 1993*.

D. Luzi (1995). 'Internet as a new distribution channel of scientific grey Literature: the case of Italian WWW servers'. In *Second International Conference on Grey Literature: Grey Exploitations in the 21st Century, 2-3 November 1995*.

D. Luzi, et al. (2003). 'The communication flow of research projects results'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information, 4-5 December 2003*.

B. H. MacDonald, et al. (2003). 'Grey Literature in the Life of GESAMP, an International Marine Scientific Advisory Body'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information, 4-5 December 2003*.

J. Mackenzie Owen (1997). 'Expanding the Horizon of Grey Literature'. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information, 13-14 November 1997*.

J. C. McDermott (1995). 'Defining grey literature for the 21st Century'. In *Second International Conference on Grey Literature: Grey Exploitations in the 21st Century, 2-3 November 1995*.

M. Nahotko (2007). 'Some Types of Grey Literature: A Polish Context'. In *Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, 10-11 December 2007*.

M. Natarajan (2006). 'Grey literature: problems and prospects for collection development in e-environment'. *The Grey Journal* **2**(2):100–105.

L. P. Pavlov (2003). 'The Commercialization of Research Findings Documented in Grey Literature'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information, 4-5 December 2003*.

L. P. Pavlov (2004). 'The Problems of Grey in the Context of Postmodern Science and Culture'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, 6-7 December 2004*.

S. L. Ranger (2004). 'Grey Literature in Special Libraries: Access and Use'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, 6-7 December 2004*.

J. Schöpfel (2006). 'Observations on the Future of Grey Literature'. *The Grey Journal* **2**(2):67–76.

M. S. Seadle (1999). 'Grey Copyrights for Grey Literature: National Assumptions, International Rights'. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature, 4-5 October 1999*.

D. B. Simpson (1995). 'Grey Literature: the challenges for an increasingly important body of research literature'. In *Second International Conference on Grey Literature: Grey Exploitations in the 21st Century, 2-3 November 1995*.

C. Stock & J. Schöpfel (2003). 'Grey Literature in an Open Context: From Certainty to New Challenges'. In D. Farace & J. Frantzen (eds.), *GL5 Conference Proceedings. Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information.*, pp. 199–202. TextRelease.

C. Stock & J. Schöpfel (2008). 'Grey literature in French digital repositories: a survey'. In *Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society, 8-9 December 2008*.

P. Sulouff, et al. (2005). 'Learning about grey literature by interviewing subject librarians'. *The Grey Journal* **1**(3):131–136.

R. H. A. Wessels (1997). 'An EAGLE's eye-view of grey literature research'. In Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information, 13-14 November 1997.

Annexe B - Survey

During previous conferences in the GL-Series, several authors have raised questions concerning the (re)definition of the term grey literature. In this exploratory survey, an effort is made to assess attitudes towards the currently held definition and to gather elements that may lead to a redefinition of the term grey literature.
This online survey will take about ten minutes of your time to complete. Your response is greatly appreciated.

Dr. Joachim Schöpfel
Head, Department of Information Sciences
University of Lille 3, France

1. What is/are the main function(s) of grey literature?

- Registration (to establish ownership and priority and to clearly date-stamp the work)
- Certification (to have the quality of the research acknowledged)
- Dissemination (to let the public know about the work)
- Archiving (to provide a permanent record of the work)
- Other (please specify)

2. The 2004 amended definition of the term grey literature reads as follows:
"Information produced and distributed on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishers i.e. where publishing is not the primary activity of the producing body".
What do you think about this definition?
The current definition of grey literature

- remains useful
- needs revision
- is precise
- doesn't fit with new technologies
- remains relevant

Strongly agree/Agree/Disagree/No Comment

3. In the past, definitions of grey literature stressed one or more aspects, e.g. editing or producing grey literature, dissemination, different types of grey literature, etc. What do you think about these aspects? Are they important or not?
When defining grey literature, I think that the following aspect is...

- production
- dissemination
- typology (document type)
- quality
- acquisition

Very important/Important/Not so important/Not at all important/No comment

- other (please specify)

4. What do you think about the following statements?

- Grey does not imply any qualification.
- The Web 2.0 can be considered as grey.
- Grey is merely a characterization of the distribution mode.
- Grey is linked to open access.
- Raw data are part of grey literature.

Strongly agree/Agree/Disagree/Strongly disagree/No Comment

5. Do you agree with the following statements?
Grey literature will continue to play a significant role alongside commercial publishing.
The borderline between "grey" and "white" (commercial) literature will become increasingly indistinct.
The proportion of "grey" documents published on the Web will continue to increase.
The Internet will encourage a greater diversity in the types of "grey" resources available.
Bibliographic control of grey literature will remain problematic.
Open archives will offer more appropriate services and functions for at least some segments of grey literature.
Some organizations will develop tools and services to aid in the efficient exploitation of grey resources on the Web.
Searching and collecting grey literature will become as straightforward as it is for journals and books.
New tools for collecting, depositing, and archiving will make GL less ephemeral and volatile than in the past.

6. Can you imagine your world without grey literature?

7. In the field of grey literature, which question most interests you?

8. You are a...
Scientist
Scholar
Information professional
Student
Other:

Endnotes:

[i] http://opensigle.inist.fr/handle/10068/697753

[ii] http://www.refdoc.fr

[iii] On the platform http://fr.surveymonkey.com/

[iv] The controversy about quality is reminiscent of the more recent IPCC Amazongate debate. See for instance, http://www.guardian.co.uk/environment/georgemonbiot/2010/jul/02/ipcc-amazongate-george-monbiot

[v] Concerning relevance of GL for local or regional community, see our recent study on academic publishing in Belgium (Schöpfel, 2008).

[vi] With the words of one respondent of our survey: "Keeping our focus. Now that GL has become 'sexy' - people expand the boundaries beyond what makes sense. Raw data is not GL. Ephemera is not GL. The entire Web is not GL. I think losing the focus and overexpanding runs the risk of losing credibility when making the arguments for the value of GL - especially in the sciences, when up against the elitism of more formal publication (…)"

[vii] "Grey literature can only be defined in its relation to formally published or 'white' literature (…)" (Artus, 2003). But then, what is white, and where does it begin?

[viii] The British Library starts to archive snapshot copies of selected Websites.

[ix] http://www.glisc.info

[x] Italics added by JS

[xi] See the recent study of Luzi (2010).

[xii] See Hey at al. (2009).

A Terminology Based Re-Definition of Grey Literature

Claudia Marzi, Gabriella Pardelli, and Manuela Sassi
Institute for Computational Linguistics "Antonio Zampolli",
CNR, National Research Council, Italy

Abstract
The conventionally accepted definition of Grey Literature, as Information produced and distributed by non-commercial publishing, does not take into consideration either the increasing availability of forms of grey knowledge, or the growing importance of computer-based encoding and management as the standard mode of creating and developing grey literature.
Semi-automated terminological analysis of almost twenty years of terminological creativity in the proceedings of eleven GL International Conferences offers the opportunity to pave the way to a bottom-up redefinition of Grey Literature stemming from attested terminological creativity and lexical innovation.
In this paper, we focus on a set of automatically-acquired terms obtained by subjecting our reference Corpus to a number of pre-processing steps of automated text analysis, such as concordances, frequency lists and lexical association scores. Acquired terms allow us to throw in sharp relief developing trends and important shifts of emphasis in the current understanding of the notion of Grey Literature.

Theme: **Redefining Grey Literature** – Proof of Concept – Research Process

Keywords: Grey Literature definition, GL Conference corpus, Terminology extraction

1. Introduction

1.1 Grey Literature definition

The Luxembourg Convention on Grey Literature held in 1997 offered the following definition of Grey Literature (expanded in New York, 2004): "Information produced *and distributed* on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing*, i.e. where publishing is not the primary activity of the producing body".*
The questions that immediately arise are the following: is this definition still valuable? Is it so far completely satisfactory? Or does it rather need important modifications?
And what about other conventionally accepted definitions and descriptions?

In considering the evolution of the role and definition of Grey Literature, Augur (1989) started from the beginning of the 20th century, where the notion of GL had been, for many years, coextensive with that of report literature: documents evolving out form research and development activities, particularly in the aircraft and aeronautics industries, were a very important means of communicating the results of research testing. In particular, World War Two had the greatest impact on report literature, transforming it into a major vehicle of communication. By the 1970s GL became the recognized medium for dissemination and promotion for many organizations and was considered an important reading throughout the world, though not easy to find. By the 1980s other scientific domains such as Social Sciences, Economics and the Humanities were included in the wide range of research reports, discussion and policy documents, working and conferences papers, etc. A huge increase in quantity as well as the advantageous effect of the flexibility and speed, however, didn't completely obscure problems of identification and acquisition; given the nature of this kind of literature, many categories contained security restrictions. In the 1990s GL attained its importance as an independent medium of communication because of an initial need for security of confidentiality classifications which prevented documents from being published in a conventional manner.
Hirtle (1991) gave a definition of GL as "the quasi-printed reports, unpublished but circulated papers, unpublished proceedings of conferences, printed programs from conferences, and the other non-unique material which seems to constitute the bulk of our modern manuscript collections".
IGLWG (*Interagency Gray Literature Working Group*) defined in 1995 GL as "open source material that usually is available through specialized channels and may not enter normal channels or systems of publication, distribution, bibliographical control, or acquisition by booksellers or subscription agents".
Debachere (1995) described GL as "a range of materials that cannot be found easily through conventionally channels [..] but which is frequently original and usually recent".
Actually, quoting Wikipedia "Grey Literature is a term used [..] to refer to a body of materials that cannot be found easily through conventional channels such as publishers [..]".

All these descriptions of Grey Literature are phrased negatively; often GL is defined by contrast to other things. In other words, we notice that particular emphasis is laid on what GL is not, rather than on what it is.

To sum up, all these definitions and descriptions of Grey Literature do not take into account those aspects that, in our view, are most strongly associated with the increasing availability and accessibility of GL materials, and the growing importance of computer-based encoding as the standard medium of creating and developing GL.

Our general idea is that a domain-specific document repository offers the possibility to pave the way to a bottom-up redefinition of Grey Literature stemming from attested terminological creativity and lexical innovation.
We intend to inquire and monitor terminological creativity over almost twenty years of technical and scientific work in the frame of the International Conference on Grey Literature, and to ground suggestions for a re-definition on those terms that appear to be consensually shared by the various disciplinary sub-fields.

1.2 Reference corpus
The empirical basis of our work is represented by the Corpus of *GreyText Inhouse Archive,* available on http://www.greynet.org/opensiglerepository.html consisting of titles, themes, keywords and full abstracts, for a total amount of around ninety thousand tokens (containing around seventy thousand word tokens).
Although comparatively small, the corpus suits the purposes of our present investigation quite nicely. First, it is fairly well structured, allowing selective search of relevant terms in a context-sensitive way. Moreover, it contains highly informative text excerpts, as titles and abstracts are, conveying document contents in a quintessential way. The traditional haystack problem in information extraction from unstructured materials is here considerably reduced, as all texts belonging to the corpus are characterized by a high density of mostly salient terms. Thirdly, the corpus presents a longitudinal selection of documents ranging over several years of intensive research in GL. This will allow a terminological trend analysis in a diachronic perspective.

2. Methodology

2.1 Research rationale
Our general idea is that an interesting re-definition of GL can be based upon careful examination of the longitudinal trend of almost twenty years of terminological creativity in the proceedings of the eleven GL international Conferences, by focussing on a set of automatically-acquired terms (both single-word and multi-word terms) obtained by subjecting our reference Corpus to a number of pre-processing steps of automated text analysis, such as concordances, frequency lists and lexical association scores (e.g. *Mutual Information* on word pairs).

Although knowledge-poor, bag-of-words approaches to text mining have proved to perform effectively in traditional tasks such as document classification and indexing, intelligent access to the contents of a document repository requires going beyond the over-simplistic notion of a text as an unordered collection of loose word tokens. Automated identification of the most relevant terms in a domain-specific document repository represents an important step in this direction. It is commonly assumed that salient domain-specific concepts and relations are conveyed in text through statistically significant terms, whether they are simple words like *computer* and *web*, or structurally more complex word sequences like *computer science* and *world wide web*. This requires that a raw text is preliminarily marked up at different levels of linguistic analysis, ranging from tokenization and part-of-speech tagging, to chunking and dependency analysis. Relevant terminological units are then tracked down 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. All text strings that fit into the targeted morpho-syntactic pattern (e.g. *networks of institutional repositories*) are then filtered out through a further step of statistical post-processing, to assess their potential for termhood.

Filtering methods considerably vary in the literature, ranging from raw frequency lists and traditional Information Retrieval measures such as TF-IDF (Baeza-Yates & Ribeiro-Neto, 1999), to more sophisticated indices like the C/NC-value (Frantzi et al., 2000) or lexical association functions such as "log likelihood" and "point-wise mutual information" (Manning & Schütze, 1999). The result of this filtering step is a list of relevant term candidates, possibly to be validated by a domain expert but already usable for advanced content indexing.

In fact, more can be done on the way to understanding their content and the role they play in a document repository. With a view to meeting these further goals, we need to take into account the particular context where terms occur, the network of textual relations they entertain with other words and the semantic roles they play. Such a finer-grained analysis can be carried out in many ways: i) manually, through inspection/classification of a relevant list of concordances of the terms of interest, ii) semi-automatically, by automatically clustering words that occur in the same contexts (Lenci et al.,

2006), and then having experts classify the resulting clusters; iii) fully-automatically, by clustering words and then discovering their semantic relations by using machine learning techniques (Mitchell, 1995).
For the present purposes, a context-sensitive analysis of relevant terms (domain specific word forms) was carried out through manual inspection of relevant list of concordances and frequency (step i above), since the other approaches require availability of a considerably larger amount of textual data. Nonetheless, we believe that our preliminary analysis illustrates the potential of the corpus-based approach to domain definition we propose here.

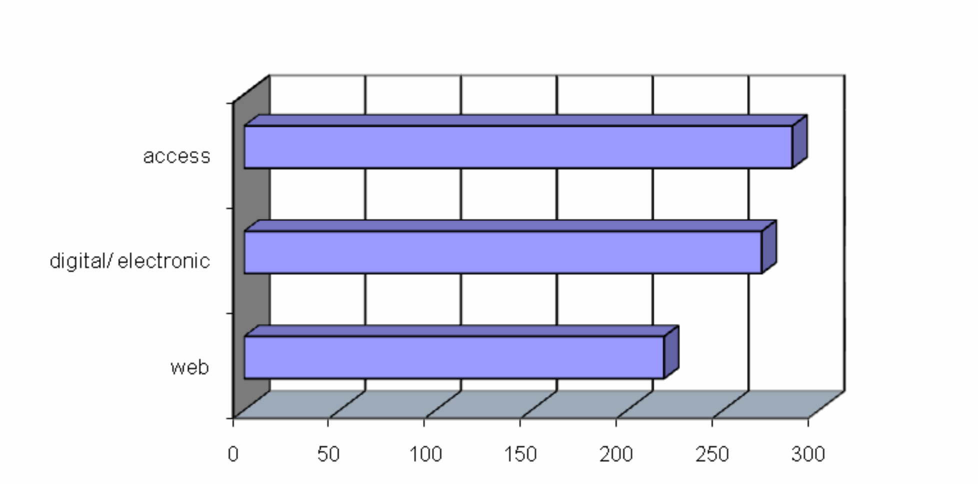
2.2 Data extraction

We started from a single-word frequency list, acquired automatically from around 90,000 tokens. Items in the list, that contains words that occur more than 9 times (an empirical threshold corresponding to 0.01% of the total size of our reference corpus), were ranked by decreasing frequency values to bring the most relevant terms to the top, as shown below.

985	grey
966	literature
737	**information**
477	research
220	**access**, conference
204	library
202	paper
191	**documents**
182	scientific
178	proceedings
162	development
159	project
142	**electronic**
137	system
130	use
128	science
128	**digital**
126	report
124	available
121	libraries, open, technical
120	data, national
119	collection
106	public
104	**knowledge**
103	publishing, university
102	study
100	results
98	international
97	resources, technology
95	metadata
92	repository
91	users
88	published
85	database, publications, **web**
84	academic, document
83	analysis
82	authors
80	communication
77	management
75	services
74	countries
73	researchers
72	government, **internet**, repositories
70	work
69	materials
68	health, projects, world
67	databases
65	community
64	bibliographic, theses
63	survey, systems
61	european, sigle, social
57	content
40	**accessible**
32	**online**
30	**www**
27	amsterdam, archive, author, economics, nature, reference, references
26	**accessibility**, agencies, citations, formats, intellectual, technological
25	networked, political, professionals, security, standards
24	communities, concept, industry, language, legal, virtual
23	italy, paradigm, physics, societies, uk
22	exchange, goal, japanese, preservation, purpose, scientists, sharing, useful
21	learning, retrieval, significant, topic
20	governmental, identification, networking, property, questions
19	historical, marine
18	india
17	school, searching, site, sites, tool, transfer
16	botswana, market, model, worldwide
15	agricultural, benefit, distance, financial, free, multicultural, multiethnic, poland, questionnaire
14	communications, companies, datasets, delivery, educational, networks, Russia
13	african, bank, catalogs, cooperation, cultural, multimedia, visibility
12	catalogue, techniques, unpublished
11	cognitive, czech, freedom, method, methodology, rural, semantic, words
10	america, benefits, bibliography, catalogue, collaboration, culture, engineering, engines, france, germany, literatures, media, participation, website

Frequency distributions were then aggregated by putting in the same frequency class quasi-synonyms or semantically closely related terms (e.g. *internet, web* and *www*). This post-processing step allowed us to assess how often a concept, or ontological entity, was used in the corpus.

The analysis has been centred on those concepts which appear innovative with respect to the traditional definitions of Grey Literature reported above. Accordingly, core notions such as *information* and *documents*, which figure prominently in our list, although undoubtedly relevant to a proper characterization of GL, are taken to be too well established to deserve further analysis. Here, we rather intend to focus on highly salient concepts that appear to be shared by various disciplinary sub-fields, and mark, in our view, important steps in the evolution of current understanding of GL. In particular, we selected three such notions: *digital/electronic*[i]*, access* and *web*. Their aggregated frequency distributions are shown in the diagram below:



As a further step in our analysis, we considered lexical association scores between salient terms (e.g. m*utual information* on word pairs), focussing on terminological usages that are closely related to the ontological entities already mentioned above.

The typical collocates of a*ccess* are: *easy, electronic, facilitate, full, grey, information, internet, journals, literature, materials, movement, multicultural, open, public, repository, research, scientific.*

Digital combines with *document, grey, library, literature, network, object, project, repository, system, technology, theses.*

Electronic keeps company with *format, grey, information, journal, literature, network, paper, publication, report, resource, technical, theses.*

Reference to the notion of *web* is typically accompanied by *access, database, grey, information, network, literature, science.*

Finally, particular emphasis should be placed, in our view, on the use of *knowledge* coupled with *base, exchange, generation, infrastructure, management, scientific, service, share, society,* and, especially, *information*.

3. Results

Term aggregation by conceptual unity and manual inspection of the most recurrent contexts of use of selected terms shed considerable light on both established and innovative notions. The steadily increasing occurrence of the attribute d*igital/electronic* bears witness to the growing importance of computer-based encoding as the *standard medium* of GL. Here, availability in digital format appears to be the outcome of an integrated system of software tools for efficient, possibly metadata-oriented document production and management, and an essential prerequisite to ubiquitous dissemination and ready accessibility.

The noun *access* (defining the process of accessing text documents), is seen in the company of adjectives like *easy*, *full* and *open*. The usage underlines important conceptual innovations in the way GL material is distributed and eventually used; e.g. *open access* focuses on the free accessibility and reusability of digital contents. Coupled with *information, document* and *repository, access* appears to point to a conception of world-wide available, structured digital contents, offering the combined advantage of ubiquitous accessibility and quality control under authoritative document management. As Farace (2006) puts it, "open access to information is the key to knowledge, both in its generation and transfer". The management of valued resources in a global environment is in fact conducive to the extraction and combination of targeted information and, eventually, to the generation of innovative knowledge. This perspective lays emphasis on the increasing importance of information management systems for GL, and casts doubts on those definitions of Grey Literature as "a mere characterization of the distribution mode" (as already pointed out by Mackenzie Owen, 1997).

Finally, systematic reference to the notion of *web* throws in relief the huge importance of the World Wide Web as the standard means of disseminating GL, and the role of networking communities, acting at the same time as providers and users of GL material in a highly distributed, collaborative scenario.

4. Concluding remarks

Grey Literature defines an innovative approach and methodology for a wide information dissemination and exchange, by offering the web-based sharing facilities and distributed access to openly available scientific and technical document repositories, possibly under authoritative content management.

An updated re-definition of GL should take into consideration the key notions of digital medium, web-based distribution channels, information access policy and access and management tools for GL. By bringing these innovative elements into the picture, we are in a position to do justice to recent developments in the evolution of GL, where traditional core notions such as *information*, *distributed access* and *electronic/digital format* appear to acquire novel, cooperative and interactive undertones, coupling the advantages of flexibility, speed and quantity, with the further bonus of ubiquitous accessibility and content quality control in a global cooperative environment. In fact, by blurring the traditional divide between providers and users of document repositories, GL not only defines a policy for distribution and access of information, but does promote new, creative modes of production and use of innovative knowledge.

At its core, Grey Literature is about producing and distributing the seeds of new knowledge.

References

AUGUR, CHARLES P. (1989). *Information Sources in Grey Literature*. Bowker-Saur, London.

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

BAAYEN R. H. (2001). *Word Frequency Distributions.* Kluwer Academic Publishers, Dordrecht.

CARROLL B. C. AND COTTER G. A. (1997). A new generation of grey literature: The impact of advanced information technologies. *Publishing Research Quarterly*, 13 (2), 5-14.

DEBACHERE, M. C. (1995). Problems in obtaining grey literature. *IFL4 Journal,* 21 (2), 94-98.

CHURCH K. W., HANKS P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics* n. 16 (I), 22-29.

FARACE D.J.(2006) (Guest Editor). Introduction: Open access to grey resources*. Publishing Research Quarterly*, 22,(1), 3.

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.

FRIEDMAN, T.L. (2005). *The world is flat. A Brief History of the Twenty-first Century.* Farrar, Straus, Giroux New York.

HIRTLE, PETER (1991). *Broadsides vs. Grey Literature.*

JEFFERY K.G., ASSERSON A. (2007). Hyperactive Grey Objects. *Publishing Research Quarterly*, 23(1), 71-77.

ERIC C. KANSA, SARAH WHITCHER KANSA, MARGIE M. BURTON AND CINDY STANKOWSKI (2010). Googling the Grey: Open Data, Web Services, and Semantics. *Archaeologies*, 6 (2), 301-326.

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.

MACKENZIE OWEN J. (1997). The expanding horizon of grey literature. *GL3 Conference Proceedings. GreyNet*, AMSTERDAM,11-13.

MARCUS A. BANKS (1996). Towards a continuum of scholarship: The eventual collapse of the distinction between grey and non-grey literature. *Publishing Research Quarterly,* 22 (1), 4-11.

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

MANNING C., SCHÜTZE H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, MA.

MITCHELL TOM M. (1995), *Machine Learning*. McGraw Hill, New York.

SCHÖPFEL J. (2006). Observations on the future of Grey Literature. *The Grey Journal,* 2, (2), 67-76.

SEYMOUR D.J. (2010). In the Trenches Around the Ivory Tower: Introduction to Black-and-White Issues About the Grey. *Archaeologies*, 6,(2), 226-232.

DBT Software (CNR patent) http://www.ilc.cnr.it/pisystem/prodotti/index.html

GREYNET WEB SITE http://www.greynet.org

[i] Contrary to our expectations and a general terminological trend, the attribute *electronic* continues to be used interchangeably with *digital* to characterize a document and/or its content; therefore we are considering both of them.

Peering through the Review Process: Towards Transparency in Grey Literature

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

Abstract
Now that grey literature is readily catalogued, referenced, cited, and openly accessible to subject based communities as well as net users, the claims that grey literature is unpublished or non-published have sufficiently been put to rest. However, now that grey literature has met these former challenges and entered mainstream publishing, it requires in the spirit of science to have a system in place for the quality control of its content. This new challenge has recently been spurred by the IPCC affaire involving the use/misuse of grey literature and is now almost a daily topic in the world media. The purpose of this study will be to explore the degree to which grey literature is reviewed and to compare similarities and differences with formal peer review carried out in various degrees by commercial publishers. This study will further distinguish the review process implemented by grey publishers from that of mavericks and vanity press, where personal opinion and pure speculation run rampant. This study looks at the body of literature on peer review and its relatedness to grey literature. Key concepts and elements in peer review form the framework for a comparative analysis, and these are examined in light of guidelines on peer review submitted by publishers. In addition, alternative models for peer review found during the course of study are compared for their relevance to grey literature.

Keywords: Peer Review, Grey Literature, Publisher

1. Introduction

If you've already read the abstract, you're aware that I begin this study with a basic assumption - being that grey literature is readily catalogued, referenced, cited, and openly accessible to subject based communities as well as net users. And, that the claims that grey literature is unpublished or non-published have sufficiently been put to rest. With this aside, I proceed to the topic at hand, that being a look at peer review as it is applies to grey literature. I think that it should not go unsaid that I was somewhat hesitant to even use the term "peer review" at the start as seen in the very formulation of the title of my presentation, where I refer instead to the review process. Only during the course of the study, do I feel more inclined to use the term peer review.

The grey literature community has systematically addressed earlier challenges and in doing so has entered mainstream publishing. The grey literature community's response to a general survey [Boekhorst et al., 2005] reveals that a majority of the respondents are of the opinion that grey literature is always subject to a review process and that the content of commercially published literature is not superior to grey literature. However five years on, in the spirit of science, and faced with the recent challenge spurred by the IPCC affaire [Clover, 2010] in which the use/misuse of grey literature has become a daily topic in the world media it now becomes necessary to examine the grey literature community's position on peer review.

Peer review could be seen as the last citadel where commercial publishers have the opportunity to discredit grey literature instead of coming to terms with it. However from the media perspective, the battlefront appears not to be between grey and commercial publishers but from another affront – that being the black literature, what Carrington [2010] describes as the blogoshpere and what Cohen [2010] refers to as the 'know-nothings'. With the IPCC affair, the blogs and tweets of this world have turned their attacks from commercial to grey publishers. And over the past year of reading such comments and remarks, I view it simply as the lack in understanding what grey literature is and is not.

Carrington [2010, p. 2] and others have identified the beginning of a revolution in the way science is being done. For them, change must come or the trust that the public places in it will be jeopardized. The credibility of science then is at stake. And, this sentiment holds true as well in the field of grey literature. Hence, the premise that commercial and particularly journal articles are peer reviewed deserves a closer look. Bornmann and Daniel [2009, p. 11] state that "What constitutes a refereed journal, varies from journal to journal." For them, the peer review process itself is more or less a black box and much more research on the black box of peer review is needed. Following then this line of argument - that it is inaccurate to assume that a formal peer-review process is followed by all commercial publishers - is it then just as inaccurate to equate the corpus of grey literature as being non-peer reviewed?

2. Purpose of the Study

The purpose of this study is neither to argue nor defend any one position, but rather to draw comparisons from both the commercial and grey publishers and suggest guidelines for the latter. Guidelines that would be commensurate to the resources and knowledge based communities in which grey literature is produced and used. Now that grey literature has entered the wider, public media, net citizens must come to understand in clear terms what it is; and its value must be made transparent for its further uses and applications.

Significant strides over the past two decades dealing with the challenges of bibliographic control, collection development, archiving, preservation, and open access have contributed to the increased use and application of grey literature. However today, the new challenge is to address its quality control, which is the purpose of this study. One may or may not agree with Carrington when he says that mainstream journalism took the climate change report out of perspective; however, in so doing it brought grey literature into the mainstream press – something that the grey literature community itself could not have done in such a warp tempo and to which we as an S&T community are forever grateful.

3. Method of Approach

The method applied here perhaps best fits the term explorative study, one that relies on a mix of methods involving literature review and comparative analysis. Metaphorically speaking, in an explorative study, a researcher anticipates being at the top of a stairwell at the close of ones study; however more than often, during the course of the study the researcher begins to take steps up as well as downward and by the study's close, the step on which the research ends forms the baseline for the beginning of a more formal research. This metaphor is characteristic of my expedition into the review process carried out by grey publishers.

4. Literature Review - Sourcing and Referencing

Some 15 references on peer review were found in GreyNet's Collection of Conference Papers in the OpenSIGLE Repository. Only three of which went beyond mention of the term in relation to grey literature. This was research carried out by Bertrum MacDonald [2004], where in his GL5 conference paper he recommends that an outline of the peer review process appear in each published report; in the GL7 conference paper by Marcus Banks [2006] where he touches on the waning primacy of peer review; and in the GL7 conference paper by Markus Weber [2006] in which he discusses a specific quality assessment system carried out at the Swiss Federal Office of Public Health - an area to which he returns in a chapter published in 2010.

In my literature search, I even came across GreyNet's early use of the PEER Module - Public Enterprise in Editing and Review - [Farace, 1994] implemented from 1993-2000. However this module was specifically used in compiling and editing publications and did not actually address the review process - different from the aim of this current study.

Further, a number of search engines and alert services were used to gain access to the available body of literature on peer review, which is staggering. Content from some 30 referenced sources were in various degrees used. Half of which were significantly relied upon.

Early in the literature review it became evident that the term peer review itself had many different uses. While this study is specific to publishing, the term is also applied to grant reviews, hospital review for cardiac procedures, drug tests, etc. In its narrowest context related to publishing and dating back to the 17th Century, peer review dealt with journal and journal articles. However, it also became clear in this study involving grey literature that a multitude of document types and a wide range of stakeholders would have to be considered in order to achieve transparency in the review process. Nevertheless, journal and journal articles provide a tangent in the comparison of peer review carried out by commercial and grey publishers. This is in part based on the results of a citation analysis carried out by Schöpfel et al. [2004], establishing that journal and journal articles are not only types of grey literature, but that these document types are increasing in use and application compared to reports, and other traditional grey literature documents.

In every literature review, a researcher hopes to find that one star publication, which provides a state of the art on the topic under investigation. Fortunately, that first star publication came early in my study – the title of which is "Peer Review and the Acceptance of New Scientific Ideas: A Discussion paper from a Working Party on equipping the public with an understanding of peer review" published by Sense about Science [2004]. Interesting enough, this fifty page paper completely circumvents the term 'grey literature', perhaps due to the fact that the study was heavily sponsored by commercial publishers such as Elsevier and Blackwell? Or, because the term grey literature had not yet appeared in Miriam Webster's

Dictionary [2005] http://www.merriam-webster.com/dictionary/gray%20literature, Wikipedia http://en.wikipedia.org/wiki/Gray_literature, or in the Encyclopedia of Library and Information Sciences [2009] http://www.informaworld.com/smpp/title~content=t917508581.
Later on in the literature search, two other articles proved of special interest to this study because of their consideration of alternative models for peer review. These will be discussed later on in this paper.

5. Literature Review - Mining for Key Terms

Nearly 100 terms were derived from the literature on peer review. Some are very specific, others general, and others appear to have been newly coined. I refer to this phase in the review of the literature as mining for key terms. An endeavour which began simultaneous with the full-text reading. This approach seeks to establish a common vocabulary of terms in order to better understand the topic under investigation. The terms identified were then alphabetically indexed with scope notes when applicable, and page numbers and/or links to the actual sources. Well after this index had been drafted, I came across an article in Nature.com in which the IAC recommended that in order "to restore some of the IPCC's lost credibility, … consistent terminology be used, especially when taken from 'grey literature' that has not been peer reviewed." The dependent clause in this statement clearly insinuates that not all grey literature is peer reviewed, which then allows for the clause to be restated to read that the whole of grey literature cannot be treated as non-peer reviewed.

6. Literature Review - Classification of Terms

In the final phase of the literature review, the key terms entered in the index were then classified into five main categories. These categories deal with

1. The **criteria (functions)** of peer review, including such terms as Validity, Significance, Originality, etc.

2. The **roles (stakeholders)** in the peer review process, including such terms as Editor, Reviewer, Publisher, etc.

3. The **attributes (characteristics)** of peer review, including such terms as Anonymous, Formal, Evidence based, etc.

4. The **process (steps)** in the review of publications, including such terms as Pre-screening, Gate keeping, Rejection, etc.

5. And, the **results (impact)** that peer review has on publications, including such terms as Plagiarism, Tenure, Sales increase, etc.

This attempt at classification not only allowed for bringing order into the vocabulary surrounding peer review, but also enabled a selection of terms from the index that best serve in the comparative analysis, which follows.

7. Comparative Analysis across Publishing Channels

The claim [Sense about Science, 2005] that unpublished research is no help to anyone is certainly supported by grey publishers. Having compiled a vocabulary of terms from the literature on peer review and having classified the terms into categories, we can now begin to compare their intended use and relationship within and across commercial and grey publishing. It is not my purpose here to establish a one-on-one relationship of terms, but instead to gain a better understanding of peer review in these two channels of publishing by focussing on five basic categories of terms related to peer review.

Peer review is a procedure, which is applied by thousands of commercial publishers worldwide. It is safe to assume that the actual procedure applied in and among these publishers is not identical or consistent. Chang and Aernoundts [2010, p.4] state that "The actual peer review processes can differ in practice" … "some journals focus on the significance while others more on the methodology". It is perhaps even safe to assume that a comparison of peer review by journals would be spread on a continuum from strict to semi-controlled. Chang and Aernoundts [2010, p.4] go on to say that "Getting all the different journals to agree to apply the same standards and criteria is likely going to be difficult, if not impossible". All the more for grey publishers. Perhaps grey literature can be understood as publications that remain in the review process, where they have been ratified rather than validated. Peer review in commercial publishing is a stamp of approval; while in grey literature it is part of scholarly communication.

It is unacceptable to label grey literature as inferior without reservation and consideration to the organizations who produce, publish and thereby ratify it. In grey literature, while the terms editor, reviewer, publisher, etc. may not always be expressly named as such their imprint on the publication is indelible.

For Cohen [2010, p.2] anonymity can help prevent personal bias, but it can also make reviewers less accountable. The reviewer or referee in grey literature need not be anonymous. They would be more willing to have their name published in the work they review. They like in journal publishing [Peer Review Survey 2009] consider their work of reviewing as a contribution to the community to which they belong. The 'review process' in grey literature is a broader more informal term, because of its coverage of numerous and diverse document types, where authors and corporate authors open their works-in-progress to subject based communities. Grey literature cannot be confused with untested opinions and speculation, but is based on empirical fact and findings and the producers are prepared to explain the standard of evidence on which they are based. In another peer review survey carried out by the Academy of Learned Societies for the Social Sciences [ALPSP 2000, p.3], when the respondent was asked "does your review process differ for different types of articles" the answers were 61 (32%) affirmative and 128 (68%) negative. I would venture to say that for grey literature this ratio would be just the opposite.

Grey literature has significantly contributed to the open access movement and as such has bolstered the public's trust in science. Grey literature lends itself to scientific and scholarly communities, because it's review process is self-correcting. In each stage along the way it is improved. Editorial 'pre-screening' does not only take place for journals but for much of grey literature. Peer review in commercial publishing takes place prior to publication; while in the grey circuit, it is seen as an integral part of the publishing process.

At times, commercial publishers are faced with the dilemma of disclosure or withholding Information, where they have to consider the damage of bad news or the promotional opportunities of good news. Grey publishers venture to publish negative results, which commercial publishing may consider necessary to screen. Grey publishers expose such results neither in a tabloid fashion nor in furore but in the spirit of science. Grey publishers would no doubt side with Rehmeyer [2010], when she writes that a negative answer would likely give a fundamentally deeper understanding of the nature of a subject area. In the discussion paper by the working party Sense about Science [2004, p. v] it is explicitly recommended that if companies are immediately obliged to report R&D results and cannot peer review their publications, they should at least produce a 'best practice guide'. Something that grey publishers would do well to consider.

8. Comparative Analysis - Publisher Guidelines

In June of 2010 the first contact was made with grey and commercial publishers requesting a copy of their "peer review guidelines". For Grey publishers, GreyNet's Distribution List was used and the Questia List http://www.questia.com/aboutQuestia/partnersPub.html was used in contacting commercial publishers. In September, after the summer holidays, a reminder was sent out.

It was expected that if grey publishers had a set of guidelines, they would be apt to respond. However, on the other hand, it was uncertain whether commercial publishers would respond to such a request. In order to adjudge the predisposition of commercial publishers to the request for their peer review guidelines, three were phoned in advance to inquire if they would be willing to do so? Two were willing and did so, and the third stated that peer review was clear and needed no further explanation for their reviewers.

The overall response from both the commercial as well as grey publishers was minimal. Three respondents from GreyNet's Distribution List of over 1000 email addresses and four responses from 30 out of 200 publishers on the Questia List were received. The guidelines that were received varied from extremely detailed to more or less a checklist of items incorporated in a standard form letter. Hence, results are inconclusive. Nevertheless, I would like to mention a few things that stood out in the brief analysis - namely, commercial publishers mention in their guidelines the amount they remunerate reviewers; there is always the question to reviewers as to the readership or market for the manuscript; and those commercial publishers, who submitted their guidelines voiced an express interest in receiving the results of this study. I would even venture to guess that without the initial phone calls, there may not have been a response from commercial publishers. A factor that would have reckoned with in a more formal research on peer review.

In mid-October, an IPCC Task Group published Notes on the Review of IPCC Processes and Procedures [Stocker, 2010], where the use of grey literature is clarified. Since the Intergovernmental Panel on Climate Change's guidelines for peer review is what spurred my study in the first place, I chose to include it in this section on Publisher Guidelines. Citing from the IPCC Task Group's Review, we read that "Much of science today relies on grey literature, and in fact some fields or areas of science must rely on grey literature; however, its [the IPCC] review process and rules on how grey literature is used should become more transparent". Interestingly enough, the Czech Delegation in the IPCC Task Group would rather ban all grey literature from IPCC publications, even when brought to their attention by other country delegates that among grey literature includes reports created at the government level sent through the legislative process, which is therefore "peer-reviewed" by individual Parliaments. Sources that may often be more valuable than an article that is published in a journal on a commercial basis.

From my brief analysis, I came to understand that for an explorative research of this nature, a distinction should have been made between the request for guidelines as opposed to a check list issued to reviewers - where the latter should have taken precedence. In the ALPSP survey [2000, p. 6], when asked if referees are provided with a checklist with which to carry out their peer review, 153 (78%) responded affirmative and 44 (22%) negative.

9. Comparative Analysis - Models for Peer Review

During the course of study, a change in direction took place. Instead of including self-publishing alongside commercial and grey publishing in the comparative analysis, the focus of attention turned to alternative models of peer review in comparison with the traditional journal model. Two alternative models were discovered. The first, in an article by Patricia Cohen [2010] dealing with an Open Peer Review Model and the second in an article by Chang and Aernoudts [2010] discussing their proposed peer-to-peer review model. In both articles no explicit mention of the term grey literature appears.

Open Peer Review Model

For Cohen [2010] the traditional (peer review) process is not so much a gold standard but an effective accommodation to the needs of the field. It represents a settlement for the particular moment, not a perfect ideal." She continues in her discourse by saying that "the goal is not necessarily to replace peer review but to use other more open methods as well."

Cohen draws on the term open as opposed to insular, the latter meaning remote or inaccessible - what Bornmann and Daniel [2009] might refer to as part of that black box in commercial publishing. Cohen further introduces the term crowd sourcing related to new stakeholders in an open-peer review model. She also draws attention to the fact "many professors are wary of turning peer review into an American Idol-like competition … worrying that … know-nothings would predominate." And to offset such a suspicion, Cohen explains how open models could count toward academic tenure, and refers to one such case in the Shakespeare Quarterly experiment. Cohen further touches on other scholarly values in an open model besides quality control – these include generating discussion, improving works-in-progress, and sharing information rapidly.

Peer-to-Peer Review Model

Another alternative to traditional peer review is the peer-to-peer review model introduced by Chang and Aernoudts [2010]. Their study arises from the question whether 'raising the awareness of and archiving research' are key elements in journal publishing? Chang and Aernoudts proceed to describe an automated peer reviewer selection system. This automated process begins with the reviewer, who in some capacity is seen as an editor.

In traditional journal publishing, editors do not make the reviewers comments accessible; however, in Chang and Aernoudts' peer-to-peer model such comments comprise the reviewer's impact factor. They then proceed to explain how the use of nicknames could assimilate blind peer review. For grey publishers it should not be the goal of mimicking every aspect of the journal peer review model, but instead to develop parameters that best serve grey literature communities and their review processes.

Chang and Aernoudts also introduce an added document type in their peer-to-peer review process called the Peer Review Report. This report is actually a peer review of the peer reviewer of a manuscript. By this time, it becomes clear that their focus of attention is on whether a peer reviewer is qualified, rather than the quality of the peer reviewed manuscript.

In fine, Chang and Aernoudts see that in the traditional journal model the author is answerable to reviewers, reviewers are answerable to the editor, and editors in turn are answerable to their readership. Whereas, in their proposed peer-to-peer model, editors are third parties who facilitate an automated system in which reviewers answer to other reviewers, to the authors, to the editors, and all other stakeholders in the peer review process. As we are made aware of in the traditional journal model, in the next generation it may be difficult to attract sufficient numbers of reviewers [Sense about Science, 2009]. Would not this concern become even more augmented in Chang and Aernoudts' proposed peer-to-peer model?

10. Some Preliminary Findings

Based on the literature review and the somewhat limited analysis carried out in this explorative study, preliminary findings indicate that

- Grey literature document types far exceed journal articles and require alternative models for peer review;
- Community sourcing lends itself to the review of grey literature;
- Grey literature focuses more on the review process than the end product;
- Grey literature is more apt to include negative results in publications;
- Commercial and grey publishing share more in common with one another than with self publishing;
- The IPCC affaire exposed grey literature to the wider public and defends the further use of grey literature; and
- Guidelines for good practice would serve to enlighten net users of the value of grey literature.

11. Concluding Remarks

The IPCC Affair is in some ways similar to WikiLeaks. Grey literature is out there in sundry formats, distributed via diverse channels, available not only to subject based communities but worldwide to net users. As such, information contained in the grey literature now comes under further scrutiny. It becomes the obligation of the grey literature community to inform the wider public of the quality of its publications. This can be accomplished through further research on the peer review process, enhanced curricula on grey literature in schools and colleges of library and information science, as well as active steps taken by corporate authors and grey publishers in making the review process available in publication, hence more transparent to its readership.

References

ALPSP/EASE (2000), Current Practice in Peer Review: Results of a survey conducted during Oct/Nov 2000. – ALPSP, Association of Learned and Professional Society Publishers ; EASE – European Association of Science Editors, and the Academy of Learned Societies for the Social Sciences, http://www.alpsp.org/ForceDownload.asp?id=140

Banks, M.A. (2006), Towards a continuum of scholarship: The Eventual Collapse of the Distinction Between Grey and non-Grey Literature? – In: GL7 Conference Proceedings.
ISBN 90-77484-06-X

Boekhorst, A.K., D.J. Farace and J. Frantzen (2005), Grey Literature Survey 2004: A Research Project Tracking Developments in the Field of GL. – In: GL6 Conference Proceedings. – Amsterdam : TextRelease, pp. 1-9. – (ISSN 1386-2316 ; No. 6). ISBN 90-77484-04-3

Bornmann, L. and H.-D. Daniel (2009) The Manuscript reviewing process: Empirical research on review requests, review sequences, and decision rules in peer review. – In: Library & Information Science Research, 32 (2010) 5-12.

Carrington, D. (2010) Climategate shows the need for openness by scientists. - In: The Observer, Sunday 11 July 2010 http://www.guardian.co.uk/environment/2010/jul/11/climategate-muir-russell-review

Chang, C. and R.H.R.M. Aernoudts (2010), Towards Scholarly Communication 2.0: Peer-to-Peer Review & Ranking in Open Access Preprint Repositories. Available at SSRN: http://ssrn.com/abstract=1681478 (Working Papers Series).

Clover, C. (2010), Sloppy science is seeping into the climate watchdog
http://www.timesonline.co.uk/tol/comment/columnists/guest_contributors/article6999815.ece

Cohen, P. (2010), Scholars test web alternatives to peer review. – In: The New York Times, August, 23 2010 http://www.nytimes.com/2010/08/24/arts/24peer.html

Farace, D.J. (1994), The PEER Module: Public Enterprise in Editing and Review. – Presentation at the 5th IASP, Thessaloniki, Greece – May 6-10, 1994

MacDonald, B.H., R.E. Cordes, and P.G. Wells (2005), Grey Literature in the Life of GESAMP, an International Marine Scientific Advisory Body. – In GL5 Conference Proceedings, pp. 50-63. – ISBN 90-77484-01-9.

Nature.com (2010) http://www.nature.com/news/2010/101019/full/467891a.html

Rehmeyer, J. (2010), Crowdsourcing peer review. – In: ScienceNews, web edition http://www.sciencenews.org/view/generic/id/63252

Schöpfel, J., C. Stock, D.J. Farace, and J. Frantzen (2005). – Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit. – In: GL6 Conference Proceedings, pp. 55-63.
ISBN 90-77484-04-3.

Sense about Science (2004), Peer Review and the Acceptance of New Scientific Ideas. – Discussion paper from a Working Party on equipping the public with an understanding of peer review. http://www.senseaboutscience.org.uk/pdf/PeerReview.pdf

Sense about Science (2009), Peer Review Survey 2009: Preliminary Findings http://www.senseaboutscience.org.uk/index.php/site/project/395

Stocker, T. (2010), Review of IPCC Processes and Procedures: Notes on the Informal Task Group on Procedures. – Thirty-Second Session of the IPCC in Busan, 11-14 October 2010

Weber, M. (2006), Grey Literature in Public Administration: An Example of a Specific Quality Assessment System. - In: GL7 Conference Proceedings. ISBN 90-77484-06-X

Weber, M. (2010), How to assure the quality of grey literature: The Case of Evaluation Reports. – In: Grey Literature in Library and Information Studies, pp. 29-38. – Berlin : De Gruyter, 2010. - ISBN 978-3-598-11793-0

Acknowledgements to Dr. Joachim Schöpfel and Jerry Frantzen in their help in the search and retrieval of literature throughout the study and to Dr. Leonid Pavlov, CITIS (RU), Markus Weber, SFOPH (CH), Aarhus University Press (DK), Amsterdam University Press (NL), Manchester University Press (UK), and the World Bank (USA) for submission of peer review guidelines.

An Analysis of Current Grey Literature Document Typology

Petra Pejšová,
National Technical Library, Czech Republic

Marcus Vaska,
University of Calgary, Canada

Abstract
This analysis is based on the classification of the international systems 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). During the analysis of the lists of document types, we have discovered that these typologies contain, besides "real" document types (reports, theses, etc.) other aspects, such as events (arrangement, organization), types of events (conferences, speeches), producers (universities, institutes), processes (translations, output), content (political documents, legal texts), location (domestic, foreign), and format (e-texts, numeric data). However, this approach is not systematic. Therefore, we have decided to create a classification scheme for document types only, and classify other aspects into various groups in order to define them more precisely. The scheme will be processed in a text version as well as schematically in mind maps.

We believe that identifying a specific typology for credible grey literature document types, particularly reports, conference proceedings, and government documents, will assist in the classification of grey literature in the fields of science, research, and education. On the other hand, grey literature also consists of various means of communication, such as telephone calls, meetings, e-mails, blogs, interviews, social networking tools, or discussions in Wiki. It is important to identify only credible document types and not use unverified information that may be unsuitable for scientific work.

The aim of this analysis is therefore 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.

Keywords: analysis, classification, documents, gray/grey literature, systems, types, typology

Introduction: Defining Typology in the Grey Literature

> "Improved access to and sharing of research information is the key to accelerating progress and breakthroughs in any field" (Brian Hitson and Lorrie Johnson, 2009)

Indeed, continued and concentrated efforts in the pursuit of grey literature has caused a transparency and ability to share documents that only a few years ago would not have been deemed possible. Despite the wealth and variance of forms of grey literature (including multimedia) in institutional repositories, there appears to be a lack of systematically classifying these documents into a universal, standard typology. In fact, Beissel-Durrant (2004) states that the typologies currently in place in the social science literature "do not necessarily categorize research methods in a systematic way, using mutually exclusive categories and hierarchies that are not necessarily complete" (p.2). An exhaustive search is therefore required to eliminate bias, as the presence of bias could potentially undermine the research retrieved.

Traditional monikers assigned to grey literature have labeled this material as mainly consisting of primary sources, focusing on theses, reports, and government publications. This leads to two key characteristics of grey literature resources, namely their ubiquitous nature, and their difficulty in being properly identified (Schöpfel, 2006). While the history of grey literature may have supported this notion, the same does not necessarily hold true today; Hjorland (2006) claims that "each sphere in society has developed its own kinds of documents." Hence, conducting an analysis of current grey literature document typology and subsequently creating and implementing a quality control system to guarantee the credibility of grey literature becomes increasingly important.

The word *typology*, first introduced in the Merriam-Webster dictionary in 1845, has origins that date back to biblical times. Originally defined as "a doctrine …holding that things in Christian belief are prefigured or symbolized by things in the Old Testament", its current association refers to a "study of or analysis or classification based on types or categories" (Merriam-Webster, 2010). Although various disciplines (i.e. anthropology, archaeology, psychology, and in particular linguistics), have a different notion of how typology fits in with their subject area, the idea of classification and organization is the same. This characteristic is significant when it comes to identifying the role that document typology plays in the realm of the grey literature.

Is typology for grey literature really irrelevant? Some researchers claim it just might be, particularly as the grey literature is moving closer to the white (Di Cesare, 2006). As the numerous document types in the *GreyNet* repository can attest to, grey literature truly does transcend boundaries and plays a role in more than merely science, research, and education. When categorizing the various types of grey literature into a single, universal typology, it seems plausible to see the wealth and versatility of grey literature sources as an information neighborhood, "an environment within which practical information seeking and orienteering information seeking, as well as both directed and undirected browsing, can take place." (Burnett, 2000) The representation of grey literature in numerous types and formats can indeed create the appearance of numerous aspects of grey literature that do not appear to hold a common purpose with each other, hence the need for a standard typology for this material to put everything in its place.

Aspects & Analysis of Grey Literature Typologies

In preparation for the analysis of a grey literature document and subsequently its typology, it is necessary to consider the purpose of the material being studied, the aspect of the grey material in question, and perhaps most importantly, "how are grey documents used?" (Schöpfel, 2009, p.7)

This paper, and subsequently the grey literature typology proposed, will focus on the following aspects: document type, event, producer, content, location, format, and periodicity. While theses, dissertations, reports, conference proceedings, working papers, and to a lesser extent, courseware, are considered fundamental types of grey literature (Schöpfel, 2009), other aspects or 'shades' of grey' must also be considered. The purpose of a document typology for grey literature adheres rather closely to Beissel-Durrant's (2004) notion of prioritizing research methods, exemplifying the focus of the research, identifying needs for additional training or research, as well as appropriate classification methods.

Creating a Grey Literature Typology Classification System

We collected a total of 241 terms used to describe grey literature typologies; these terms have been categorized into one of six system typologies as follows:

- 133 terms from GreyNet
- 35 terms from NRGL
- 30 terms from ASEP
- 17 terms from OpenSIGLE
- 14 terms from RIV
- 12 terms from ROAR

After removing duplicates, we obtained a final list of 193 original terms. While no single term was present in all six typologies, two terms, namely theses and research reports, appeared in five of the analyzed system typologies (theses in GreyNet, OpenSIGLE, ASEP, NGRL, ROAR and research reports in GreyNet, RIV, ASEP, NGRL, ROAR), while 28 terms occurred twice, and six terms materialized three times. It is interesting to note that four instances were observed where no common term was found among any of the system typologies studied.

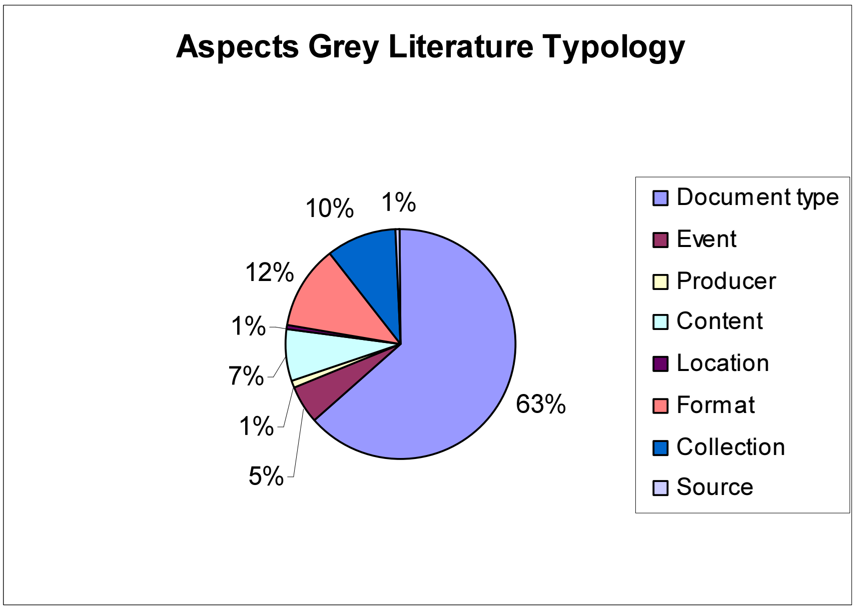
Our examination of the various aspects in grey literature typologies ascertained 121 document types which we subsequently organized into 19 collections. For example, a collection labeled with the broader term ***REPORT*** is comprised of narrower term document types such as annual report, business report, bank report, and so on. The 19 collections identified as narrower term document types are listed in a mind map appearing at the end of this paper (Appendix 1).

In addition to document type, we have proposed six additional aspects of GL typologies in order to more succinctly classify the GL terms we analyzed. These include:

- **Format,** describing the type of presentation (e.g. electronic document, e-text, multimedia, and so on). – 23 terms
- **Content**, referring to the type of information in the document (e.g. computer program description, policy document, product data, and so on) – 14 terms

- **Event,** depicting the occasion on which the document was issued (e.g. conference, workshop, lecture) – 10 terms
- **Producer,** denoting the organization producing the document (e.g. legislation) – 2 terms
- **Location,** indicating the place where the document is situated (e.g. board) – 1 term
- **Source**, representing the source data for the document (e.g. survey) – 1 term

The graph below illustrates the occurrence of the 193 original terms used to describe grey literature amongst the proposed aspects in GL typologies that have been identified.

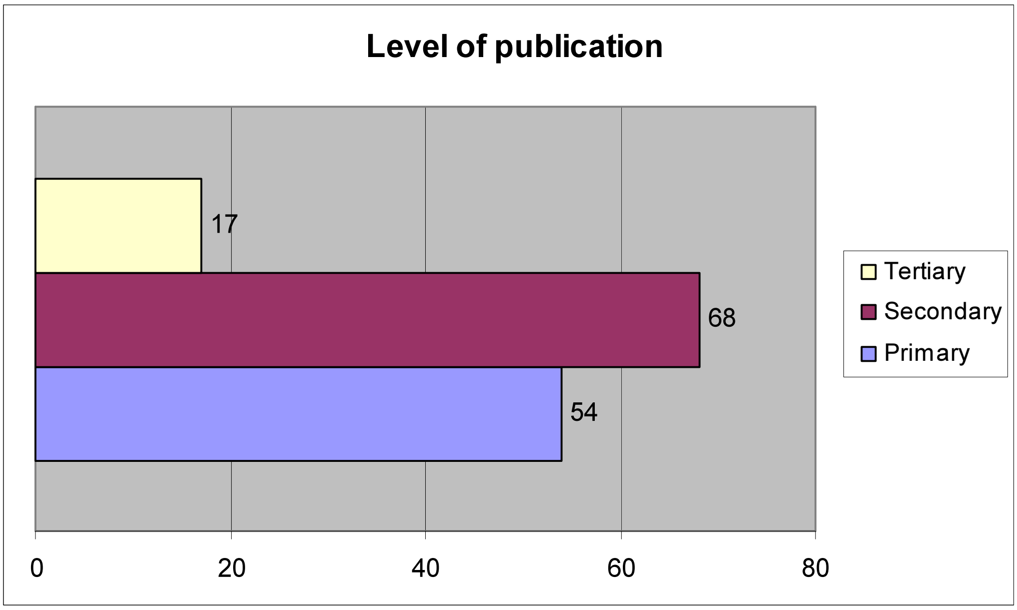


Many of the analyzed terms contain more than one aspect, such as a project information document, (comprised of both and information and a project document), a computer program description, etc. We also identified 44 terms that expressed a similar meaning, requiring us to select preferable terms in order to maintain consistency amongst our classification scheme. For instance, *annual* was deemed the preferred term when describing a yearbook or other such yearly document, while *leaflet* was favored over flyer when classifying items of an advertising nature. However, some terms were difficult to analyze, due to lack of description or unknown reasons as to why they were introduced. This was particularly indicative of website reviews, where questions arose whether the material was a review of a website or rather a review presented on a website. We therefore require and look forward to comments and feedback from the grey literature community on these matters.

Distribution terms to broader (Collections) and narrower terms document types is a basis for creating a new and well-structured grey literature typology. The National Technical Library in Prague will prepare a first draft version that will be available for feedback from the grey literature community. After incorporating the comments received, a subsequent step will implement the aspects identified in the analysis as an optional description of the document types. These optional descriptions are necessary to precisely express the document type without having to constantly introduce new types of documents, thereby unduly expanding the typology of terms that are used only for specialized areas. This approach is necessary for creating a functional general grey literature typology.

In addition, we identified three levels of publication in our analysis, according to the Library of Congress specifications for identifying primary, secondary, and tertiary sources, and Hjorland's (2006) revised version of the UNISIST typology model used for classifying scientific and technical documents :

- 54 primary
- 68 secondary
- 17 tertiary



At this point, this data is being used solely for informative purposes; there are no plans to include levels of publication for creating the first version of a new grey literature typology.

Implementing a Quality Control System to Guarantee Credibility of Grey Literature

While the existence of any electronic media or non-traditional resource should be analyzed for its acceptance as a grey literature document, the adage of effective searching, whereby key websites and the Internet in general are consulted, needs to be considered as well (Giustini, 2010). Social media, blogs, phone, fax, e-mail; all of these forms of grey literature can greatly enhance the searching process. The recent H1N1 epidemic indicated the importance that common social networking sites, primarily Twitter, played in disseminating information in a timely, virtually instantaneous manner. While social networking tools or Wiki contributions should not be relied on exclusively, "the key is not to rule anything out, not even tweets" (Giustini, 2010).

Although the rapid "information-at-your-fingertips" approach is impressive, there is a danger of accepting unverifiable information as fact without further investigation. A case in point is an article that appeared in the United Kingdom's *Daily Mail* newspaper earlier this year (Rose, 2010). Climate change is a global issue, with numerous organizations, such as the World Wildlife Fund (WWF) and the Intergovernmental Panel on Climate Change (IPCC) taking a firm stance and commitment to reducing our carbon footprint and preserving the environment for generations to come. In 2007, a report produced by the IPCC, which was subsequently awarded the Nobel Prize, claimed that the Himalayan glaciers would melt by 2035 (Rose, 2010). Despite repeated objections by some glacial experts as to the accuracy and authenticity of this report, the claim was not refuted until January 23, 2010, when Dr. Murai Lai, a scientist at IPCC, admitted that the statement was "included purely to put political pressure on world leaders." Further investigation revealed that no peer-reviewed scientific research had been carried out to support the Himalayan glacier melting assertion; "the 2035 melting date seems to have been plucked from thin air", and was due to an arithmetical error by the WWF. Unfortunately, this error not only damaged the reputation of the IPCC, it also questioned the quality and qualifications of those producing grey literature, particularly since this material is often not peer-reviewed.

Content contained within open access repositories still prefers theses and dissertations as key primary material. Other grey literature documents, such as conference papers, reports, even works in progress, are slowly buckling the trend; this is comforting, especially since Schöpfel (2009) reports in his study that "100% of the institutional archives give access to grey material." (p.15)

Lack of bibliographic control is the primary reason why grey literature can be difficult to locate; it must be easily retrievable in order to be useful. While most repositories, including GreyNet and OpenSIGLE, are making conscientious efforts to classify the material they store, there are others that still do not do so.

Despite the dissolving of EAGLE in 2005, and SIGLE's dormancy, some researchers believe that each association will classify its documents in-house, without adhering to any bibliographic standards. However, this may not necessarily be as troubling as originally thought: indeed the rapid expanse of the Internet has enabled increasing numbers of grey literature documents to be made available to the public for the first time, which puts additional pressure on cataloguing this material in a uniform matter.

Nevertheless, the Web has also created greater awareness that these types of material exist and can be retrieved. The trade-off between access and awareness is a never-ending challenge.

Evaluating Grey Literature Document Types

Presently, open access repositories of grey literature are maintained and/or funded by either an academic institution (typically a University), or by a pubic research association (Schöpfel, 2009). As the analysis of the following types of documents found in repositories will indicate, a majority of these storehouses of information centre on more than one domain, often playing a multidisciplinary role. Barely seven years old, the notion of Open Access, whereby scholarly output is freely available to the general public, has taken academia by storm. Focusing on institutional repositories and various policies regarding unrestricted access, "the GreyNet community intensified its research activities on the impact of the open access movement on the grey literature." (Schöpfel, 2009, p.4)

Numerous studies have been launched in an effort to determine why research is primarily disseminated via a report, thesis, or conference proceeding, despite the various types and formats that exist today. Schöpfel (2006) ponders this dilemma and provides his own reasons: "research results are often more detailed in reports, doctoral theses, and conference proceedings than in journals…they are distributed in these forms up to 12 or even 18 months before being published elsewhere. " (p. 68)

GreyNet

While the aim of this paper is to analyze, identify, and create a typology for credible types of grey literature, namely reports, conference proceedings, and government documents, the impact of virtual communities cannot be discounted. The notion of creating awareness is often commented upon in papers on grey literature, and social networking tools or Wiki collaborations certainly have a role to play in this pursuit. Announcements, one of the aspects categorized in the *GreyNet* repository is considered to be a common activity in online communication, with some even saying that these information updates "play a significant role in the informational economics of the community" (Burnett, 2000). Information is not just given away; it becomes invaluable to the person seeking it.

Founded in 1992, The Grey Literature Network Service, or GreyNet as it is commonly called, is "dedicated to research, publication, open access, and education in the field of grey literature" (GreyNet, 2010). For nearly two decades, this organization has strived towards seeking, identifying and disseminating grey literature to as wide an audience as possible. Recent technological advances and increasing acceptance and adherence to the Open Access movement have strengthened the awareness of the importance of grey literature among several disciplines. GreyNet has certainly achieved a number of milestones in a relatively short period of time; its goal of facilitating "dialog and communication between persons and organizations in the field of grey literature" will undoubtedly grow exponentially in the coming years.

In 2004, GreyNet developed a GL Survey, whereby visitors to the site were invited to contribute to a list of document types that they felt best constituted grey literature, the purpose of which was to best describe the type of document it embodies. As a result, 133 different document types in grey literature have been identified.

OpenSIGLE

OpenSIGLE, the System for Information on Grey Literature, functions as a repository for the collection of scientific, technical, economic, and humanities documents produced across Europe. The 13 member countries include Belgium, the Czech Republic, France, Germany, Hungary, Italy, Latvia, Luxembourg, Portugal, Russia, Slovakia, Spain, and the United Kingdom (OpenSIGLE, 2010). The partnership between GreyNet and OpenSIGLE has ensured that preprints, PowerPoint presentations, abstracts, and biographical notes from previous international conferences on Grey Literature are included. OpenSIGLE is thus growing at a remarkable rate. This past year, "700 000 records of the unique European database on grey literature SIGLE migrated to an open access environment" (Giustini, 2010).

SIGLE, the predecessor to OpenSIGLE, assigned nearly 96% of its contents to one of three categories: reports, theses, and conferences. While this classification supports the traditional definition of grey literature along with the most common types of grey material, it can be problematic when creating, analyzing or defining a grey literature typology. For instance, there are several subcategories of reports, ranging from those produced by institutions to annual reports to logs generated by a specific activity. Further, Schöpfel (2006) argues that the theses and conference proceedings distinction fails to take into account unpublished manuscripts, newsletters, presentations, working papers, preprints, lecture notes, and even personal communications. These documents, regardless of the format they may be presented in, are all types of grey literature, and need to be distinguished as such.

Registry of Open Access Repositories (ROAR)

Founded in 2003, ROAR's key role is providing information concerning the growth and status of open access repositories around the world. Consisting primarily of dissertations and preprints/postprints of peer-reviewed articles, (Digital Library Federation, 2005) ROAR also contains documents in a wide-variety of formats, including multimedia archives. Offering the user an opportunity to present material to the Editorial Review committee for possible inclusion in the repository, ROAR is growing at an exponential rate; as of November 6, 2010, there are a total of 1988 items in the repository, organized into one of 9 repository types (Registry of Open Access Repositories, 2010).

National Repository of Grey Literature (NRGL)
In 2008, a four-year study, *The Digital Library for Grey Literature: Functional Model and Pilot Implementation,* the goal of which is to create a National Repository of Grey Literature (NRGL), was launched. The NRGL originated as an idea in 2005 due to the termination of SIGLE, which greatly affected the state of grey literature in the Czech Republic. Supported by the National Technical Library in Prague (NTK), this project's main goals are the systematic collection, long-term archiving and provision of access to specialized grey literature, especially with regards to research and development, civil service, and education, as well as from the business sphere, marketing "open access" at the national level. To support this goal, the NTK created a network of partner organizations, a functional 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. These consist primarily of a recommended metadata format, exchangeable designs and templates, examples of licensing models and legal issues, resolved preservation, methodology, archiving, and the provision of access to digital data (National Repository of Grey Literature, 2010). Further details on this project can be found on the NRGL website, http://nrgl.techlib.cz.

Since the end of 2009, a NRGL central user interface has been available to search for grey literature in the Czech Republic. This NRGL central search interface offers a user-friendly system for searching data thanks to data visualization and dynamic contextual navigation. All institutions in the NRGL network are gradually integrated into this interface. At the end of November 2010 there were over 51 000 grey literature records. The interface is available at www.nusl.cz .

Information Register of R&D results (RIV)

RIV is part of the R&D Information System in the Czech Republic. Since 1993, RIV has collected information about the results of R & D long-term intentions and projects supported by different state and other public budgets.

The data available in RIV has been made possible by contributions from public sponsors, namely different ministries and other state offices with the responsibility for a state-run R&D long-term intention and/or R&D project, providing financial aid. This includes the Grant Agency of the Czech Republic, the Academy of Science of the Czech Republic, and local authorities (Research and Development Council, 2006).

RIV refers to the conveyance of data to an informational research system, experimental development, and innovation. According to a disclaimer or "law", the support of experimental development is provided only under the supposition of the truthful publication of pertinent data. As such, RIV contains information about all research results. Nevertheless, RIV does maintain its right to remove information about any results forwarded by experimental institutions, in the event that the data are incorrect, or otherwise deemed inadmissible. In addition, the general terms available for a description of data for RIV include: central evidence of activities, research activity, provider, receiver, other participants, proposal from a research organization, creator, etc.

Register of Publication Activity of the AS CR (ASEP)

The ASEP system is produced under the auspices of the Library of the Academy of Sciences of the Czech Republic. Together with RIV, types of publications are listed according to their form and incorporation. These include monographs, conference contributions, dissertations, electronic documents, conference volumes, temporary publications, articles in a professional periodical, prototypes, norms and rules, specialized maps, certified methods, software, chapters/sections of books, newspaper articles, patents, reviews, translations, workshops, exhibitions, research reports, and several others. The ASEP system contains bibliographical records concerning research results at institutes of the Academy of Sciences of the Czech Republic from 1985. The ASEP system publication records are also sent to the RIV database (Evidence publikací v AV ČR, 2010). The User interface of the ASEP system is available at www.lib.cas.cz/en/ASEP.

Concluding Thoughts & Future Directions
"Grey is global…grey is growing…grey is good." (Hitson, 2009). Without a doubt, grey literature is here to stay, with the border between grey and white becoming more and more transparent, in response to the increasing number of grey material being posted on the Web. As the above arguments suggest, there is not one set rule of classifying and organizing the grey literature. The typology that has been suggested in this paper is merely one notion of how a typology for this type of material can be implemented; it is

certainly not the only one. Nevertheless, subjecting a piece of grey literature that is neither a theses nor a conference paper or report to a 'miscellaneous' or 'other' distinction makes identifying and gathering grey literature that much more challenging.
Implementing a quality control system to guarantee the credibility of grey literature is of vital importance, despite the many challenges.

Following this analysis, we established an international Grey Literature Typology Working Group for creating a grey literature typology. The six-member group consists of experts in grey literature, ontological engineering, data modeling, and knowledge organization systems. The Google project platform CODE, a tool for collaborative development, was used for this study. Information about the activities of the working group as well as a link to the grey literature typology project can be found on both the Czech NRGL and GreyNet websites.

A key goal of defining and describing grey literature document types has already been partially addressed; the proposed grey literature typology discussed in this paper has been transformed into open standard machine-readable format as an open available web application, the details of which are available in *Publishing the Vocabulary of the Types of Grey Literature as Linked Data,* a poster presented at the GL12 conference. "The typology of grey literature will be a controlled vocabulary in RDF (Resource Description Framework) expressed as SKOS (Simple Knowledge Organization System) concept scheme. This description of the document types of grey literature has a loose structure with hierarchical relations. Each type will be provided with a definition and a prototype example of a document for which it can be used. By design, it 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)

The first draft of a grey literature typology will be issued in January 2011. The online platform CODE, http://code.google.com/p/grey-literature-typology/, will be open for comments from the grey literature community until the end of March 2011. The Working Group will analyze and respond to comments regarding a clear and structured typology, and compile the materials for incorporation into a draft grey literature typology by the end of May 2011. This recommendation will be included in the first version of a grey literature typology in the SKOS concept scheme, to be published on June 30, 2011. The Development Cycle will be conducted bi-annually, and all versions will be issued via the Web as sustainable linked versions. Future plans and directions will involve the translation of the proposed typology into other languages.

Most in the grey literature community would undoubtedly agree that the use of this material varies in frequency among different disciplines; some subject areas make use of it on a daily basis, while others have not yet skimmed the surface. Nevertheless, grey literature definitely holds a place in the information-seeking behavior of today's researchers, more so than it ever has before. As the future of grey literature information seeking turns towards alternate formats such as multimedia and datasets, the fundamental purpose of grey literature remains the same: increasing awareness of the grey by opening the information doorway and providing unrestricted access to what lies beyond. As Schöpfel (2006) argues, "however diverse, these documents all have one point in common: they contain unique and significant…information that is often never published elsewhere."

References

Beissel-Durrant, G. (2004). *A typology of research methods within the social sciences: NCRM working paper.* Retrieved September 12, 2010 from http://www.ncrm.ac.uk/publications/documents/NCRMResearchMethodsTypology.pdf

Burnett, G. (2000). Information exchange in virtual communities: A typology. *Information Research*, *5*(4). Retrieved September 12, 2010 from http://informationr.net/ir/5-4/paper82.html

Di Cesare, R., and Ruggieri, R. (2006). *Evaluation of grey literature using bibliometric indicators: A methodological proposal.* Retrieved September 12, 2010 from https://darchive.mblwhoilibrary.org/bitstream/handle/1912/1322/proc06057.pdf?sequence=1

Digital Library Federation. (2005). *Review of resources.* Retrieved November 6, 2010 from http://www.diglib.org/pubs/dlf106/DLF106part3b.html

Evidence publikací v AV ČR: ASEP (2010). Retrieved November 10, 2010 from http://www.iach.cz/knav/databaze_cz.htm

Giustini, D. (2010). *Got app? Top mobile medicine sites 2010*. Retrieved September 12, 2010 from http://blogs.ubc.ca/dean/2010/04/got-app-top-mobile-medicine-sites-2010/

Grey-literature-typology (2010). *Vocabulary of the types of grey literature*. Retrieved November 20, 2010 from http://code.google.com/p/grey-literature-typology/

GreyNet: Grey Literature Network Service (2010). Retrieved September 12, 2010 from http://www.greynet.org/

Hitson, B.A., and Johnson, L.A. (2009). *WorldWideScience.org: Bringing light to grey*. Retrieved September 12, 2010 from http://opensigle.inist.fr/bitstream/10068/697997/2/GL10,_Hitson_and_Johnson,_Conference_Preprint.pdf

Hjorland, B. (2006). *Document typology*. Retrieved September 12, 2010 from http://www.iva.dk/bh/core%20concepts%20in%20lis/articles%20a-z/document_typology.htm

Library of the Academy of Sciences of the Czech Republic (2007). *ASEP – publication records.* Retrieved January 4, 2011 from http://www.lib.cas.cz/en/ASEP

National Repository of Grey Literature (2010). Retrieved November 9, 2010 from http://nrgl.techlib.cz/

OpenSIGLE: System for Information on Grey Literature in Europe (2010). Retrieved September 12, 2010 from http://opensigle.inist.fr/

Registry of Open Access Repositories, ROAR. (2010). Retrieved November 6, 2010 from http://roar.eprints.org/

Research and Development Council (2006). *RIV* : *Information Register of R&D Results*. Retrieved November 19, 2010 from http://www.vyzkum.cz/FrontClanek.aspx?idsekce=1028

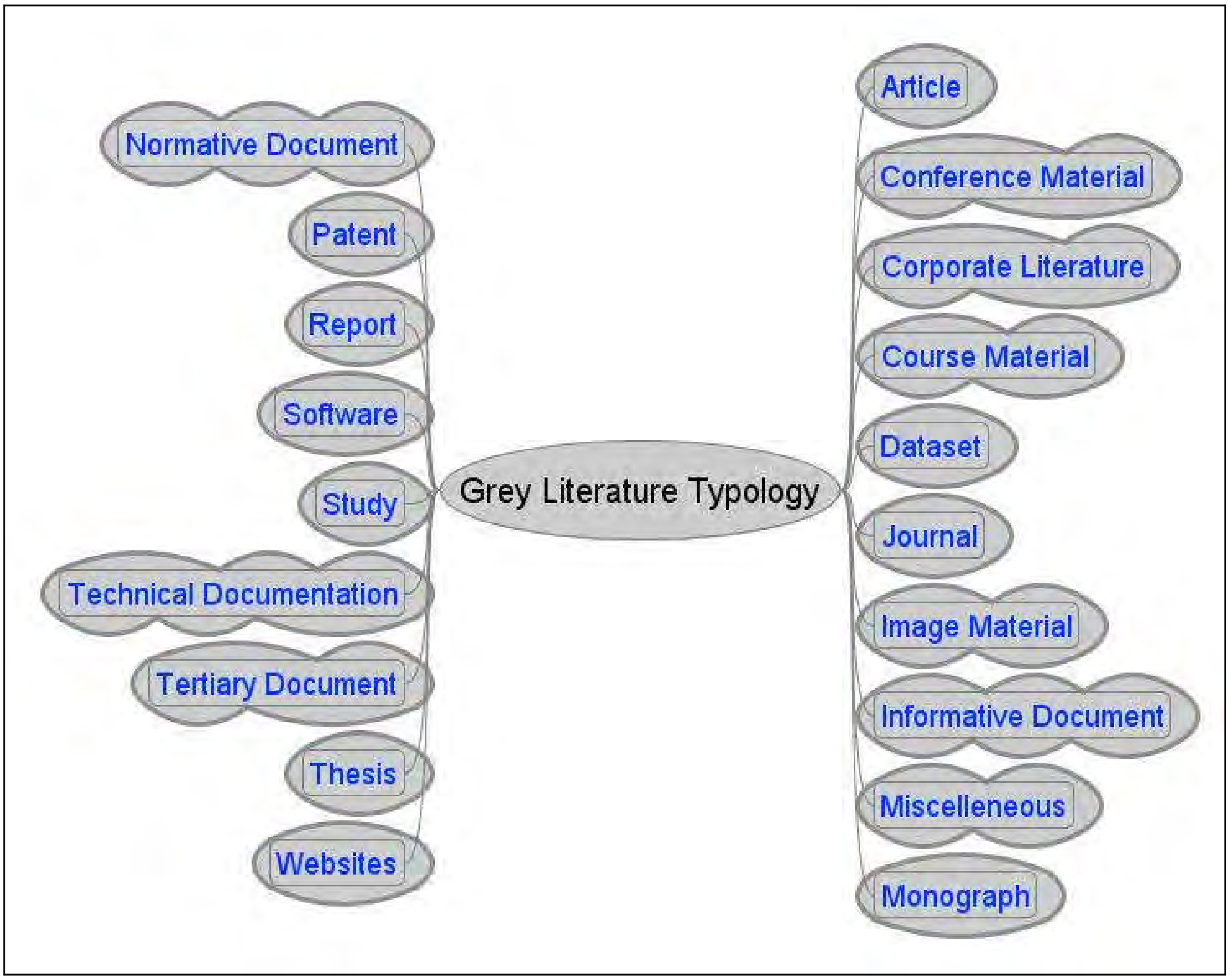
Rose, D. (2010, January 24). Glacier scientist: I knew data hadn't been verified. *Daily Mail Online.* Retrieved November 6, 2010 from http://www.dailymail.co.uk/news/article-1245636/Glacier-scientists-says-knew-data-verified.html

Schöpfel, J. (2006). Observations on the future of grey literature. *The Grey Journal, 2*(2), 67-76.

Schöpfel, J., and Stock, C. (2009). Grey literature in French digital repositories: A survey. Retrieved September 12, 2010 from http://archivesic.ccsd.cnrs.fr/docs/00/37/92/32/PDF/GL10_Schöpfel_Stock_final.pdf

University of North Carolina, Wilmington: Randall Library (n.d.). *Identifying primary, secondary, and tertiary sources*. Retrieved December 31, 2010 from http://www.lib.auburn.edu/bi/typesofsources.htm

Appendix 1: Mind map of the 19 terms identified as narrower term document types in analysis



NYAM

Global Grey Literature in Health: Identification, Preservation, and Sustainability

Janie Kaplan, Lea Myohanen, and Elizabeth Taylor
The New York Academy of Medicine Library, United States

The New York Academy of Medicine (NYAM) Library has been aggregating grey literature in public and global health since 1999 through print and web. The *Grey Literature Report's* primary focus has been on the United States with a rapidly growing interest in the more global environment. NYAM has a growing interest in expanding the global perspective of the report which is motivated by the desire to increase capacity to identify information produced in the global world by non-governmental organizations (NGOs) and international agencies as a result of the engagement of the world in technology. The NYAM Library holds a common concern in sharing best practices and lessons learned via the dissemination of information that is often hard to find and a recognition that the capture of this information and the preservation of it, whether print or digital, may, in many cases, be the only record of documents in the future. The NYAM Library believes that its report is a benefit to developing transparency of grey literature as an "added value" to research in health, and particularly global urban, public, environmental, and social health. Finally, and equally important to all this is the Library's interest in maintaining an alignment with the mission and interests of NYAM.

This paper outlines some first identified steps that the NYAM Library is taking to broaden the coverage of the *Grey Literature Report* to represent global coverage in undeveloped countries. We are focused on the countries of the global south where health issues are critical to future development. Because many of these countries do not have stable municipal infrastructures that allow for traditional approaches to literature identifications, aggregation, and dissemination, NYAM will be developing strategies to identify grey literature in many different ways. We will document our approaches to these new geographic areas covering how we identified where we will begin; the analysis of what we needed to do; and the steps we took to implement this analysis. We will share the workflow that we create from these strategies and the successes and disappointments that may develop from these approaches within our newly expanded geographic focus. This paper relates our first attempt at establishing collaboration with four African universities to collect grey literature in their respective countries.

Background:
NYAM's Grey Literature Project produces a bimonthly report, *the Grey Literature Report*, which has been published on the web since 1999. The report provides recently identified grey literature in the areas of health and science policy, urban health, public health, health services research and health of special populations. Over the past year, the project, with input from colleagues from the International Society of Urban Health, has begun to collect grey literature in these categories more broadly seeking to develop links to enable us to identify such information in developing countries.

Goal and Brief Methodology of the Report
Currently, NYAM librarians scan over 750 organizations to identify material for inclusion in the *Report.* The organizations are both national and international producers of grey literature. NYAM librarians select and acquire print and digital resources, as defined by scope and collection criteria for the project, and catalog the documents into OCLC and the library's catalogue. The Report can be accessed at http://www.nyam.org/library/pages/grey_literature_report. We are the only library in the United States aggregating this health information formally and offering accessibility to researchers through an indexed method. Over 1300 individuals and organizations are regular subscribers, and we collaborate currently with a Kaiser.edu project (Kaiser Family Foundation) to co-brand and promote the information to users. Subscriptions are currently free

Expanding collaboration

The Grey Literature Project (GLP) team was and is seeking to partner with librarians in other countries who would be interested in contributing their grey literature knowledge and information to this global project. It would allow freely available access to information that is hard to find and of importance not only to users in the specific international country, but to the world. In turn, the Grey Literature *Report* indexes grey information from other countries and it is this that we would hope to expand.

In conjunction with the 2010 annual meeting of the International Society on Urban Health, held in New York City and hosted at NYAM on October 27 - 29, 2010, we sought to identify a number of librarians from African countries and invite them to attend the conference as an opportunity to discuss the thinking behind the grey literature project, determine their needs and interest, and, if there was enthusiasm for the project, proceed to organize an effective strategy to develop and activate an a online network as the first step in what we hope may become a global grey literature network.

Since the Project is a library project, we were seeking librarians who understand the processes of information sciences, such as selection, indexing, copyright, standardization, etc. It does not preclude individuals who are working in a library without library science training, but have a strong understanding of library process. It was and is not limited to any particular category of library, but may include hospital and health libraries, government agencies, non-profit agencies, university and academic libraries and/or departmental libraries collecting in the defined areas. It may even interest public libraries where the staff is knowledgeable of where and how information is produced in their country.

We were fortunate to connect to the Carnegie Corporation's Project for African Librarians and obtain travel funding from the corporation to invite librarians from the institutions that Carnegie currently works with to attend the aforementioned conference and at the same time meet with us on our Grey Literature Project.

In July 2010, The New York Academy of Medicine Library approached fifteen African academic libraries that receive support from the Carnegie Corporation seeking in-person discussions about their interest in forming collaborative partnerships to expand the content of the NYAM *Grey Literature Report* project. Although there was much interest, many could not attend due to the short length of time to obtain visas or other commitments. NYAM was very excited however to host six librarians from the following four universities:

- University of Witwatersrand, South Africa
- University of KwaZulu-Natal, South Africa
- University of Dar Es Salaam, Tanzania
- University of Ghana-Legon, Balme Library and the College of Health Sciences

Outcomes

Within the framework of the larger conference program, we set aside specific time to meet and learn about these African libraries and their representatives' interests in grey literature in public health and to share with them our hopes for expansion through collaboration for the *Grey Literature Report.*

After three days of a very intensive agenda of discussions and questions and learning about each other, the group came together with some specific goals and objectives to move the expansion of the *Grey Literature Report* forward in partnership.

The goals for this potential expansion fell into specific areas. The first topic addressed was the aggregation of both retrospective and current materials available in these universities and their countries. The group agreed that a major task is to initiate the aggregation of African grey literature, both retrospective and current, into NYAM's *Grey Literature Report* for dissemination internationally to users seeking grey information in global health. The steps to make this happen were listed as such:

- Training by NYAM Library staff on selection processes and scanning techniques with African libraries' staff
- Identification and selection of grey literature material by African librarians
- Scanning and conversion of selected documents by African Librarians
- Indexing and assignment of MESH headings
- PDF conversion of documents
- Deposit into respective Institutional Repository , if desired
- Links sent to NYAM Library (most likely by e-mail) for inclusion of material in the *Grey Literature Report*
- NYAM catalogs material into bimonthly reports and adds producers to the list of scanned resources
- Credit to all contributing institutions

Recognizing that this is a very ambitious undertaking, it was acknowledged that there were some basic steps to be put into place. All the African libraries were in need of scanners, recommendations on scanning equipment, and funding to purchase the hardware. NYAM librarians needed to produce a manual that will assist the African librarians in the selection and identification process of grey literature for the report using the criteria for selection that is currently in place, if possible. Most critical, prior to next steps, is the administrative approval from each of their universities to participate in the project with NYAM.

The group also identified as a collaborative task the promotion of the concept of grey literature in Africa and identification of agencies in Africa that publish such information. The creation of such identified lists would lead to culling of grey information from their publications and then the expanded access and discovery for international users through the *Grey Literature Report*. Much of this as described would be dependent on the contributions of our guest librarians to identify the producers of grey literature in their respective countries. Specific outcomes from achieving this would result in the following:

- Identification of other sources of African material (e.g. from Health Systems Trust and other NGOs)
- Creation of lists of NGO and other agencies by country
- Identification of other possible collaborating partners, most likely via SANHIP (the existing South African National Health Information Partnership)
- Incorporation of other partners' material as suggested

As noted earlier, all the collaborative work will require formal Memorandum of Understanding to be signed between individual institutions and NYAM; the identified work would require funding with regards to hardware requirements, software development, and personnel support to expand the grey literature project. As of this writing, we have received notice from two of our colleagues, that they have administrative approval to proceed.

In looking to the future, the list below reflects the advantages of this project to African libraries, their home institutions, and NYAM, as written by one of our African colleagues who summarized the meeting. The collaborative project will:

- Make unpublished research in the broad field of public health and health policy available internationally, thus increasing individual institutions' international footprints
- Increased visibility of partners "unpublished" research has potential for moving organizations forward in terms of the vision of institution as a top university
- Populate the individual organization's Institutional Repository
- Preserve valuable material from African collections internationally, as documents will be archived for digital preservation on several international servers
- Offer cost savings (in equipment and human resources) as libraries will not need to back up their own digitized material
- Create an opportunity to begin a digitization project on a small scale, with minimal capital or human resource expenditure
- Afford a mechanism for incorporating previously non-catalogued, ephemeral public health material into the local and international body of literature
- Make available previously unavailable, or hard to find, material to international researchers
- Afford health policy makers access to information to inform national health policy decisions
- Offer the prestige of collaborating together
- Explore possible grant funding for archiving and digitizing of all libraries' materials, should the project grow, based on learned skills acquired through this project.

The group as a whole felt that we had accomplished much in three short days. We knew that there was much work ahead and the challenges of returning to our respective institutions and being engulfed with other priorities loomed ahead, especially with distance and communication issues to deal with as part of the work. NYAM also expressed the desire to proceed slowly as the *Grey Literature Report* was in the midst of a site redesign and database enhancement due for release in early spring 2011. We are all realistic about the length of time such a project may take to implement.

To date, follow up has proceeded with a report to Carnegie Corporation submitted, a Google Group established for librarians to communicate with each other, and shared scanner information distributed. NYAM currently is finishing up the redesign of the *Grey Literature Report.* We encourage all to stay tuned over the next year to follow our progress.

For further information, contact:

Janie Kaplan, The New York Academy of Medicine , Library
jkaplan@nyam.org

Grey Literature Past and Present: The Evolution of Library Annual Reports

Susan E. Searing
University of Illinois, Urbana-Champaign, United States

Introduction

Librarianship was professionalized in the late nineteenth century, when university-level programs of study were established to train librarians in the United States. As the field of library management developed, the annual report became a vehicle for communicating the library's mission and accomplishments to its stakeholders, including funders (both private and governmental) and the general public. Writing and reading annual reports also allowed librarians to share professional knowledge and to benefit from each others' experience. Today, library historians and local history researchers can discover in rich veins of source material in annual report series that may date back a hundred years or more.

Many American libraries still produce reports every year. Today's reports are different in form and content from their ancestors and are just one means among many that librarians now use to communicate with library users, donors, and government officials. Nonetheless, this genre of grey literature is worthy of collection and preservation.

Annual reports as a grey genre

The annual report is a genre rooted in the world of commerce:

> A printed publication, usually less than 100 pages in length, submitted each year by the officers of a publicly held company to its board of directors (or other governing body) and issued in softcover for distribution to current and prospective shareholders, describing the firm's activities during the preceding fiscal year and its current financial position. … Some nonprofit organizations also publish annual reports. (Reitz)

In the United States, state laws usually require governmental agencies such as libraries to submit annual reports. Beginning in the second half of the 19th century, legislation that authorized the establishment of tax-supported public libraries often mandated annual reports as an accountability measure, to assure taxpayers and government officials that their money was being spent well (Cannon 1923). Today, American librarians still compile reports to meet legal requirements imposed by local, state, or national governments.

Early library annual reports served an important additional purpose. In the fledgling days of the profession, before the emergence of widely-read professional journals, American librarians exchanged annual reports in order to learn about innovative facilities and programs. Extensive statistical data was compiled and used to benchmark a library's performance compared to libraries in similar communities and to justify requests for increased funding (Lear 2006). Library annual reports played a similar role in the dissemination of professional knowledge in Great Britain (Peatling and Baggs 2004; Peatling and Baggs 2005).

Despite the importance of annual reports in the development of public librarianship, many librarians dreaded the yearly exercise of writing them. In 1896, Caroline Harwood Garland, a public librarian in Dover, New Hampshire, wrote, "It is usually with a profound sigh that the librarian sits down to write his annual report" (Garland 1896). Librarians were reluctant to take time away from more pressing work to compose a report that, they suspected, few people would actually read (Munn 1923; Rosser 1930; Danton 1949). Modern librarians still question the utility of annual reports, yet they continue to produce them and strive to make them informative and engaging.

Today it is generally believed that improving public relations is the primary purpose of library annual reports (Brown 2003). However, an examination of twenty-five recent English-language books about library public relations, communication, and marketing found that three-quarters of them did not mention annual reports at all.[1] Of the six that did, only one devoted more than a single page to the topic (Hartzell 2003). Have annual reports become such a routine component of library management that lengthy discussions in the professional literature are not needed? Or have such reports diminished in importance? Today's librarians have multiple media at their disposal to share information with the public, and professional journals, email groups, conferences and online communities facilitate networking to share ideas within the profession. Nonetheless, the library annual report persists as a unique sub-genre of grey literature.

Review of the literature

Although public libraries have been producing annual reports for more than a century, and there has been a fair amount of writing *about* them over the years, little substantive research exists. The available literature is of several types. The largest portion consists of short pieces with tips for producing high-quality, attention-getting reports. These sources offer advice on content, organization, design and distribution. However, the last complete, detailed instructions for report preparation appeared more than two decades ago (Bradley and Bradley 1988).

Back in the 19th century, annual reports were so important to the training of future librarians that schools, like the Library School at the University of Illinois, collected and preserved them (Curtis 1912). Library science students in the 1890s were assigned to analyze library annual reports as part of their curriculum (Carman 1895; Warwick 1896). Annual reports were so highly valued as sources of information about library development that the American Library Association (ALA) in 1913 published an index that covered the reports of 170 library commissions, state libraries, university libraries and public libraries in the United States and Canada (Moody 1913).

Both British and American library historians have written convincingly about the value of older annual reports as historical evidence (Peatling and Baggs 2004; Peatling and Baggs 2005; Lear 2006). However, these remarkably rich sources of information remain under-utilized by historians, perhaps because they have been difficult to obtain and tedious to examine. As more and more older documents are digitized, this will hopefully change.

Until the 1980s, professional magazines often carried notices and reviews of exemplary annual reports, especially ones that were creatively designed. For example, *Library Journal*, a long-running magazine for American librarians, drew attention in the late 20th century to reports that mimicked a recipe card, a federal income tax form, and a restaurant menu ("An annual report worth cribbing" 1985; "A taxing recipe" 1989).

A small number of studies have gone beyond mere description and compared reports, usually in order to derive a set of best practices for content and format. In 1934, Ridley and Miles rated library annual reports from nineteen American cities. Their criteria included promptness, visual appeal, readability, length, cost, and breadth of distribution, as well as elements of content, such as diagrams, pictures, the balance of topics, and evidence of "social significance," i.e. impact upon the community (Ridley and Miles 1934). The most comprehensive comparative study was conducted in the mid-twentieth century by Madeline S. Riffey, a librarian at the University of Illinois. She examined fifty current annual reports from American public libraries, choosing a sample that "represent[ed] attractive and readable reports from libraries of various sizes and regions of the country" (Riffey 1952, p.1). Riffey conducted a detailed analysis of the reports, assessing them on numerous dimensions of organization, content, format, and distribution. Riffey's carefully chosen sample forms the basis for my own study, which looks at digital representations of 21st-century reports from the same set of libraries.

Identifying a sample

There are over nine thousand public libraries in the United States (American Library Association). To identify a representative sample, I relied on Riffey's earlier work. I visited the websites of the fifty libraries in her 1952 study and searched for their annual reports. (All of the libraries are still in existence, although some have changed names or merged.) I was able to discover digital versions of current annual reports for only eighteen (36%). "Current" in most cases is 2009 or 2008. Typically the annual report is linked directly from the "About Us" section of the library website. In two cases, there were no visible links to the report, but I was able to find it by using the site's search engine. All but one of the reports are PDF documents; a few are also available as interactive web files.

What can annual reports tell us?

Statistical information

Today's annual reports present statistical data very selectively. The report writers choose numbers that will impress the average reader, such as the total number of books borrowed during the year or the number of people who accessed the internet on library computers. Historically, librarians counted many more things, and statistics were a constant topic of discussion among early librarians. In 1902, the journal *Public Libraries* published a model form for annual statistical reporting. It was four pages long. Among other data, it required librarians to track the number of books added to their collections in 23 categories, including broad subject classes such as "religion" and "useful arts," document types such as government publications, and books in the languages of major immigrant groups, such as German, Norwegian, and Bohemian ("Official reports for libraries" 1902).

The following year, a paper presented at the 1903 conference of the American Library Association outlined a staggering total of 131 questions to be answered in an annual report, most of them statistical in nature (Foster 1903). Repeated proposals were made in both Great Britain and the United States to standardize library statistics. The American Library Association issued a succession of guidelines, and for years many library annual reports stated explicitly that their statistics followed the format recommended by ALA. However, the ALA guidelines were not adopted universally.

Today, the federal Institute for Museum and Library Services (IMLS) collects statistics from over 9,000 public libraries. The data covers facilities, staff, revenue, expenditures, collections, and services, and the IMLS website displays a handy tool for generating comparisons between libraries (IMLS library statistics). The Public Library Association, a division of the ALA, also collects statistics from a representative sample of 800 libraries and publishes an annual compilation which complements the IMLS data. (Public Library Association 1991-). Consisting solely of numerical data, these national-level reports lack the rich narratives of local annual reports but do facilitate comparisons and benchmarking.

One component of older annual reports that is rarely seen nowadays is a statistical breakdown of the items library users checked out. The most concise reports simply presented the numbers or percentages of fiction books versus non-fiction books that were borrowed during the year. But in other reports, the totals, especially for non-fiction, were further divided by subject, such as history or science, or genre, such as biography. Thus library annual reports depict the changing reading habits of American citizens over time and may reveal local and regional differences in literary taste.

With the software available today, it is a simple matter to record statistics and generate eye-catching, informative graphics. But even before the computer era, some librarians tried to lend their statistical reports some visual impact. The reports of the Salt Lake County, Utah, Public Library, for instance, although inexpensively mimeographed on low-quality paper in the 1940s and 1950s, made frequent use of illustrations and photographs to liven up the quantitative data. A more recent approach to making the numbers interesting involves calculating the value that public library services would have on the open market – the cost of purchasing a book, attending a cultural program, renting meeting space, and so on – to demonstrate a high return on a small investment of tax dollars.

Narrative information

Some library annual reports provide only the most concise, factual narratives, while others are long and opinionated. From a historian's viewpoint, the verbose reports are the best. Bernadette Lear writes, "It is exhilarating to read about the rise of open shelves; business, technical, and other special collections; young adult services; bookmobiles; and other advances…One can often glean important dates, staffing changes, milestones, city ordinances, a sense of how the library operated…" (Lear 2006, p.467) . Whereas statistics recorded how many books were borrowed, the textual portion of a report might bemoan the fact that borrowers prefer light fiction to more serious books, or might speculate on why books on certain subjects were suddenly popular. Library annual reports also reflect broader trends in American history. For example, in our present economic downturn, it lifts one's spirit to read how public libraries coped with budget cuts and survived during the Great Depression of the 1930s.

Testimonials from members of the community rarely appeared in older reports, but they are a very common feature today. Photographs of happy library users are accompanied by their own words recounting the positive impact the library has had on them and their families. Today's reports often include long lists of individual and corporate donors. Such lists were not unknown in earlier years, but they were not so common. Because public libraries, although still supported by taxes, increasingly rely on donations and grants, the annual report has evolved into an important tool for fundraising. For this reason, perhaps, the names, quotations, and images of library users are prominent in today's library annual reports.

Staff activities, on the other hand, have largely disappeared from annual reports. Members of the public are simply not interested in the staff's professional development activities or individual accomplishments. Older reports often identified every staff member by name, even janitors and part-time employees, but today's reports typically list only directors and top-level administrators. Volunteers are more likely to be recognized by name than are front-line staff members, especially in reports from large multi-branch systems like the New York Public Library or the Denver Public Library. Members of boards of trustees are *always* listed.

In general, library annual reports have always focused more on the past year than on the future, but goals and plans are sometimes discussed. Today, some reports enumerate strategic goals, and many reprint the library's mission statement. Despite widespread budget cuts and staff reductions, public library reports for 2008 and 2009 are relentlessly upbeat in emphasizing the library's success. In some older reports, by contrast, the librarians honestly acknowledge disappointments and areas needing improvement. For example, Milton Ferguson, director of the Brooklyn Public Library in the 1940s, wrote eloquently year after year about the desperate need for salary increases and shorter work hours for librarians.

Visual information

Today's annual reports are treasure houses of visual information. Nearly all of them feature full-color photographs of people using the library – choosing books from the shelf, peering at computer screens, listening raptly to storytellers. Racial, ethnic, gender and age diversity are evident in the photographs, signaling that the library is a resource for everyone in the community.

Library buildings are also depicted, especially new or renovated ones. Whereas casual photos of users were less common in the past, photos and drawings of buildings have been a staple of library annual reports for more than a century, making them a source of information for architectural historians. Occasionally, library reports even contain original floor plans.

Snapshots of interiors reveal the library furnishings of earlier eras. Some photos show new service locations in a pristine state, but others depict the active life of the library as readers interact with staff and use the collections. When a library acquired new equipment, such as a microfilm reader or audiotape player, a photograph of it was often inserted in the next annual report. Less frequently, photographs depict services offered outside the walls of the library, such as book delivery to hospitals or storytelling in outdoor parks.

Library annual reports also contain images of the people who work in and use public libraries. Photographs of staff appear with some regularity in older reports, often working behind the scenes in cataloging or binding departments. In 1947, the Buffalo (New York) Public Library used its annual report both to celebrate its fiftieth anniversary and to make the case for better facilities. "What is the Buffalo Public Library?" the report asked.

> The Buffalo Public Library is not the drab and crowded buildings with their dark, congested book stacks. Nor is it the books alone, for books unread are no books at all. No, the Buffalo Public Library is the people. It's the people who use the Library, you and I and our children, over 2,000 of us a day, every day. It's also the people who make the Library tick. It is the men and women—professional librarians, clerks, bookbinders, and maintenance men—who make up the small but smoothly functioning team we call the Library Staff (Galt and Foster 1948, p.8-9).

Pictures of the staff at work in cramped spaces were eloquent visual arguments for a new building. Today, by contrast, when staff members are pictured, they are nearly always shown interacting with users in public spaces. Usually, however, library users are depicted utilizing the library without any assistance.

When it comes to graphic design, library annual reports continue to mine ideas from advertising and other media. For example, the latest report from the Salt Lake County, Utah, Library is formatted like a fotonovela, while other libraries use color, multiple fonts, and creative layouts to generate excitement on the page.

Future research

The present research is just the beginning of understanding library annual reports as a genre of grey literature that carries information not readily found in books or in librarians' professional journals. Restricted to a population of libraries selected in 1952, the present study is necessarily limited in scope. Riffey's original set of criteria for comparing reports needs to be updated to note the presence of content such as URLs, mission statements, and return-on-investment analyses. A rubric for evaluating visual information should also be developed and applied.

From Riffey's original group of fifty, only 18 libraries have mounted current annual reports on their websites. Therefore, a future direction for this research project will involve contacting librarians at those libraries from the original sample who do not make their reports available on the web, in order to learn whether they still produce annual reports and, if so, how their reports are distributed.

Further, given the clear indication from contents and design that today's library annual report is intended as a marketing and fundraising tool, it could be enlightening to interview librarians who prepare annual reports, in order to discover what audiences they have in mind and what outcomes they perceive from their efforts.

Academic libraries in the United States, in particular those at large universities, also issue annual reports. The available research on academic library annual reports is very limited (Swanepoel and Smit 2003; Sauer 2006). A future phase of my research will identify a sample of annual reports from academic libraries for evaluation and comparison.

Implications for collecting library annual reports

I do not doubt that there are many more questions that one could ask about this body of grey literature, and the answers would reveal overlooked aspects of library management, library outreach, the library's idealized self-image, and the role and position of the library vis-à-vis its community of users and supporters. And a larger question remains as well: how and where will such reports be preserved for future scholars and librarians to study and enjoy? The University of Illinois Library ceased systematically collecting such reports in the 1970s. Some continue to arrive as gifts, but circulation statistics and informal observation suggest strongly that current library annual reports are not used by today's students and researchers. In the past, a student preparing for a job interview at a public library might seek out and read its annual report; today that student will explore the library's website instead. Thus, under increasing pressure to reduce the labor-intensive processing of gift materials, reports that have a web presence are no longer cataloged and retained in print at my university. Can we count on the originating libraries to retain archival copies? Will historic annual reports be digitized as part of mass digitization efforts? Alas, although the University of Illinois is a partner in the Google Books project and the Open Content Alliance, many of the historic annual reports are in fragile condition or too tightly bound to be scanned.

The University of Wales Aberystwyth has preserved a large collection of British public library annual reports from the period between 1850 and 1919, and created a bibliographic database of the collection (Peatling and Baggs 2000). No such project has been undertaken in the U.S. Nor is there, to my knowledge, any systematic effort underway to capture and preserve the current output of annual reports. Neither DLIST, the US-based digital repository for library and information science, nor E-LIS, its more successful European counterpart, have made an effort to incorporate annual reports, though I believe they would not be unwelcome.

The question of preservation becomes more complicated as libraries adopt new media for their reporting. Consider the 2009 report of the Topeka and Shawnee County (Kansas) Public Library, which is in the form of a multimedia website. Incorporating audio and video content and page-turning technology, as well as multiple static web pages, it is a lively portrait of a library energetically engaged with its community. What access will future historians have to this document? Because this same question is being asked about digital grey literature in other areas, I am hopeful that a solution will be found that will enable today's library annual reports to be as useful to future historians as the old ones are to us today.

Notes

[1] I examined twenty-five English-language books in the University of Illinois collection, published since 2000 and bearing the subject headings "Libraries—Public relations," "Libraries—Marketing," or "Communication in library science."

References

An annual report worth cribbing. 1985. *Library Journal* 110 (32): 32.

American Library Association. ALA library fact sheet 1 -- number of libraries in the United States. Available from http://www.ala.org/ala/professionalresources/libfactsheets/alalibraryfactsheet01.cfm (accessed 11/11/2010).

Bradley, Jana, and Larry Bradley. 1988. *Improving written communication in libraries*. Chicago: American Library Association.

Brown, David. 2003. Annual reports: A powerful public relations tool. *Mississippi Libraries* 67 (2): 46-7.

Cannon, Lucius H. 1923. Public library annual reports and the law--I. *Library Journal* 48: 747-51.

Carman, Maude Wheeler (Straight). 1895. "Library Reports." [Chicago].

Curtis, Florence Rising. 1912. *List of library reports and bulletins in the collection of the University of Illinois Library School.* University of Illinois bulletin vol. 9, no. 12. Urbana: University of Illinois.

Danton, Emily Miller. 1949. Just who reads our annual reports? *Library Journal* 74 (11): 880-8.

Foster, William E. 1903. The essentials of a library annual report. *Library Journal* 28 (7): 76-81.

Galt, Alexander, and John T. Foster. 1948. *This is your library: Fifty years of service to Buffalo.* Buffalo, N.Y.

Garland, Caroline Harwood. 1896. The librarian's annual report. *Library Journal* 21 (12): 65-6.

Hartzell, Gary N. 2003. *Building influence for the school librarian: Tenets, targets & tactics*. 2nd ed. Worthington, Ohio: Linworth Pub.

IMLS library statistics - overview. Available from http://harvester.census.gov/imls/index.asp (accessed 10/28/2010).

Lear, Bernadette A. 2006. "'Tis better to be brief than tedious": The evolution of the American public library annual report, 1876-2004. *Libraries & the Cultural Record* 41 (4): 462-86.

Moody, Katharine Twining. 1913. *Index to library reports.* Chicago: American Library Association Publishing Board.

Munn, Ralph. 1923. Library reports. *Library Journal* 48: 413-4.

Official reports for libraries. 1902. *Public Libraries* 7: 466-9.

Peatling, G. K., and Chris Baggs. 2000. Creating a database of British public library annual reports, 1850-1919. *Journal of the Association for History and Computing* 3 (3). Available from http://hdl.handle.net/2027/spo.3310410.0003.302 (accessed 11/11/2010).

Peatling, G. K., and Chris Baggs. 2004. Early British public library annual reports: Then and now: Part I. *Library History* 20 (3): 223-38.

Peatling, G. K., and Chris Baggs. 2005. Early British public library annual reports: Then and now: Part II. *Library History* 21 (1): 29-45.

Public Library Association. Public Library Data Service Advisory Committee, and American Library Association. 1991- . Public library data service statistical report, 1988- .

Reitz, Joan M. ODLIS: Online dictionary for library and information science. Available from http://lu.com/odlis/ (accessed 11/11/2010).

Ridley, Clarence E., and Arnold Miles. 1934. Evaluating library reports. *Bulletin of the American Library Association* 28 (7): 394-6.

Riffey, Madeline S. 1952. *Annual reports for public libraries.* Urbana, Illinois: University of Illinois Library School. (Occasional papers, 28).

Rosser, E. M. 1930. A readable library report. *Bulletin of the American Library Association* 24 (1): 8-9.

Sauer, James. 2006. The annual report. *College & Undergraduate Libraries* 13 (2): 21-34.

Swanepoel, A. J., and I. P. J. Smit. 2003. Towards an understanding of annual reports of university and technikon libraries in South Africa. *College and Research Libraries* 64 (2): 141-50.

A taxing recipe: How to get their attention with the annual report. 1989. *Library Journal* 114 (18): 30-1.

Elma Warwick. 1896. "Librarian's Report." [Chicago].

I search

I find

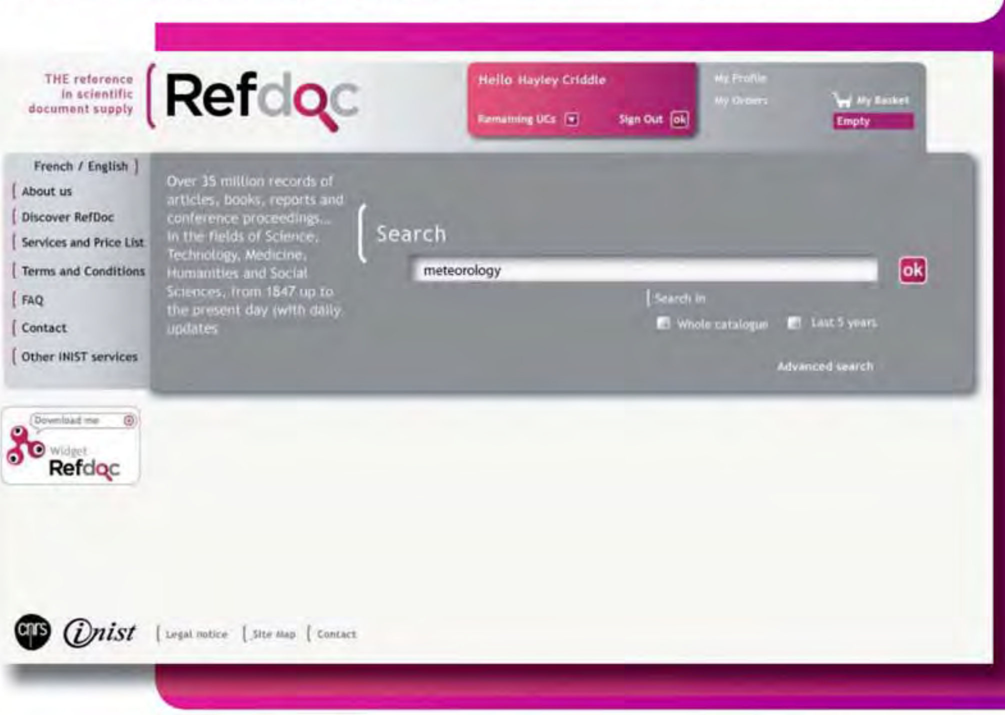
I order

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



Over 35 million records of articles, books, reports and conference proceedings... In the fields of Science, Technology, Medicine, Humanities and Social Sciences, from 1847 up to the present day (with daily updates

Search: meteorology

INIST DIFFUSION

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

bbcom 5677

From OpenSIGLE to OpenGrey
Changes and Continuity

Christiane Stock and Nathalie Henrot
Institute for Scientific and Technical Information, INIST-CNRS
France

First presented at the GL8[i] conference in New Orleans 2006 as a prototype, OpenSIGLE went life in December 2007. After three years of existence, the results are beyond all expectations. OpenSIGLE has become a reference source for grey literature, and its user community has grown constantly, especially from outside Europe. The integration of the GL conference preprints into the repository from 2008 onwards not only added research papers on the topic of grey literature to its contents, but also permitted OpenSIGLE to be accepted in the "Directory of Open Access Repositories" (OpenDOAR*).
In spite of the success of OpenSIGLE it's not wise to rest on one's laurels. The change of name to "OpenGrey" signifies a shift in the content of the repository as well as in its physical appearance. Besides providing a new look and a more convenient technological environment, OpenGrey closes the gap between the close of the SIGLE* database and today, including recent records and links to the full text.
The paper presents the new website which includes numerous facilities requested by users such as OAI-PMH*, the possibility to export records and an improved access to the document itself. OpenGrey also takes into account a changed user behaviour, where visitors arrive after searching Google or Google Scholar* and want all relevant information at a glance. The paper further explains input procedures and gives other information for the ongoing updates of the repository. Finally we call former SIGLE members and new partners to contribute to OpenGrey.

From SIGLE to OpenSIGLE

SIGLE (System for Information on Grey Literature in Europe) was a unique European database of bibliographic records in grey literature. It was produced between 1980 and 2005 by initially seven and in the end fifteen members of the European Union, represented by major libraries and research organizations. Its contents covered all scientific disciplines (pure and applied science and technology, economics, social sciences and humanities).
As a commercial product SIGLE was accessible through subscription to hosts, e.g. STN International, and available on a CD-ROM produced by Silverplatter/Ovid.

INIST decided to transfer the results of 25 years of work[ii] onto an open access platform. As a result OpenSIGLE went live in December 2007 with almost 700 000 bibliographic records, using DSpace* technology and its qualified Dublin Core metadata* format. The operation was to be low cost and easily feasible.

It was very important that key information found in every SIGLE record was preserved during the migration:

- English title or keywords
- SIGLE classification code
- Availability statement, in order to facilitate the order of a paper copy

One of the main goals of SIGLE was to facilitate access to the paper document. Most of today's requests for assistance from users concern the document availability.

OpenSIGLE – its evolution and its usage

Though not perfect[iii], with its lack of flexibility for the user, and no possibility of exporting either search results or records, OpenSIGLE found an ever growing audience and new visitors from – nearly - all around the world.

OpenSIGLE was included in OpenDOAR (Directory of Open Access Repositories) in November 2009, after its integration in the WorldWideScience.org* portal a year earlier. Indexed in Google and Google Scholar since summer 2008, the database is reached through the Google search engine by an increasing number of users, and visits via Google Scholar amount to 30% per month. For 2010 the average number of monthly visits exceeds 35 000. The audience is definitely worldwide, not only European: 10 % of the visitors come from Asia, 20% from North America, continents not involved in the database production. These proportions have remained fairly stable over the past year.

The following statistics (see below) are based on php/MyVisites*, an open source software using a tracker, similar to Google analytics. As a result, figures are lower than with a log analyser[iv].
Two years of statistical data collection reveal a considerable evolution in numbers and the influence of the Google generation. Between 2009 and 2010 the number of visits almost tripled, and the number of page views more than doubled. Simultaneously the average duration of a visit decreased as well as the number of pages viewed in a visit.

We take the data as a sure indication of a change of behaviour in our visitors. Many of them reach the OpenSIGLE after using a search engine, spend a short time looking for relevant information and leave. However some of our users tried to replicate former search strategies in OpenSIGLE and were successful.

Average per month	2009	2010
Visits	12 000	35 000
Pages viewed	35 450	88 580
Length of visit	107 sec	91 sec
Number of pages viewed per visit	3.4	2.8

The most important highlight for 2010 relates to the Greynet collection[v]. As of October 2010 the complete collection of conference preprints from the GL conferences are accessible in full text and in open access. GL5 to GL11 were provided by GreyNet in electronic format and the first conference became available in full text on OpenSIGLE as early as May 2008. GL1 to GL4 were received in paper form, digitized by INIST and uploaded to OpenSIGLE in September 2009.

When OpenSIGLE became OAI-PMH compliant, we decided to be more explicit about the rights and adopted the Creative Commons licence*, in order to inform the users and partners about the different conditions of use of the bibliographic data and GL proceedings.

Time for change

Several reasons lead us to think of changing the system.
Right from the beginning of OpenSIGLE we were facing limits in the technical performance of DSpace, especially when uploading and indexing GL collections. The problems were even noticeable for smaller updates. DSpace obviously reached its limits with 700 000 records in the database. Besides, the website layout was rather simple (low key) and today no longer answers current needs for referencing.
The most important reason for thinking of a change was the absence of features requested over the years by OpenSIGLE users, such as the possibility to export search results for state of the art studies.

An important number of emails from users request how to obtain the document, a service which was one of the principal goals of SIGLE and OpenSIGLE. This essential information should be seen at first glance.
In fact, the availability statement is « hidden » in the full record display of OpenSIGLE. However, few users take the time to look it up. 55 percent of the visits only display one page in OpenSIGLE before leaving again, according to our statistics. In addition, in the context of the Google generation we presume that the brief record display becomes more and more the "entry page" to the database, due to the Google Scholar search, and half of the time the unique page-view. The new website should therefore improve the access to information on document supply through the general layout, improved record display and additional links.

Early on, INIST-CNRS intended to re-open OpenSIGLE for "new" input. This can be done in two ways:
- add records from 2005 onwards from former EAGLE* members (INIST-CNRS included).
- open the database to new European partners.

However, the technical limits mentioned above would have been a major hindrance, hence the need for a change of software.

In order to reinforce the changes in the technical environment as well as in the contents and policy for the database, INIST-CNRS decided on a change of name: "OpenGrey". Several new domain names have been acquired for OpenGrey with the extensions .eu, .fr, .net and .org. The new domain name will be *opengrey.eu*.

The OpenGrey homepage

The homepage of the new website OpenGrey aims to meet current needs of the users as well as to facilitate referencing. It is divided into three parts with different groups of information. The upper field holds the name and the logo, - inspired by the lemniscates or infinity symbol -, and provides information on the contents for referencing. Tabs allow the user to access further information, e.g. on the partners and former EAGLE members as well as to choose the language for the user interface.
Besides the "Google" like search field, the centre includes three blocs for short texts: a mini "about", a "focus" on specific subjects and a "news" bloc.
The bottom part, separated into four blocs, provides more information on search and help, a choice of export tools (highly requested by users), legal mentions, and further information (about, tools for partners etc.).

Change and continuity:

The major change from OpenSIGLE to OpenGrey relates to the software: The DSpace platform used for OpenSIGLE is replaced by Exalead® * as the search engine used for the database, completed with in-house developments using php and MySql software for the user interface.

Persistent identifiers are essential to guarantee perennial access to records or documents. In OpenSIGLE each record as well as the communities and collections of the DSpace architecture are identified by a unique identifier, the handle*. The handle system allows to using the URLs as persistent identifiers which remain, even in case of server changes. Many websites linking to OpenSIGLE or a particular collection (e.g. GL11) use the handle. In order to assure continuity the handles of the present OpenSIGLE records will migrate to the new system. A redirection is planned for the handles identifying communities and collections of OpenSIGLE which cease to exist in the same way in OpenGrey.

Another change comes with the Exalead software. The user interface for search provides the Google-like field for full text record search. Exalead also offers the possibility to refine the results through faceted search. The criteria chosen for refinements are similar to the Czech repository NUSL/NRGL[vi] (author, subject, date, language…). This feature has a major consequence for the backoffice: metadata which were merged for the simpler DSpace format in OpenSIGLE need to be detailed once more, and even new controlled fields must be added to allow the refinements and to insure consistency.

As mentioned before, we know from usage statistics that the brief record display is the entry page for many visitors and the one and only page viewed by half of them. Therefore the development of features based on this record display has become a necessity. All relevant information must be available at first glance or invite the user to click further. For example, we intend to make an explicit distinction between:
  Access to the paper copy
  Access to the online full text
In addition we must provide easy links to both fields and to further information, e.g. on how to obtain a copy of the paper document or on the terms and conditions of the partner organization.

The partner homepage

OpenGrey will provide each partner with a space for information on its institution, a kind of "homepage", accessible from a tab "Partners" on the OpenGrey homepage. OpenSIGLE already offers information on former EAGLE members in the collection page of the "Country Community"[vii]. This content will be transferred to the new partner homepage as a starting point, namely:
- General information of the organization, with a logo, if available.
- Information on how to obtain a copy from old SIGLE records.

Each partner is strongly encouraged to propose any improvement or information about its organization, its policy of document delivery and the fees attached etc… Any information on national grey literature initiatives, databases, recent updates, or any information on conferences relating to the subject could be announced in the "News" section of the Homepage.

Contents, input and updates of OpenGrey:

OpenGrey will include all records from the present OpenSIGLE database. In addition, we intend to add new records (created since 2005) from partner organizations, including French grey literature as mentioned before. OpenGrey will not host the documents themselves. The new cooperation will be based on formal agreements with each partner organization. The partner retains the control over the grey documents. An open access policy is encouraged for the full text documents, items which remain under the control of the partner. The bibliographic records as well as the OpenGrey website will be placed under the Creative Commons Licence.

OpenGrey metadata should be as rich as possible. This is why we prefer to receive files in xml format via ftp, although metadata harvesting with OAI-PMH is not excluded. However, harvested material will most likely be in basic Dublin Core format, which is poor material for Exalead®'s refinement feature.
We defined a metadata scheme based on the DSpace scheme (qualified Dublin Core), but taking into account new needs (e.g. for Exalead®'s refinements) and ideas from similar repositories. Suggestions for additions are very welcome.

Updates to the OpenGrey database are done by batch upload of files, excluding direct deposit through the interface, contrary to institutional repositories[viii].
Tools for participants will be made available through the website. It might be interesting to open a community space dedicated to partners. On a first meeting it was suggested to create an advisory board to discuss arising questions.

Outlook:

Knowing than many documents referenced in the former SIGLE database have since been digitized or became available in electronic format, leads to another point for improvement. Our plans for future developments include the addition – if possible through batch upload – of links to the now available full text in the partner repository.
Another project planned with Greynet is to add persistent links to datasets in the bibliographic records, when they are available in a repository. The initiative might start with the next GL conference.

Glossary:

Creative Commons: The Creative Commons copyright licenses and tools forge a balance inside the traditional "all rights reserved" setting that copyright law creates: http://creativecommons.org/licenses/

DSpace: DSpace is a software for academic, non-profit, and commercial organizations building open digital repositories: http://www.dspace.org/

Dublin Core Metadata: The Dublin Core Metadata Initiative (DCMI) is an open organization, incorporated in Singapore as a public, not-for-profit Company engaged in the development of interoperable metadata standards that support a broad range of purposes and business models: http://dublincore.org/specifications/

EAGLE: European Association for Grey Literature Exploitation**,** producer of the SIGLE database: http://en.wikipedia.org/wiki/European_Association_for_Grey_Literature_Exploitation

Exalead® is a global software provider in the enterprise and Web search markets: http://www.exalead.com/software/

Google Scholar: Google Scholar provides a simple way to broadly search for scholarly literature ...: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites: http://scholar.google.com/

Handle: The Handle System provides resolution services for unique and persistent identifiers of digital objects, and is a component of CNRI's Digital object (Corporation for National Research Initiatives® is a not-for-profit organization formed to undertake, foster, and promote research in the public interest) : http://www.handle.net/

OAI-PMH: The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH in the remainder of this document) provides an application-independent interoperability framework based on metadata harvesting*:* http://www.openarchives.org/OAI/openarchivesprotocol.htm

OpenDOAR: Directory of Open Access Repositories : The OpenDOAR service provides a quality-assured listing of open access repositories around the world: http://www.opendoar.org/find.php

OpenSIGLE: http://opensigle.inist.fr

SIGLE: System for Information on Grey Literature in Europe: SIGLE was an online, pan-European electronic bibliographic database and document delivery system for grey literature. http://en.wikipedia.org/wiki/SIGLE

php/MyVisites: is a free and open source (GNU/GPL) software for websites statistics and audience measurements: http://www.phpmyvisites.us/

WorldWideScience.org: a global science gateway comprised of national and international scientific databases and portals: http://worldwidescience.org

References

[i] http://opensigle.inist.fr/handle/10068/697774
[ii] the main effort relating to the identification and collection of documents
[iii] mainly because it used standard DSpace features without further development
[iv] older software based on simple log analysis would yield numbers at least four times higher
[v] http://opensigle.inist.fr/handle/10068/697753
[vi] http://nrgl.techlib.cz/index.php/Main_Page
[vii] E.g. NTK : http://opensigle.inist.fr/handle/10068/20
[viii] However, a special development might allow partner administrators from partners to directly edit individual records.

Slovak Centre of Scientific and Technical Information

Our mission is to support the development of science, technology and education in the Slovak Republic.

- ❑ Operating and improvement of complex information systems for research and development – Central Information Portal for Research, Development and Innovation. (more info: www.vedatechnika.sk)
- ❑ Popularization of science and technology in the Slovak Republic – The National Centre for Popularization of Science and Technology in Society. (more info: www.vedatechnika.sk)
- ❑ Implementation of the national projects, which are aimed at improving of information security of research and development and services for research and development. (more info: www.cvtisr.sk, http://nispez.cvitsr.sk, http://dc.cvtisr.sk)
- ❑ Library and information activities – national information center for support of science and technology and specialized scientific public library of the Slovak Republic. (more info: www.cvtisr.sk)
- ❑ Methodical and analytical activities, which are aimed at supporting the management and evaluation of research, development and higher education for the Ministry of Education, Science, Research and Sport of the Slovak Republic.

Contact address:

Slovak Centre of Scientific and Technical Information
Lamačská cesta 8/A, 811 04 Bratislava, Slovak Republic
Tel.: +421 2 69253 102
Fax: +421 2 69253 180
E-mail: cvti@cvtisr.sk
www.cvtisr.sk

Integration of an Automatic Indexing System within the Document Flow of a Grey Literature Repository

Jindřich Mynarz and Ctibor Škuta,
National Technical Library, Czech Republic

Abstract

The Web empowered the authors of grey literature to publish their work on their own. In case of self-published works their author is also their indexer. And because not many of the grey literature authors are professional indexers, this may result in poor or no indexing.

Even though the Web made publishing easier, indexing is still hard. Nevertheless, we believe that the web technologies and machine learning algorithms may help to reduce the cognitive overhead involved in indexing, and make it eventually as easy as publishing on the Web is.

To help overcome the issue of quality and consistency of subject indexing automatic indexing systems can be used. Given enough full-texts already equipped with the terms from the controlled vocabulary that is to be used, machine learning algorithms can be employed.

Our aim is to provide *human-competitive* automatic indexing to authors and producers of grey literature. We demonstrate how an automatic indexing system based on machine learning can be integrated into the document flow in an open source digital repository of grey literature. We build upon open source tools and a controlled subject headings vocabulary available in an open standard format.

We will be using *Maui Indexer* as an automatic indexing system, *CDS Invenio* as a digital repository software, and *Polythematic Structured Subject Heading System* (PSH) as a knowledge organisation system. Both Maui Indexer and CDS Invenio are open source, and CDS Invenio's modular architecture makes it possible to extend it with new functionality. Maui Indexer works with controlled vocabularies expressed in Simple Knowledge Organisation System format in which the PSH is available.

From these components combined we will try to put together a solution for automatic indexing aimed at grey literature in the Czech language environment. Maui Indexer is domain and language independent so it is possible to adapt it for the field of Czech grey literature. The document samples we will test on will come from the *National Repository of Grey Literature* which is maintained by the *National Technical Library* of Czech Republic.

In the end, we will discuss integration of the automatic indexing component from the user perspective and sketch out how the user can interact with it through the user interface. We will also provide details around the actual implementation of the proposed system. The conclusion will deal with the evaluation of benefits of the implemented system for grey literature authors.

1 Introduction

The Web offers a publishing model that empowers masses of users to publish their works on their own. As it was stated in the previous literature, grey literature *"does not imply any qualification, is merely a characterisation of the distribution mode"*[3], and the Web can be considered as the single most significant distribution mechanism for grey literature. The sheer volume of documents available on-line constitutes a significant part of the grey literature publishing landscape. The Web has made *self-publishing* easier by lowering both the financial hurdles and the amount of know-how necessary for the publication process and thus it enabled a kind of *"do it yourself"* publishing.

While this is a tremendous benefit without which the *open access* movement would not have been established as firmly as it was, there are also drawbacks to it. By contrast to this mode of making documents accessible, the traditional publication models have procedures in place that go along with publishing that are not replicated well in grey literature publishing on the Web. The part of traditional publishing process that can be neglected in publishing on the Web is *subject indexing*. While it is now for the most part clear how to publish documents on the Web, the approaches to subject indexing are less established and available for use to non-professional users. This topic will be discussed in our paper and our focus will be to show how *self-indexing* of documents published on the Web via a digital repository can be accomplished; much in the same way users were endowed with the ability to *self-publish* their works.

2 Indexing of Grey Literature

Grey literature is characterized by a way of publishing that outputs documents with limited visibility. It may be hard to find such documents because they are distributed in a way that does not use established document access mechanisms, such as commercial databases and the like. This aspect of grey literature makes it difficult to be searched for either through libraries or web search engines. Also, the field of grey literature is closed tied to the *open access* way of publishing. However, if a document *cannot be found* there is no use of it being released in the *open access* way.

As we will argue, additional *subject indexing* terms can make grey literature documents searchable in a meaningful way so that they eventually become more prominent in the search results of library or web search engines. Subject indexing can be seen as an essential requirement to make documents findable [8], even though there are powerful search engines that enable documents to be found even without carefully crafted indexing. It can help in making grey literature documents more visible. Moreover, there is a consensus that it is necessary for useful navigation interfaces that can be built on top of digital document repositories (e.g., a faceted navigation) [7].

Subject indexing metadata enriches documents with *affordances* that allow to do more with them. It can support navigation interfaces that make it possible to browse the document collection in a useful manner based on *navigation paths* taken from the structure of the knowledge organisation system used for indexing. In this way the connections within the subject indexing system, such as hierarchical or associative relationships, may be harnessed as a *"map"* to the document collection. The subject indexing places the documents in a logical space constructed by the knowledge organisation system's structure and allows the user to browse the documents organized in such way by following the relationships between indexing terms. Every subject indexing term that is assigned to a document constitutes an *entry point* through which the document can be found and accessed.

Thus, we see subject indexing as an important enrichment to the document that enables to build interfaces for document repositories that can be navigated in a meaningful way. In this paper, we are interested in subject indexing done by the grey literature authors, and we have to admit that there are barriers that can make subject indexing a difficult task for them.

Grey literature is often published directly by the documents' authors, which may imply that, if it contains any subject indexing terms at all, it is the indexing done by the *authors themselves*. Subject indexing may be difficult for non-professionals and therefore this situation may result in no or poor indexing.

The established best practise for subject indexing is to use a *controlled vocabulary*. Authors know best about the contents of their works but they might not be familiar with the controlled vocabulary that they are supposed to use to express their works' content. They may not know how best to use the subject indexing system to describe their documents and therefore it constitutes a barrier for them, because first they have to learn how to use it.

This is why professional indexers are able to produce better indexing - they know how to *use* the subject indexing terms, especially with respect to the whole document collection. This background knowledge of the indexing system and the document corpus is a fundamental requirement for high-quality indexing.

In automatic indexing this background knowledge is in a sense captured in the *indexing model* on which the automatic system operates. And thus it turns out that, if configured properly, automatic indexing might come up with subject terms of reasonable quality and consistency. In the following sections we will describe how to reach that goal, while we will deal in detail with an application of automatic indexing to grey literature.

3 Automatic Indexing

In this paper we will investigate the option of non-professional *self-indexing* of grey literature on the example of an automatic indexing system for a digital document repository. We propose a semi-automatic indexing system that incorporates human feedback for the final selection of indexing terms. The system suggests a set of pre-selected indexing terms from a controlled vocabulary that may be used to describe document's content. These terms can be refined in the next step by the user interacting with a selection via interface that enables to remove or add new indexing terms.

The main help of our proposed approach is to lessen the cognitive overhead involved in intellectual indexing. Rather than completely automating the process of subject indexing we decided to use the automation for suggestions of indexing terms that can be amended and validated by the human user.

In this way we strive to provide grey literature authors with a tool that makes it possible for them to come up with non-professional indexing that is of high quality and consistency. The reason for such a goal is that the inconsistency of *user-generated indexing*, often done with freely created keywords, is one of its main drawbacks. This is what we try to alleviate by preprocessing the document and suggesting indexing terms in an automated manner. Also, we argue that this approach might help to increase the *scalability* of subject indexing.

The intellectual indexing carried on by professional indexers does not *scale*. Given that the size of the grey literature published on the Web is ever-increasing, there is a need for indexing system that is able to scale. The traditional solutions for subject indexing based on manual examination of each processed document do not provide a way to scale them up and therefore they are not the best answer for the requirements of the current grey literature publishing.

Scaling is a difficult problem in any situation and we do not strive to provide a definitive answer on how to scale up subject indexing in general or even in the case of grey literature. We propose that this issue can be alleviated by the use of automatic indexing. This solution scales because the time that is being used for subject indexing is *machine time* instead of human time. The processes that are carried out automatically with computers can be scaled up by assigning more computational power to them. Although in fact, our focus is not to achieve full automation of the indexing process so that it may be scaled on demand, but rather *to scale up the number of people* that are able to do the indexing while retaining a reasonably high quality.

Now that we have described the aim we try to achieve with automatic indexing, we will continue with a basic description of what automatic indexing is. It is a process of assigning indexing terms to a document in an automated fashion. It can use techniques based on analysis of language corpora. A common form of automatic indexing relies on "simple" statistical analysis of the full-text. In the case of our indexing system, we have used automatic *term assignment* that selects terms from a controlled vocabulary for a particular document based on the analysis of the document's content. We have employed a *machine learning* approach that is based on computing conditional probabilities.

The approach we have chosen employs *supervised machine learning* that gathers feedback from users of the indexing system. Machine learning modifies the automatic indexing functionality with regard to the set of training data on which it "learns" how to do good indexing. The approach of supervised learning adjusts the indexing algorithm based on the newly acquired information about the way the indexing system is being used. Each time user approves of a set of automatically generated subject terms this decision is fed back into the indexing model and makes it more aligned with the specifics of the document collection.

One of the fundamental requirements of automatic indexing is the access to the *full-text* of the examined document. However, this is not a problem in the use case we have described. If the indexing of a document is done by its author, the access to the document's full-text is granted.

Automatic indexing consists of a sequence of processes. During the course of automatic indexing the full-text is processed by a number of procedures that are collectively referred to as the machine processing *pipeline*. The full-text is sent through a sequence of processes that take the text as their input and pass their output to the process that is directly after them in the pipeline's sequence.

In our case we start with a procedure that yields a *plain-text* of the document in question which might have been in another format, such as PDF or MS Word. Once we have acquired a plain-text embodying the content of the document a series of normalizations and helper procedures are run on it.

One of them removes the *stop words* - the words that do not affect the meaning of the document, such as prepositions or conjunctions. The system contains a list of stop words that are automatically excluded from further processing.

Another common technique that we take advantage of is *stemming* which reduces words to their root forms. In this way, we get rid of inflections, plural suffixes, and other characters that differentiate among the derivatives of the same root form. This method is supposed to collapse the different forms that refer to the same meaning to one word form so that a more effective computation can be done with it.

After these pre-processing steps automatic indexer carries on with its main suite of functions; it analyses the full-text and outputs a set of suggested indexing terms that can be assigned to the processed document. Since this paper is not about automatic indexing itself, but rather its application for grey literature, we will not discuss this part in detail and instead, we will move to the actual implementation of the indexing system.

4 Implementation

After having described the field of automatic indexing in general we will now proceed to provide an overview of the way we have implemented the automatic indexing system and put the methods of automatic indexing for grey literature into practice.

The guiding principle of our implementation was *re-use* of existing components, which we combined together in a document processing pipeline, or extended them in the cases where there was a need for it. This way of development would not have been possible if the parts we were building with had not provided access to their source code. Hence, their *open source* nature enabled us to modify them and extend their functionality. The combination of the constituent parts was possible due to their modular architecture that enabled them to be joined in a chain of processing procedures, which are applied on the examined document.

Not only the software that we have used in the automatic indexing system was open source, the data is communicated in this system in *open formats*. To illustrate this point, the subject headings system we have used was already available in RDF[i] data format expressed with SKOS,[ii] an established standard for representing knowledge organisation system, such as thesauri, subject headings systems, or systematic classifications. It has gone through the standardization process and has reached the status of a recommendation of the World Wide Web Consortium.[iii] This open standard is well supported by the indexer we have used, which eliminated the necessity of data conversion to a suitable format.

In order to manage the flow of control in the system a unifying data communication format is used. For this purpose we have adopted *JSON*, a light-weight data communication format.[iv] The parts of the system exchange short JSON messages to pass the data needed for the indexing process to another part of the system. In this way we have harnessed another standard format to glue the components of the system together.

We wanted to preserve high modularity of the individual components in the system as a whole as well. Therefore we have exposed most of the functionality of the resulting system as a *web service*, which encourages loose coupling and re-use of the system's parts.

Now that we have described the overall architecture and design of the system we will move to the discussion of the individual parts. In the section that follows we will present an overview of the components that are involved in the automatic indexing system we have built.

4.1 Components

We will briefly describe each of the components that together make up for the whole automatic subject indexing system. These are not strictly limited to *software* but they also include *data* that is used in the process of subject indexing – the subject headings system from which the indexing terms are drawn, and the full-text corpus which serves for machine learning algorithms.

According to the design goal we have stated previously, the automatic indexer pipeline is composed mainly of already existing applications that we have re-used for this purpose. The parts that are new serve to connect the re-purposed components. We have written the "*glue code*" that ties the parts that are used in the process of automatic indexing. In order to connect all the components together and set up the indexer for processing Czech language, only a few additional functions had to be implemented.

4.1.1 Subject Headings System

One of the core components of the system is the controlled vocabulary of subject headings we have used. It provides indexing terms that are assigned to documents, which enables to maintain a degree of indexing's consistency by referring only to indexing terms that are *authorized* by the subject headings system. The use of a controlled vocabulary implies that the suggested indexing terms are more consistent compared to the keyword extraction techniques.

The subject headings system we have used is the *Polythematic Structured Subject Heading System* (further abbreviated as PSH).[v] PSH is a bilingual Czech-English controlled vocabulary maintained and used at the *National Technical Library*.[vi] It is a universal system and it consists of headings describing all major aspects of human knowledge. Its structure is similar to thesauri with hierarchical, associative, and equivalence relationships. PSH is primarily expressed in the *MARC 21 Format for Authority Data*,[vii] but it was also converted to RDF data format, expressed with SKOS, which is more suitable for the automatic indexer we have employed.

4.1.2 Digital Repository

The "*host environment*" in which the system of automatic indexing is built in is the digital repository software. The software we have used for this purpose is *CDS Invenio*.[viii] This software's modular architecture enabled us to extend its functionality with a new plug-in that does the automatic indexing.

Invenio processes newly submitted documents in a series of step that are referred to as the document workflow. We have inserted the automatic indexing pipeline into the workflow so that a document can go through this additional suite of procedures to be enriched with subject indexing terms.

The user interface of the automatic indexing system is included in the repository. To achieve this level of integration not only we had to extend the functionality of the modified software but also alter the presentation interface to enable the user to access the added new functionality. This was possible due to the clear separation of the code responsible for the repository's core functions and the templates that build up the interface the user is interacting with.

4.1.3 Automatic Indexer

The component that is responsible for the main function of the system is the automatic indexer. We have chosen to use *Maui Indexer*.[ix] The author of this software claims that it produces *human-competitive indexing*[4] and it seems correct from the results of the comparative studies done with this indexer.[x] This assertion is based on the implicit presumption that the subject terms assigned by humans are the standard with which the quality of automatic indexing is compared. The indexing produced by Maui Indexer is thus comparable to human indexing both in terms of quality and consistency, which is precisely the result we are looking for in our automatic indexing system.

The software is described as being independent of the domain and language for which it is used. However, to achieve the best precision some adjustments need to be done. Because of the language of the document collection, for which the automatic indexer was intended, we wanted to adapt Maui Indexer for the Czech language. The modifications involved changing of the indexer's parts: the list of stop words and the stemmer code.

In the ideal case, stop words would be based on the corpus of documents that we want to index. We chose to use the *Czech National Corpus*[xi] instead to create a list of the most frequent words that may be used as stop words. Czech National Corpus is a vast document collection reflecting the contemporary written Czech [2]. Thus, it served well to establish a good baseline with respect to our document collection.

To reduce words in an analysed full-text to their root forms we have taken over the *aggressive Czech stemmer* [1], which we have adapted in a way so that it can be plugged in the Maui Indexer's source code.[xii] The aggressive nature of the stemmer is based on the approach it takes to stemming non-root word forms. It addresses the morphological characteristics of the Czech language to normalize the irregularities of inflection, consonant alterations and the like. However, in some situations, it may remove characters that are necessary for the distinction of the word sense and thus create the same root form from multiple words that do not share such a root. This feature may compromise the quality of resulting indexing and it is the reason why we consider it as an immediate target for further refinement of our indexing solution.

4.1.4 Text Corpus

The procedure built by combining the previously mentioned components was applied to a text corpus of the document collection of grey literature documents stored and maintained in our digital repository. In our case, we have applied the automatic indexing system we have built to the *National Repository of Grey Literature*.[xiii] This repository, maintained at the *National Technical Library*, collects grey literature from the network of cooperating partner institutions, ranging from the institutes of the Academy of Sciences of Czech Republic to public universities [6].

The contents of the documents included in this repository are mostly in Czech. This was the primary drive behind the decision to enhance the functionality of the automatic indexer towards the Czech language. We have also taken into account that the contents of the repository are produced in collaboration with the partner institutions. Its long-time goal is to have the co-operating institutions produce the document's descriptive metadata, including subject indexing, on their own without a need for central co-ordination. This intention made for an adequate use case of the author-generated subject indexing.

4.2 User interface design considerations

While the quality of the underlying procedures is certainly crucial to produce results of a reasonable quality, the part of the indexing system that has a comparable significance is the *user interface*. The importance of the user interface design stems from the necessity of user feedback. The way how the users provide feedback on the automatically generated set of indexing terms needs to be designed carefully to take advantage of the author's knowledge about the indexed document. The resulting design has some notable features that may have a significant influence on the user experience with the indexing system.

We have decided to provide the automatic indexing as an *opt-in* procedure, which means that the user has to actively declare that the document entered into the repository should be processed with the automatic indexer. If the user checks in a box for automatic indexing, during the next step in the document workflow there will be an additional screen containing the suggested indexing terms and the functionality that allows to modify them.

The primary functionality of the automatic indexing system we have developed is to suggest a list of subject headings that in some way describe the processed document. Its objective is to facilitate non-professional indexing while maintaining a high level of consistency. It is not meant to serve as a *replacement* for the person doing the indexing, but in fact, it is a start for the indexing process controlled by the human user.

We still see the main value being added to the processed documents by the user of the system rather than generated by the system itself. For this purpose we have enhanced the user interface with helper functions that are meant to facilitate more effective indexing.

One of these functions is the *auto-complete* feature. To bridge the gap between the language of the user and the language of the knowledge organisation system used for indexing the user interface displays a list of suggested headings based on a heading's fragment supplied by the user.

To assist the user in deciding on the subject heading's adequacy we have added a utility that shows citations of the example documents indexed with the subject heading user considers to use. When user highlights certain subject heading a short list of links to the documents which have the same heading attached is presented in the interface. In this way, the user can *learn by example* to consider the applicability of a given subject heading for a particular document.

We had to make alterations to the *search interface* as well. There would be no value in added indexing terms if one could not use them to search and navigate the document collection in which they are used. To reflect this added structure a change to the user interface needs to be made so that it is possible to harness the indexing terms to access documents. In our case, responding to this requirement consisted in adding a new search index built for the subject headings and appending a new field to the search form to access this index.

5 Future Possibilities and Challenges

The work we have done with the automatic indexing system for grey literature is by no means finished and we are aware that there are further possibilities for improvement and challenges that have to be solved to deliver a better system. In the specific use case we have described in this paper there are certain aspects of the automatic indexing that are worth underscoring.

It is important to note that there is no use in adding subject indexing terms if such metadata cannot be harnessed via the search or navigation interface for the document collection. If the indexing is not reflected in end-user interfaces it does not provide any additional value. Thus, the indexing must be represented in user interfaces to have an impact on the overall functionality of a digital repository. As we have mentioned in the previous section, we have extended the digital repository with a new search field to access the documents through subject headings to address this issue.

We have to take into account that the indexing we are dealing with here is still a *user-generated indexing*, even though it is somehow refined by the system we have implemented. This implies that the resulting indexing terms might need further verification by a professional if we want to have a quality control in place. As we have sketched in the paper, this is the way our system works and will work for the near term future. Nonetheless, our goal is eventually to get rid of this necessity once we will have a higher level of confidence in the non-professional subject indexing that is fed into the repository.

Due to the modular nature of the whole system, there is a plethora of ways how it can be enhanced and developed further. Every part of the document processing pipeline can be considered for an improvement. Our aim is first to focus on the parts that affect the quality of the results of indexing the most. After every change we want to check if it leads to an increase in the system's precision by comparing it with the results the system had on the same document with the previous configuration.

We argue that the system not only benefits grey literature authors and maintainers of digital repositories, but, moreover, it can also benefit the individual components it is made of because it reflects on the way the component is being *used*. This can be applied on the knowledge organisation system from which the indexing terms are drawn. The data about its usage in practise can be crucial for its development and further evolution reflecting the changing needs of the user community and a shift in the way its concepts are perceived.

6 Conclusions

This project would not have been possible if there were not *open standards* that govern the field of subject indexing. They enabled the re-use of existing components adhering to certain standards and their combination in a novel way. The layer provided by open standards constituted an environment for interoperability and systems built with an open architecture in mind.

All the parts we have put together in this open framework are *open source*. This means they are open to modifications and extensions, and those were necessary for the system to work as a whole. Moreover, due to the their modular character it was possible to switch one part for another or plug in a new component.

The indexing system that came out of this way of development is applied to the grey literature documents. It was designed to reflect the nature of grey literature. We have argued that the situation of subject indexing of grey literature is unsatisfactory and we have expressed our view of the causes for such state. Our motivation was to react to the current conditions and propose an approach that may lead to an improvement of the way subject indexing is done for grey literature.

References

[1] DOLAMIC, Ljiljana; SAVOY, Jacques. Indexing and stemming approaches for the Czech language. *Information Processing & Management*. November 2009, vol. 45, iss. 6, p. 714 – 720. ISSN 0306-4573. DOI 10.1016/j.ipm.2009.06.001.

[2] KUŘERA, Karel. The Czech National Corpus : principles, design, and results. *Literary and Linguistic Computing.* 2002, vol. 17, no. 2, p. 245 – 257. ISSN 0268-1145.

[3] MACKENZIE OWEN, John S. The expanding horizon of Grey Literature. In *Perspectives on the design and transfer of scientific and technical information : proceedings of the 3rd international conference on grey literature*. Amsterdam : TransAtlantic, 1998, pp. 9–13. Also available from WWW: <http://cf.hum.uva.nl/bai/home/jmackenzie/pubs/glpaper.htm>.

[4] MEDELYAN, Olena. *Human-competitive automatic topic indexing*. Waikato, 2009. 214 p. Dissertation thesis (PhD.). University of Waikato, Department of computer science, 2009. Also available from WWW: <http://www.cs.waikato.ac.nz/~olena/publications/olena_medelyan_phd_thesis_July2009.pdf>.

[5] PEPE, A. [et al.]. CERN Document Server Software : the integrated digital library. In DOBREVA, Milena; ENGELEN, Jan (eds.). *9th ICCC International Conference on Electronic Publishing : from author to reader : challenges for the digital content chain.* Leuven : Peeters, 2005. ISBN 90-429-1645-1.

[6] PEJŠOVÁ, Petra (ed.). *Grey literature repositories*. Zlín : VeRBuM, 2010, 156 p. ISBN 978-80-904273-6-5.

[7] RIBEIRO, Fernanda. Subject indexing and authority control in archives : the need for subject indexing in archives and for an indexing policy using controlled language. *Journal of the Society of Archivists.* April 1996, vol. 17, iss. 1, p. 27 – 65. ISSN 0037-9816.

[8] SYKES, Jan. *The value of indexing : a white paper prepared for Factiva, a Dow Jones and Reuters Company* [online]. February 2001 [cit. 2010-12-02]. Available from WWW: <http://4info-management.com/pdf/indexingwhitepaper.pdf>.

[9] VLACHIDIS, Andreas [et al.]. *Excavating grey literature : a case study on rich indexing of archaeological documents by the use of natural language processing techniques and knowledge based resources* [online]. 2009 [cit. 2009-09-30]. Preprint for ISKO UK 2009 : Content Architecture: Exploiting and Managing Diverse Resources. Available from WWW: <http://www.iskouk.org/conf2009/papers/vlachidis_ISKOUK2009.pdf>.

[i] Resource Description Framework. <http://www.w3.org/TR/rdf-concepts/>.

[ii] Simple Knowledge Organisation System. <http://www.w3.org/TR/skos-reference/>.

[iii] http://www.w3.org/

[iv] http://www.json.org/

[v] The on-line version is available at http://psh.ntkcz.cz/skos/

[vi] http://www.techlib.cz/en/

[vii] http://www.loc.gov/marc/bibliographic/

[viii] The project's website is at http://invenio-software.org/ and our installation of Invenio is available at http://invenio.ntkcz.cz/.

[ix] http://code.google.com/p/maui-indexer/

[x] Examples can be found at http://code.google.com/p/maui-indexer/wiki/Examples

[xi] http://ucnk.ff.cuni.cz/english/index.php

[xii] The stemmer we have used is available at http://members.unine.ch/jacques.savoy/clef/CzechStemmerAgressive.txt.

[xiii] http://nrgl.techlib.cz

A Profile of Italian Working Papers in RePEc

Rosa Di Cesare, Daniela Luzi, Marta Ricci
Institute of Research on Population and Social Policies, IRPPS/CNR, Italy

Roberta Ruggieri
Senato della Repubblica, Italy

Abstract
This paper describes the results of an analysis of Italian Working papers (WP) available both in RePEc (Research Papers in Economics) and in Institutional Repositories (IR) and websites. Given that RePEc is a disciplinary repository based on the active involvement of economic institutions, rather than authors, our analysis intends to explore the institutions' propensity for making their collections available both in disciplinary and Institutional repositories. Therefore, the paper provides a profile of the Italian Economics institutions participating in RePEc as well as an in-depth analysis of the their availability WPs and WP series. Moreover, IRs and websites of the Italian institutions participating in RePEc were analysed to compare the scientific contents available in these important sources of free access information (RePEc, IRs and websites).

1. Introduction

RePEc (Research Papers in Economic) is one of most important disciplinary repositories, which covers different aspects of research in Economics, and gathers the largest collection of working papers. Founded in 1997, it provides users with a variety of services, ranging from searching facilities for document (IDEAS) as well as research institutions profiles (EDIRC) to a provision of access statistics for items and authors (LogEC) as well as tools for citation analysis (CitEc). This decentralized repository is primarily based on an interconnected network of over 1000 interoperable archives supported by an eclectic mix of participants, from the major commercial publishers, university presses, research centres, central banks to university departments in 80 countries worldwide. This makes RePEc different from other disciplinary repositories. It is based on the collaboration with Economics institutions, which make their collections retrievable by RePEc using both a common bibliographic template for content descriptions and a protocol to exchange data. Only recently has RePEc set up a service that allows single authors to submit their publications, but they are allowed to do it only when they belong to institutions lacking a RePEc archive.

In the framework of the Open access movement, Institutional and disciplinary repositories represent complementary communication channels to enhance the visibility and impact of scientific results. Generally, disciplinary repositories tend to be populated with a greater number of papers compared with the ones available in Institutional repositories (IR). This fact has been underlined in several studies [Swan, 2005; Kingsley 2008; Bjoerk et al., 2010] pointing out that authors show a greater propensity to submit their work to thematic archives rather than self-archiving their works in IRs. Different explanations have been given, ranging from publication practice of specific scientific communities to the capacity of institutions to actively involve the different stakeholders (authors, librarians, information managers, etc.) of the research lifecycle.
Given that RePEc is based on the active involvement of economic institutions, rather than authors, our analysis intends to explore the institutions' propensity for making their collections available both in disciplinary and institutional repositories. For this reason this paper provides first a profile of both Italian economics institutions and their production in terms of Working papers (WP) and WP series listed in RePEc (paragraphs 4.1. – 4.3.) and then analyses whether these institutions make their WP series also available in their IRs and/or web pages (paragraphs 4.4 - 4.5). Therefore, our analysis intends to contribute to the identification of successful strategies to increase the impact of research results within an open access environment.

2. RePEc characteristics

Even if RePEc is currently listed among the most successful disciplinary repositories (the largest after the well-known arXiv.org), its own founders generally refers to it as a bibliographic service or database [Zimmermann, 2009], a de-centralised non-commercial digital library [Barrueco et al., 2000], a decentralised academic publish system, an open library [Krichel, 2001]. This is due principally to two separate, main features: namely its historical development and its organisational model.

RePEc dates back to 1997 and is based on the previous NetEC project founded in 1993 by Thomas Krichel as a collection of projects (BibEc, WoPEc, CodEc, WebEc, BizEc and HoPEc) aiming at distributing information relevant to Economics and in particular focused on WPs diffused via Internet. It is interesting to read about its development in the pre-history of WWW [Krichel, 1997; Karlsson, 1999] because each project represented the effort of both tracking different types of information (from the print working papers in BibEc to the first digital ones in WoPEc, software codes used in Economics, collection of web pages in WebEc, etc.) and progressively establishing the active involvement of different people and institutions in sharing their resources taking advantage of the new information technologies (at the time, gophers, mailing lists etc.).

Moreover the current RePEc is also the result of constructing a cooperation model, which is suited to its scientific community and its publishing practice. For instance the development in 1997 of a centralised "Economic Working Paper Archive" at Washington University based on the model imported from the High Energy community was abandoned in favour of a decentralised database. The economists' "built-in distrust of monopolies" reported by Krichel [2001] coupled with, at that time, over 200 retrievable archives of working papers made more practical to let each institution manage its own collection locally and then make them centrally accessible on a common interface. Therefore, the establishment of a common bibliographic template for content descriptions (called ReDIF) as well as the development of a protocol (called Guildford) to exchange data represented one of the most important achievements, which are still the basic architectural framework of the current RePEc.

Under this perspective RePEc can be considered a distributed digital library, a "collection of metadata records" each one identified by a unique handle, which allows the linkage between records. Hence, the definition of a decentralised bibliographic database, based on the relations between "resource" (i.e. any output of an academic activity: research documents, datasets, computer programs), the resource logical grouping in "collection" (i.e. working paper series and journals), as well as "person" and "institution" [Barrueco, 1999].

These four elements are the core RePEc services, built upon the archives made available by the collaborating institutions on an http or ftp server. They are briefly described hereafter.

- EDIRC (Economics Departments, Institutes and Research Centers in the World) is the service that indexes economics institutions worldwide by countries and fields. It provides detailed information on the institution structure, listing the affiliated "sub-entities" (for instance in the case of universities EDIRC reports the belonging departments, institutes, and/or research centres connected with economics studies). Moreover, EDIRC provides also a list called "*Top 25% Institutions and Economists*" that ranks institutions in each participating country according to a set of criteria described in detail in Zimmermann [2009].
- IDEAS (Internet Documents in Economics Access Service) is the service that provides the user interface to browse and search RePEc scientific contents (journals, working papers, books, book chapters, software components);
- RePEc Author Service allows authors to register in RePEc, proving information on their name variations so that metadata matching as well as work attribution are facilitated. Authors also provide information on their affiliations. Only when registered is an author ranked together with his/her institution and obtains notifications of new citations found in RePEc.
- MPRA (Munich Personal RePEc Archive) is the only central archive where researchers can submit their works only when their affiliated institution does not participate in RePEc.

Besides organising this collection of archives and making their data freely available, RePEc provides a set of additional intermediary user services:

- NEP (New Economics Papers) is a notification service of new downloadable WPs for over 40 specific fields. Voluntary editors compile subject specific reports that filter RePEc new additions to provide subscribers with update information constituting a "simple form of peer review" [Bátiz-Lazo, 2005].
- LogEC is a service that provides access statistics for each item.
- CitEC (Citation in Economics) is the service that provides an autonomous citation index of many electronic documents distributed by RePEc (74% at the time of writing). This service is maintained by Josè Manuel Barrueco at the University of Valencia.

The variety of services provided by RePEc coupled with different functionalities and information needs succeed to "describe the discipline, rather than simply the documents" [Krichel, 2001] produced by the Economics scientific community.

3. Objectives and Methods

Our analysis was driven by the type of organisation model that characterises the RePEc interconnected network and in particular by the active role played by Economics institutions in making their collection retrievable, thus taking advantage of RePEc services to increase their visibility and impact within an international scientific community. Under this perspective RePEc is worth analysing for the following main reasons: a) unlike the majority of e-print archives, it is based on collaboration with different Economics institutions which make their collections retrievable by RePEc, or more precisely by the IDEAS Service, b) it is the largest open source of WPs in Economics and related fields, and c) WPs contribute to determine institutions' and authors' ranking positions measured by RePEc bibliometric analysis. For these reasons RePEc as well as the IRs and websites maintained by the Italian institutions participating in RePEC are valid sources of analysis to achieve our overall objective, that is to explore the institutions' propensity to make their collections available both in disciplinary and institutional repositories in Italy.

Our analysis is divided into two parts. In the first we have analysed the Italian contribution to RePEc in order to identify, on the one hand types of economics institutions participating in RePEc, and on the other, features and characteristics of the WPs made available in this disciplinary repository. This analysis was performed using the following RePEc data sources:

a) EDIRC list of entity and sub-entities participating in RePEc;
b) IDEAS list of Italian WP series;

The analysis of both Italian institutions participating in RePEc and their WP collections was performed at two levels. In the first data was gathered from the entire set of WP series listed in IDEAS, while in the second, data was collected from the WP series reported in the list of *"Top 25% Institutions and Economists in Italy".* This list ranks institutions on the basis of the number of authors registered in RePEc Author service, Institutions listed in EDIRC, bibliographic data collected by RePEc, as well as citations counted by CitEc and access statistics registered by LogEc. The comparison of the data of the entire set of Italian contribution with that contained in the list of the best-ranked institutions allowed us to identify the share production of the best-ranked Italian institutions, in terms of consistency, stability over time, etc.

In particular, the Italian WPs contribution in RePEc was analysed in terms of:

- WP series characteristics, considering:
  - Number of series registered;
  - Longevity (live, dead series),
  - Vitality (young and new-born series);
- WP characteristics, considering:
  - Number of WP for each series;
  - Average number of WPs per series and per year.

Within the Italian WP series we excluded those produced by and/or in collaboration with commercial publishers that turned out to be journal collections.

As RePEc is continuously updated, the analysis of Italian WP series referred to the month of July 2010.

In the second part of our study the best-ranked institutions listed in the *Top 25%* were considered in order to determine whether:

- These Institutions had an IRs;
- The series listed in IDEAS were available also in the IRs and/or in websites;
- The documents' temporal coverage in IRs and/or websites corresponded to that registered in RePEc;
- IRs and/or websites provided access to other types of GL documents.

This analysis also included those institutions that are listed in the *Top 25%*, but do not register any WP series in RePEc.

4.Results

4.1. Italian WPs in RePEc in a European context

The production of more than 80 countries worldwide is retrievable in RePEc. To give an overview of the Italian production of WPs at international level, we selected the WPs made available by United Kingdom, Germany, France and Spain. These countries were selected because they were among the major contributors to the SIGLE database.

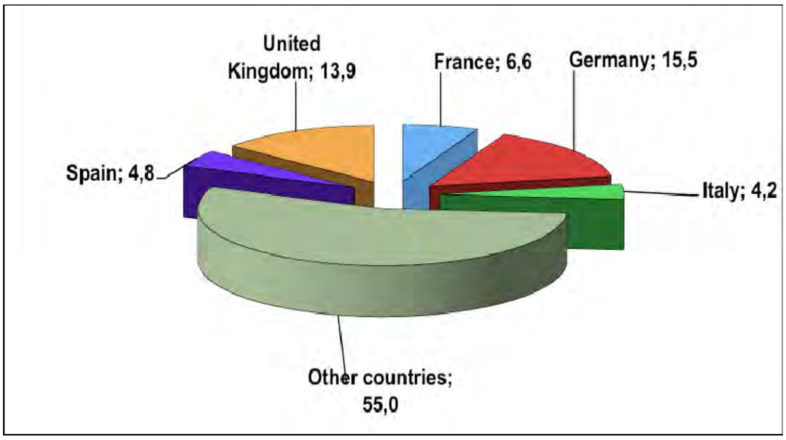


Figure 1 . - WPs of some European countries in RePEc (%)

Italy participates in RePEc with 15064 WPs, which is 4% of the entire database. The major WP producers in our sample are: Germany with 55062 WPs and the United Kingdom with 49425 WPs.
The French and Spanish contributions are closer to that of Italy.

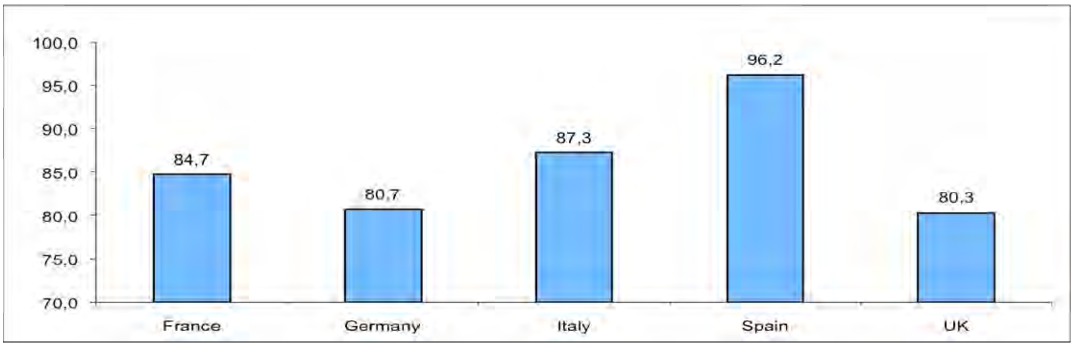


Figure 2 . – Online WPs of some European countries in RePEc (%)

If we consider the online WPs (fig. 2) we can notice that most of them are available in full text, with Spain reaching the highest percentage, followed by Italy and France.

4.2. Italian institutions in RePEc

Table 1 shows the composition of the Italian participation in RePEc. Within a total number of 369 providers, the major contributors are universities represented by their departments, faculties, and centres, reaching all together 71.2%. Other types of organisations are Research centres, Foundations, Governmental institutions and Associations.

Table 1. – Number and percentage of Italian institutions in REPEC and in Top 25% by type of providers

			Top 25% institutions	
Provider	**No.**	**%**	**No.**	**%**
University department	130	35.2	45	51.1
University centre	82	22.2	10	11.4
University faculty	51	13.8	20	22.7
Research centre	31	8.4	8	9.1
Foundation	22	6.0	4	4.5
International organization	8	2.2	1	1.1
Governmental institution	13	3.5	–	--
Association & Society	32	8.7	–	--
Total	**369**	**100.0**	**88**	**100.0**

Considering the composition of the Italian providers listed in the *Top 25%* (table 1), there are 88 best-ranked institutions; the majority of them are University departments, followed by faculties and University centres. Compared with the figures of the total number of Italian institutions, we can notice that the presence of University sub-entities is more relevant increasing in particular in the case of University departments (51.1% vs. 35.2%), Faculties (22.7% vs. 13.8%) as well as for Research centres (9.1% vs. 8.4%).

4.3. Italian WP collections in RePEc

Working papers have long been a mainstay of scientific output in Economics. Like other types of GL documents, WPs attenuate the gap of publication delay and editorial space limits, thus proving updated and comprehensive information on specific research areas. As reported in Krichel [2001], "Economists do not issue preprints as individuals; rather economic departments and research organizations issue working papers". Additionally it is important to point out that Economics institutions generally organize this type of report literature in collections, providing for each of them a title that specifies the topic of the series and progressive numbers in the collection (known as report number). This editorial activity differs from the one carried out by commercial publishers in two aspects: there is no predefined number of issues/WPs to be published in one year, and generally there is no formal peer-review procedure that is generally dependent on internal institutional rules and practices. Nevertheless, many of these WPs represent important reference points in the scientific literature: they are often cited in journals and may represent the official position of important institutions in key issues. Hence, the importance of their free, rapid and wide diffusion through different channels, Institutional repositories, websites and of course disciplinary repositories like RePEc that allows users to gain a more comprehensive picture of the production of WPs at international level.

Given these premises, our hypothesis is that the editorial activity carried out within an Economics institution in issuing and managing WP collections is comparable with the commercial publishing process of scientific journals. Therefore, in the measurement of the consistency of WPs and WP series we focused our attention in particular on:

- The *stability* of collections, i.e. their continuity over time, that according to our hypothesis indicates a well consolidated research area, whose results are progressively issued and circulated under a specific series title;
- The *novelty* of collections, i.e. the issue of new collections that indicates the setting up of new research areas in which institutions are acquiring scientific results that need to be diffused, generating an *ad hoc* new series. That is a similar procedure to that which is adopted when new journals are launched.

To describe these characteristics we used two indicators: *longevity* and *vitality*.
To measure "longevity", we classified WP series as following:

- *Live WP series*, that is series that are still available in RePEc for the years 2009 or 2010;
- *Dead WP series*, that is series that are no more available either for the years 2009 or 2010.

To measure "vitality", within the live series, we classified WP series as following:

- *Young series*, that is series available in RePEc from 2007 on;
- *New born series*, that is series available in RePEc since 2009 or 2010.

Moreover, considering that each provider produces different numbers of WPs over different periods of time, we measured the WP consistency in terms of average number of WPs within each series and the average annual contribution, as following:

- *Series' average weight*: is the average number of WPs contained in a series;
- *Annual average contribution:* is the average number of series provided by an institution in one year.

These indicators have been applied to the entire set of Italian WP series as well as to the series produced by the best-ranked institutions.

4.3.1. Characteristics of WPs series: "longevity"

There are 145 WP series listed in RePEc (table 2), the major producers are University departments (53.8%), Research centres (14.5%) and International organisations (15.9%). According to the previously provided classification of *longevity*, the majority of RePEc providers contribute with series that have a stable and continuous production over the years (109 out of 145, equal to 75.2 are *live* series), and this is particularly evident in the case of Foundations, Governmental institutions, University departments, and Research centres. The only exception is represented by International organisations that have a high percentage of series that finish in 2008 or before. These collections are produced by FAO and UNICEF, the latter in particular has a high number of *dead series* that contain project results. Therefore we can assume that a collection is closed when the project is finished, following the project lifecycle.

Table 2. - RePEc series according to "longevity"

Provider	RePEc series		Live		Dead	
	No.	%	No.	%	No.	%
University department	78	53.8	64	82.1	14	17.9
University centre	13	9.0	11	84.6	2	15.4
University faculty	3	2.1	2	66.7	1	33.3
Research centre	21	14.5	18	85.7	3	14.3
Foundation	4	2.8	4	100.0	--	--
International organization	23	15.9	8	34.8	15	65.2
Governmental institution	1	0.7	1	100.0	--	--
Association & Society	2	1.4	1	50.0	1	50.0
Total	**145**	**100.0**	**109**	**75.2**	**36**	**24.8**

Table 3. – Best-ranked RePEc series according to "longevity"

Provider	RePEc series		Live		Dead	
	No.	%	No.	%	No.	%
University department	41	58.6	37	90.2	4	9.8
University centre	9	12.9	9	100.0	--	--
University faculty	--	--	--	--	--	--
Research centre	14	20.0	12	85.7	3	14.3
Foundation	4	5.7	4	100.0	--	--
International organization	2	2.9	1	50.0	1	50.0
Total	**70**	**100.0**	**63**	**90.0**	**7**	**10.0**

The analysis of the longevity indicator applied to WP series produced by the best-ranked institutions is reported in table 3. Comparing these data with the entire set of WP series in RePEc, we can notice that the percentage of live series increases reaching 90% in total providing a higher degree of stability and continuity in production of the best-ranked institutions.

Moreover, it is worth underlining that 70 WP series are produced by the best-ranked institutions representing almost 50% of the total number of the Italian WP series listed in RePEc and this confirms that WP production contribute to determine the positions of the institutions' ranking.

4.3.2. Characteristics of WPs series: "vitality"

The analysis of the vitality indicator applied to the entire set of Italian RePEc WP series is reported in table 4.

Table 4. - RePEc series according to "vitality"

Provider	Live RePEc series		Young		New-born	
	No.	%	No.	%	No.	%
University department	64	58.7	8	7.3	3	2.8
University centre	11	10.1	1	0.9	1	0.9
University faculty	2	1.8	--	--	--	--
Research centre	18	16.5	1	0.9	4	3.7
Foundation	4	3.7	--	--	--	--
International organization	8	7.3	1	0.9	--	--
Governmental institution	1	0.7	--	--	--	--
Association & Society	1	0.9	--	--	--	--
Total	**109**	**100.0**	**11**	**10.1**	**8**	**7.3**

Among the live series, there is a low percentage of young series (10.1%), which are mostly produced by University departments, while new-born series (7.3%) are in particular produced by Research centres (3.7%). This highlights that such production is concentrated within research environments.

Table 5. – Best-ranked RePEc series according to "vitality"

Provider	Live RePEc series		*Young*		*New-born*	
	No.	**%**	**No.**	**%**	**No.**	**%**
University department	37	58.7	4	6.3	1	1.6
University centre	9	14.3	2	3.2	--	–
University faculty	--	--	--	--	--	–
Research centre	12	19.0	1	1.6	3	4.8
Foundation	4	6.3	--	--	--	–
International organization	1	1.6	--	--	--	–
Total	**63**	**100.0**	**7**	**11.1**	**4**	**6.3**

The same classification applied to the series produced by the best-ranked institutions (table 5) gives similar results. There is a small increase in the percentage of *young* WP series, while *new born* have similar values compared to the results obtained in the entire set of Italian WP series in RePEc.

4.3.3 Characteristics of WPs series: WPs production over time

As seen in paragraphs 4.3.1. and 4.3.2., Italian WP series in RePEc represent stable collections, which provide scientific results on institutional consolidated research areas over a continuous period of time. Now, in order to verify whether WP collections produced by Italian Economics institutions in RePEc also have features similar to scientific journals, we analysed the number of WPs within each series and measured the average number of WPs within each series as well as the average annual contribution.

Table 6. – RePEc annual Italian contribution by type of provider

Provider	No. of series	No. of WP	Series' average weight	Annual average contribution
University department	77	6671	86.6	11.7
University centre	14	1542	110.1	15.0
University faculty	3	140	46.7	4.2
Research centre	21	1651	78.6	11.0
Foundation	4	1380	345.0	37.3
International organization	23	614	26.7	4.5
Governmental institution	1	18	18.0	18.0
Association & Society	2	73	36.5	9.1
Total	**145**	**12089**	**83.4**	**11.6**

Table 6 shows the results of this analysis. 145 series contain a total number of 12089 WPs. The general average number of WPs is 83 papers for each series, while the average of WPs produced in each year for each series is 11,6. Considering now the different types of providers, Foundations produce on average a very high number of WPs, followed by University centres and University departments. We encountered a similar result when we measured the annual average contribution.

Table 7. - RePEc annual Italian contribution by type of best-ranked provider

Provider	No. of series	No. of WP	Series' average weight	Annual average contribution
University department	41	4573	111.5	12.5
University centre	9	1305	145.0	15.5
University faculty	--	--	–	--
Research centre	14	1496	106.9	14.4
Foundation	4	1380	345.0	37.3
International organization	2	58	29.0	5.3
Total	**70**	**8812**	**125.9**	**14.6**

The same measure applied to the top ranked institutions (table 7) shows first of all that the general average is much higher, with a higher number of WPs for each series (125.9), confirming that Foundations and University centres produce a number of WPs above the average. Further, the average

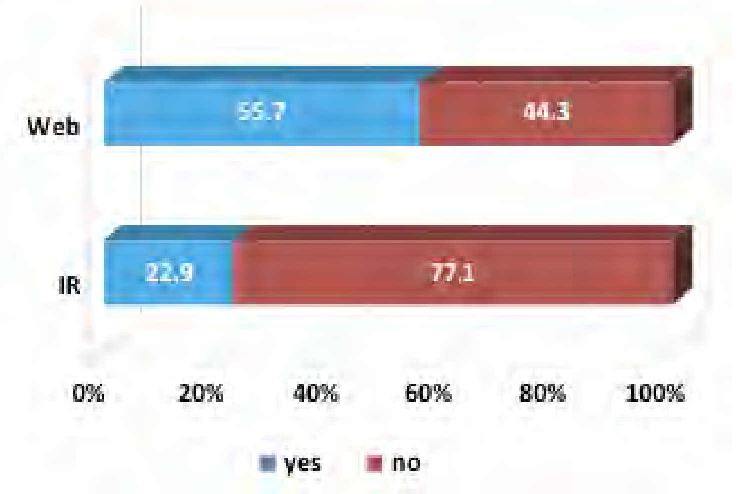
production increases to 14 WPs per year; University centres, Research centres (15.5) and Foundations (37.3) produce an above general average number of WPs.
Given this data we can compare WPs series in Economics with commercial journals considering that journals produce at most 12 issues per year and that the number of articles in one journal issue is generally much lower that the number of WPs produced in one series.

Moreover, we can notice that 73% of WPs (i.e. 8812 out of 12089 total number of WPs) is produced by the best-ranked institutions and this confirms once again that this production contributes to determining their position in the ranking, being part of evaluation metrics used in RePEc.

4.4. Italian Institutions in RePEc and their IRs and websites

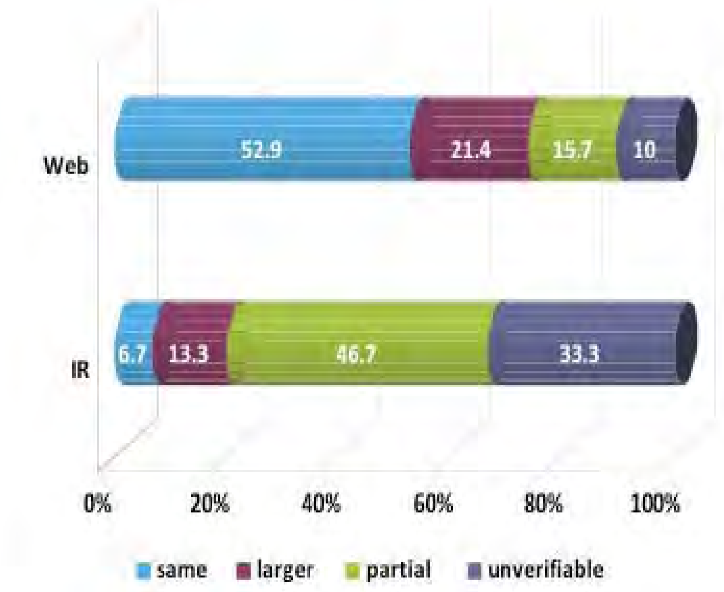
In this part of the study we analysed whether the Institutions participating in RePEc also populate their own IRs and/or diffuse their WP series within their websites. For this reason, sources of our analyses were the IRs of the 88 Italian top ranked institutions as well as their websites.

Fig. 3. Availability of RePEc series in local IRs and websites



Considering University departments, centres and faculties belonging to the same university, there are 35 RePEc providers that can potentially submit their scientific production in the 17 local IRs. Only 22.9% (equal to 8 providers) makes RePEc series also available in their local IRs. Analysing the websites of the 88 providers listed in the *Top 25%*, the situation is different. 55.7% of them (equal to 49 out of 88 providers) list the RePEc WP series in their websites too (fig. 3).

Fig. 4. – Temporal coverage of RePEc WP series in local IRs and Websites



A more detailed analysis was performed on the WPs series available in local IRs and websites in order to verify if there is uniform information content in RePEc, local IRs and websites. To do so, we checked the temporal coverage of each series within local IRs and websites (fig. 4).

Considering Institutional repositories, 6.7% of series produced by the best-ranked institutions in RePEc have the same temporal coverage compared with that available in RePEc, 13.3% a larger temporal coverage, while the majority of these institutions (46.7%) provide a limited temporal range of WP series in their IRs.

For 33.3% of them this data could not be retrieved, and this depends on how repositories organise the bibliographical data of their collections as well as on the ways the scientific community (i.e. University departments, centres and/or scientific groups) are associated with the collections they produce.

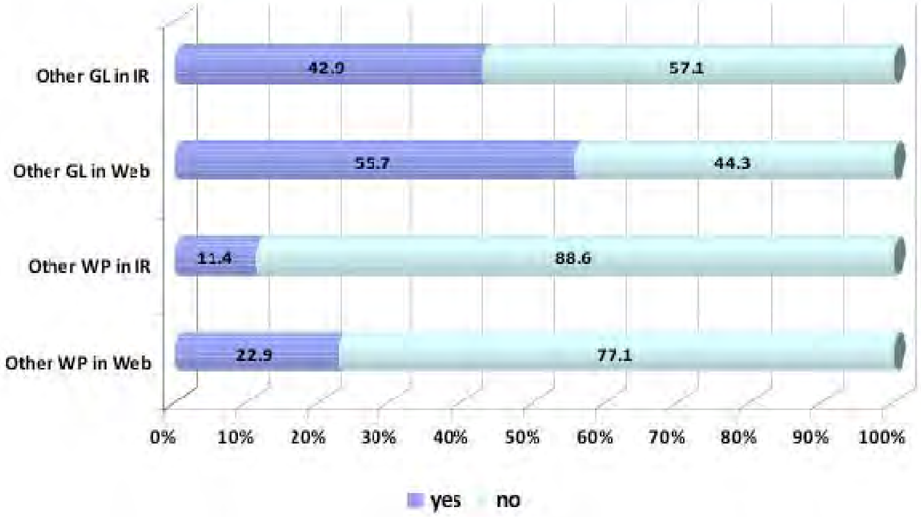
Considering websites, 52.9% of WP series have the same temporal coverage as in RePEc, while 21.4% make their WP series available for a longer period. The differences in temporal coverage can be interpreted in different ways. The richness of information provided in websites can be explained by the fact that research teams generally directly manage their own websites and are obviously directly interested in providing an updated and comprehensive picture of their research activities and results. IRs are unfortunately often managed "outside the scientific community" and further efforts in building efficient collaboration among the different stakeholders (researchers, librarians, information managers) still need to be made. Similar results were also obtained in [Bjoerk et al., 2010].

4.5. Availability of other GL information in local IRs and websites

To conclude our analysis we also checked both IRs and websites to identify whether they contained other types of GL documents. Our intention was to verify if the information providers select types of documents and/or scientific content to be made available locally. We found that both IRs and websites provide information on other GL documents (fig. 5). In the case of IRs (42.9%), they are generally course materials and theses, while in 55.7% of websites they are mainly proceedings, workshops and data set.

If we consider other WPs series available both in IRs and websites, we can see that there is no evidence of information providers selecting the WPs to be retrievable on their local systems, only 11.4% provide information on other WPs series in IRs and 22.9% in websites.

Fig. 5. – Availability of other GL documents in IRs and websites



5 . Conclusions

This study provided a profile of the Italian Economics institutions participating in RePEc associated with the analysis of their production of WPs and WP series. The results of our analysis indicate that among the 369 institutions participating in RePEc the major contributors are academies represented by their belonging to University departments, faculties and centres. They also constitute the core of the institutions listed in the *Top 25%* together with Research centres and Foundations. The analysis of WP series made available in RePEc confirms the important role played by this type of GL document. They contribute to determining the ranking position of the institutions (75% of the total number of WPs are produced by the best-ranked Italian institutions) constituting a large part of the free documents, on which RePEc bibliometric services measure institutions' and authors' impact, using various criteria to determine citations and access statistics [Zimmermann 2009].

Moreover, they represent the main research areas covered by scientific institutions whose results are diffused in a continuous and stable way within WP collections: the majority of WP series are *live*, current collections. Our hypothesis that the institutional activity of editing Economics WP series has features similar to the publication of journals was confirmed by the analysis of WPs contained in the series. In fact, measuring the average number of WPs per series and per year, we found that they contain on average a high number of WPs as well as figures comparable with the publication of issues within scientific journals. This figure is even higher if we consider the WPs produced by the best-ranked Italian institutions. They publish on average more than 14 WPs per year within a series, with peaks above the average in the cases of University centres, Research centres and Foundation.

The results obtained in the second part of the study that aims to analyse the availability of WPs in both Institutional repositories and websites maintained by the 88 best-ranked institutions in RePEc provided a more fragmented picture. In fact, IRs do not seem to represent a preferential channel for WP diffusion: a limited number of Italian institutions additionally makes RePEc series available in their local repositories, the temporal coverage of these collections in IRs is in most cases briefer than in RePEc. Moreover, the identification of RePEc series associated with their producers (University departments, centres or research groups) within IRs is not always a straightforward procedure, depending on the way IRs organise their scientific contents. These aspects may bias the retrieval of this important source of information as well as the visibility of their direct producers.

On the contrary, the websites maintained by the 88 top ranked institutions reflect their scientific production well. In fact, there is a more complete correspondence of WPs listed in RePEc, both in terms of collections available and temporal coverage. Of course the direct involvement of research teams in the development of websites contribute to the achievement of a rich and comprehensive description of their scientific activities and products.

To conclude, our study has outlined a contradictory picture of the propensity to diffuse scientific contents in both disciplinary and local IRs. On the one hand, Italian institutions voluntarily contribute with their WPs and WP collections to build in RePEc a critical mass of information relevant to Economics, on the other, they seem to neglect the role of local IRs in diffusing their scientific results. However, the richness and comprehensiveness of the information available in websites supports the hypothesis that in Italy IRs have not yet succeeded in achieving an active and efficient collaboration among the main stakeholders of the research lifecycle (researchers, librarians, information managers, IR's developers). Of course, this hypothesis has to be confirmed carrying out future, *ad hoc* analysis that can contribute to foster Open access also at local levels.

References

Barrueco Cruz Josè Manuel, Krichel Thomas, (1999). "Cataloging economics preprints: an introduction to the RePEc project", *RePEc and ReDIf documentation*, Shankari, RePEc Team, available at: http://openlib.org/home/krichel/papers/shankari.html

Barrueco Cruz Josè Manuel, Krichel Thomas, (2000). "Distributed cataloging on the Internet: the RePEc project". In: *Metadata and organizing educational resources on the internet*. Haworth Information Press. pp.227-241, available at: http://eprints.rclis.org/bitstream/10760/4010/1/jagt.html

Barrueco Cruz Josè Manuel, Krichel Thomas, (2005). "Building an autonomous citation index for grey literature: the economics working papers case". *The grey journal, an international journal on grey literature*, 1 (2).

Batiz-Lazo Bernardo, Krichel Thomas, (2005). "On-line distribution of working papers through NEP: A Brief Business History". *EconWPA*, available at: http://129.3.20.41/eps/eh/papers/0505/0505002.pdf

Björk B-C, Welling P, Laakso M, Majlender P, Hedlund T, et al. (2010). "Open access to the scientific Journal Literature: Situation 2009". *PLoS ONE* 5(6), available at: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0011273

Foster Nancy Fried, Gibbons Susan, (2005). "Understanding faculty to improve content recruitment for Institutional Repositories". *D-Lib Magazine*, 11 (1), available at: http://www.dlib.org/dlib/january05/foster/01foster.html

Karlsson Sune, Krichel Thomas, (1999). "RePEc and S-WoPEc: Internet access to electronic preprints in Economics". *RePEc and ReDIf documentation*, RePEc Team, available at: http://openlib.org/home/krichel/papers/lindi.html

Kingsley Danny, (2008). "Repositories, research and reporting: the conflict between institutional and disciplinary needs". Paper presented at the VALA 2008: *Libraries/changing spaces, virtual places*, available at:

http://dspace-prod1.anu.edu.au/bitstream/1885/46096/1/117_Kingsley_Final.pdf

Krichel Thomas, (1997). "About NetEC, with special reference to WoPEc". *Computers in higher education economics review,* 11 (1), available at: http://www.economicsnetwork.ac.uk/cheer/ch11_1/ch11_1p19.htm

Krichel Thomas, (2001). "RePEc, an open library for Economics". In: *The Economics and usage of digital library collections*, Ann Arbor, Michigan (US), Revised version 2001 available at:

http://eprints.rclis.org/bitstream/10760/12154/1/salisbury.a4.pdf

Krichel Thomas, Zimmermann Christian (2009). "The economics of open bibliographic data provision". *Economic analysis and policy*, 39 (1), available at: http://ideas.repec.org/a/eap/articl/v39y2009i1p143-152.html

Swan Alma, Brown Sheridan, (2005). "Open access self-archiving: An author study". [Departmental technical report], available at: http://cogprints.org/4385/1/jisc2.pdf

Zimmermann Christian, (2009). "Academic rankings with RePEc". Working papers 2007-36, University of Connecticut, Department of Economics, revised Mar 2009, available at: http://www.econ.uconn.edu/working/2007-36r.pdf

Scientific Data: Increasing Transparency and Reducing the Grey

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

Abstract
The foundation for all scientific research begins with data, however most scientific datasets are not publicly available and are an increasingly important part of the body of scientific grey literature.
We provide an overview of the current scientific data landscape, primarily in the United States with regard to both policies and tactical approaches to better scientific data management and access. This includes how to improve bibliographic control (a metric for the definition of grey) as well as approaches to make datasets more usable. Since few scientific datasets are commercially produced, the historical framework for differentiating grey literature must adapt to new ways of making scientific datasets less grey.
To illustrate how the identification, collection, management and access to scientific data is increasingly transparent (less grey), we will examine a couple case studies from communities of practice including medicine and earth sciences. We will examine what has been done to make them less grey and more discoverable, and the policies that are driving the change. We will conclude with a look to the future – the policies and technologies that will facilitate additional progress.

Keywords: Scientific Data, Information Policies, Grey Literature, Data Discovery

Introduction

Science in the 21st century will be conducted in a fully digital world. The results of research and development are born digital and have digital life cycles. From numbers to text to images and audio, all knowledge is reduced to bits. New ways of accessing these bits and new concepts of returning investment on their generation had generated an increased emphasis on scientific datasets which are often the elementary particles of science. In today's world, collaborative science demands the sharing of data. Advances in computing technologies allow for the collection of and analysis of data on a previously unimagined scale. Researchers and scientists are finding new ways to use old data (re-purpose) in which the original creator may not have even considered, and the idea that datasets be made available for the benefit of all science has gained significant momentum in recent years. A dynamic push-pull relationship drives the need for scientific data. The technology that enables data creation is the "push" while demand for data intensive science is the "pull" (Figure 1).

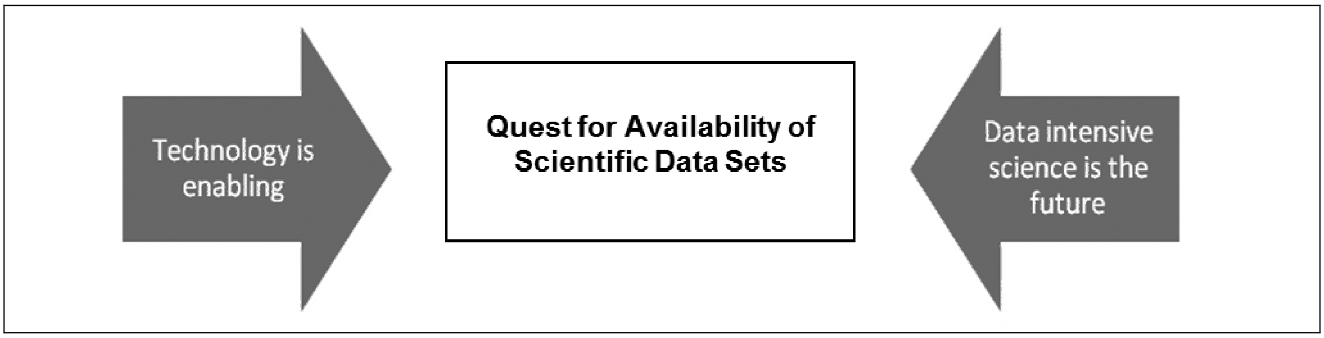

Technology is enabling
Quest for Availability of Scientific Data Sets
Data intensive science is the future

Figure 1. The drivers of scientific data**.**

More and more, researchers and policy makers alike adhere to the notion that "data are not consumed by the ideas and innovation they spark but are an endless fuel for creativity."[i] In many countries around the world, scientific data management (SDM) is being discussed at the strategic and policy level. The focus is to bring together the best practices to provide an overarching framework for SDM that includes the types of data and their expected impact; the relevant standards; and the provisions for protection, access, and continuing preservation.[ii]

Why are data grey?

The exponential growth of scientific research in relation to the multitude of distribution channels has helped create a situation that unsurprisingly became a challenge to professionals in both information and scientific communities of practice. In order to set the stage of datasets as grey literature and to look at how transparency is being increased, it is important to review a definition of the term grey literature. The most widely utilized definition today is the following:

Grey Literature -- "information produced on all levels of government, academics, business, and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body." (Luxembourg, 1997 - Expanded in New York, 2004)[iii]

Although not formally stated, the implicit nature of grey is also "obscure" or difficult to find and use. Essentially, a plethora of data "publishers," both traditional and new, have been operating without a uniform bibliographic control mechanism and without common methods to identify and access datasets that exist. This is a classic grey situation.

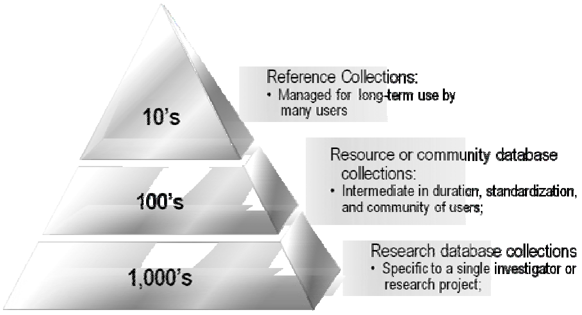

10's
100's
1,000's
Reference Collections:
• Managed for long-term use by many users
Resource or community database collections:
• Intermediate in duration, standardization, and community of users;
Research database collections
• Specific to a single investigator or research project;

Figure 2. The types of data collections.

Scientific Data Landscape

The Challenge (Data Complexity: Heterogeneity and Volume)

The combination of two interwoven factors, data volume and heterogeneity, have created a very complex data landscape. Much has been discussed in a variety of disciplines about the magnitude of data and information creation. The supply and demand of digital data have exponentially increased in recent years, and all areas of science (i.e., experimental, observational, theoretical, modeling) have been transformed by the continuous cycle of data generation, access, and storage of an ever increasing volume of digital data[4]. Data volumes in 2005, for example, were growing at a doubling rate, and datasets were forecasted to reach petabyte size in the near future[5]. As a result, scientists have predicted dramatic changes in the way science is and will be conducted, and there are speculations that "few traditional processes will survive in their current form by 2020[6]". The term exaflood has been applied to the challenges presented by massive amounts of data being generated. Within the scientific spectrum, there are many contributing examples to the exaflood. The Large Hadron Collider (LHC), for example, produces 15 petabytes annually. This avalanche of digitally stored data provides a foundational platform for data to be dissected and analyzed, shared, and combined in innovative ways to better quantify unique characteristics of matter properties[7].

The challenge to data management created by volume is compounded by the disparate forms and types of datasets that constitute the heterogeneous scientific data landscape. In the 2005 National Science Board report, "Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century" [8], datasets were characterized as (Figure 2):

- Research database collections that are specific to a single investigator or research project.
- Resource or community database collections that are intermediate in duration, standardization, and community of users
- Reference collections that are managed for long-term use by many users.

Each of these types requires very different management approaches and considerations but is rarely distinguished in discussion of "data." There is also a factor of heterogeneity across data types such as remotely sensed, field data, large and small scale experimental, and model data. Finally, each discipline has uniqueness in its data types and formats.

Publishers of Scientific Data

Part of the data landscape developed in efforts to increase transparency of datasets is a variety of venues in which scientific data are published. Currently, some of the primary sources that publish scientific data include:

1) Commercial publishers *(e.g. ProQuest)*;
2) Professional societies *(e.g., Ecological Society of America (ESA), Optical Society of America (OSA), International Council for Science: Committee on Data for Science and Technology (CODATA))*;
3) Repositories/clearinghouses/data archives *(e.g. Dryad, Sustainable Digital Data Preservation and Access Network (DataNet))*;
4) Information analysis centers *(e.g. Carbon Dioxide Information Analysis Center (CDIAC))*;
5) Research centers and researchers;
6) Metadata clearinghouses *(National Biological Information Infrastructure (NBII) Metadata Clearinghouse, Mercury Information Clearinghouse, Electronic Cultural Atlas Initiative (ECAI) Clearinghouse, Data.gov)*

Each of these has varying missions, goals, and reasons for existence. Traditional proprietary publishers (like ProQuest) have long competed in the marketplace by offering "value added" services designed to simplify the experiences of the end user. ProQuest offers a tool called "Deep Indexing" that allows users to search and retrieve information normally embedded in scholarly publications. This includes tables, figures, graphs, and illustrations within publications that have not previously been utilized separate of published material[9].

Professional societies have also historically published scholarly articles for the benefit of their paying members and affiliations. With the advent of newer technologies, many of the societies have embraced the open access concept, and/or participated in unique partnerships and collaborative agreements to increase transparency of scientific datasets. For example, the Ecological Society of America (ESA) established a data archive for publication material associated with ESA journals or partnering publishers. Three types of published content are archived for search and retrieval: appendices, supplements, and data papers. Additionally, authors are encouraged to officially register their published data to accompany archived content in efforts to strengthen the announcement of the existence of their data to a broader audience[10].

Repositories have long existed to house print format, and are equally essential as part of the digital landscape as well. These institutions may be clearinghouses or data archive centers, but serve essentially the same function. One significant recent initiative for digital data is Dryad, a collaborative partnership that originated in the research triangle of North Carolina, whose mission is to uphold an international repository that enables scientists to explore current and new methodologies of bioscience research, endorse published findings for intramural and extramural research initiatives, and potentially repurpose archived data for new applicable focus areas[11].

Another is National Science Foundation's (NSF) DataNet (Sustainable Digital Data Preservation and Access Network) Program that originated in 2007. To date, DataNet has funded two major projects that focus on integrating a plethora of diverse science areas (i.e., library and archival sciences, cyberinfrastructure, computer and information sciences) to enable and launch the development of long term sustainability of digital access, discovery, and preservation of science and engineering data. One of DataNet's additional initiatives is to accommodate the evolutionary pace of technologies and ensure reliable, long-term services to its audience[12].

The Information Analysis Center (IAC) is a concept proposed by the Weinberg Report[13] in 1963 that recommended the creation of centers of excellence where data were collected and analyzed to create new science from existing knowledge. This was a forerunner of today's data intensive science in a world without the digital technologies of today. There was a proliferation of IACs in the 1960s. On center that was a product of the movement and remains a leader in scientific data today is the Carbon Dioxide Information Analysis Center (CDIAC) [14] at the Oak Ridge National Laboratory (ORNL). In many cases, because of the highly technical focus of these organizations, access to scientific data from these sources has been fairly limited to interested researchers and employees affiliated with the Center. However, today's computing technologies (i.e., portals, embedded links) are aiding in increasing transparency above and beyond the traditional users of said data. Today, CDIAC has been a significant contributor to global climate change research and its data are easily found and accessed by scientists throughout the world.

Research centers and individual researchers have also historically created data but little was actually "published" or made available for reuse. In the internet world, more of these are in fact "published" or made available if you know how to find them and use them. This syndrome is classic grey literature and reminds us of the state of technical reports in the 1950s and 60s.

Metadata clearinghouses are a final source of published scientific datasets. Although Data.gov[15] compiles data sets from the United States government above and beyond the field of science, it is an example of a metadata clearinghouse for datasets and clearly illustrates the national policy level of commitment to better provide bibliographic and metadata control over as well as access to scientific data sets. Other examples that are more topical in nature include: the National Biological Information Infrastructure (NBII) Clearinghouse[16], and the Electronic Cultural Atlas Initiative (ECAI) Clearinghouse[17]. More specific to the scientific community is Mercury[18], and this enterprise is discussed in more detail later.

Key Thrusts that Reduce the Grey

It is clear that no single entity can solve the digital data dilemmas alone and results are often achieved through leveraging a unique set of perspectives and efforts by multiple stakeholders and Communities of Practice. Senior management is ultimately responsible for the "top down" approach by developing data policy and for data management planning. A recent workshop that addressed scientific data management (SDM) in the United States[19] brought federal agencies together to look at data policy and planning across the federal sector as an initiative that will help develop an effective top down approach. Professionals engaged in the everyday routines of manipulating scientific data represent the "bottoms up" approach and have emphasized data citation and digital object identifiers or other types of persistent identification of datasets as an essential element in their communities of practice. The information management and technology communities of practice have helped by developing improved data discovery tools and expanding capabilities in their quest for a "semantic web." One recent initiative undertaken by CODATA (the ICSU Committee on Data) was to form a Task Group on Data Citation Standards and Practices [20]. The work of this group should help in increasing the transparency of datasets by creating guidelines on how they should be identified and cited.

A look at Figure 3 below will illustrate a brief snapshot of the scientific data landscape with regards to policies, technologies, and efforts by stakeholders to increase transparency of scientific datasets.

Policy, Culture and Management	Technology Trends & Applications
National Policy – government taxpayers funded projects should be accessible	Digital object management technology
Enhanced metadata	Growth of scientific workflow software
Journals supporting links to some published datasets	Adaptation of "netcentric" way of doing business
"People getting the message that data has to be accessible."	Use of embedded links in publications
Increased involvement of libraries and lifecycle management of data	Increased number of portals serving datasets
Younger generation post data as they go – expectation that data should be shared	

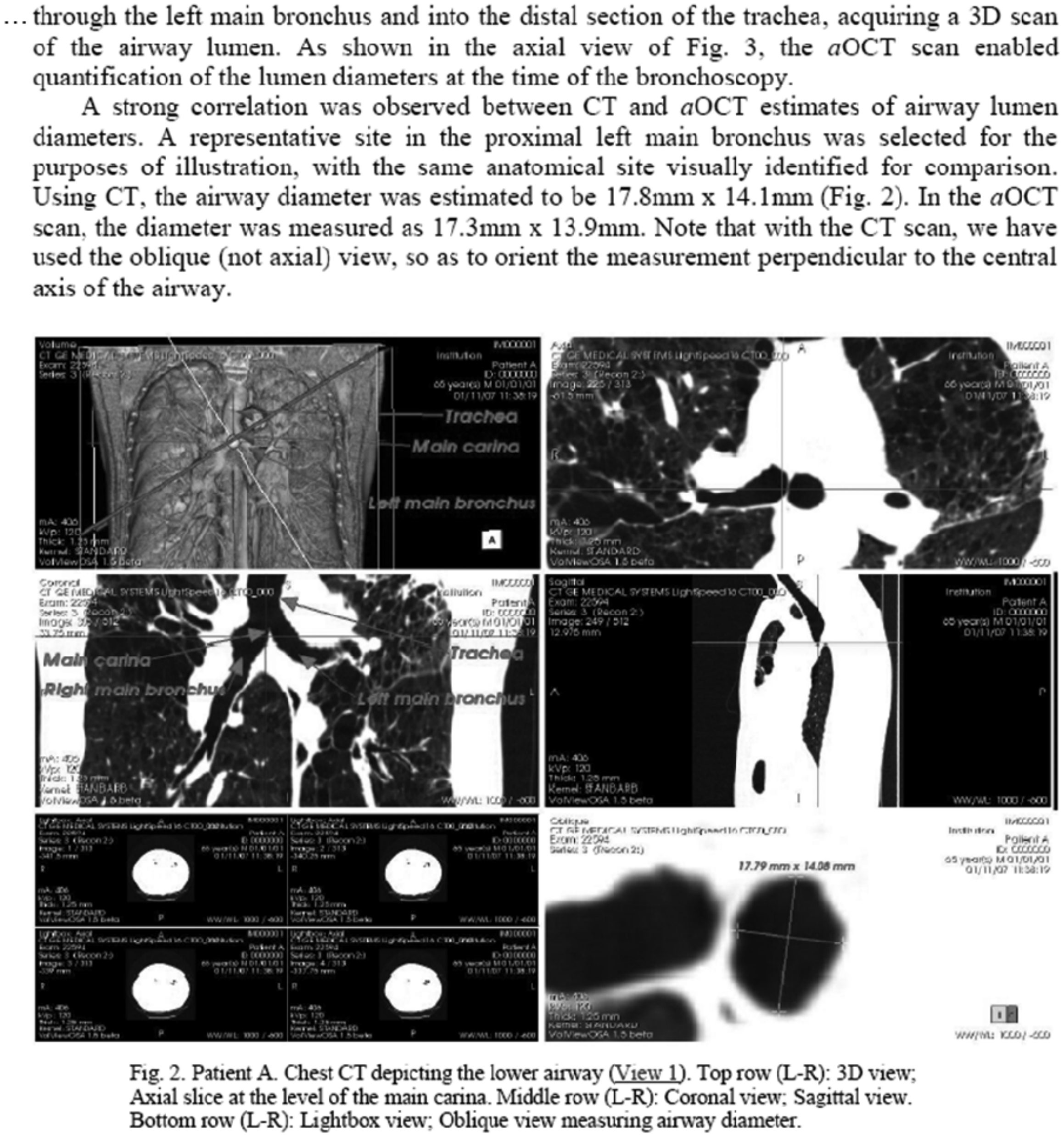
Figure 3. A high level view of the current scientific data landscape.

Data Activities in the United States/Case Studies

The following are a few case studies of initiatives that are working toward the better management of scientific data. With better management comes increased transparency.

Interactive Science Publishing (ISP) --- Medical

An innovative approach to scientific images and corresponding datasets has been implemented in a joint effort between the Optical Society of America (OSA) and the National Library of Medicine (NLM) entitled Interactive Science Publishing (ISP) [21]. OSA developed the ISP software in coordination with Kitware, Inc. and with support from the NLM and allows authors to publish large 2D and 3D datasets accessible through scientific articles. End-users can view and interact with original source data by downloading the ISP software. Ultimately, this software allows both readers and editors the ability to view, analyze, and interact with source data published in conjunction with an article. See figure 4 below for illustration.



… through the left main bronchus and into the distal section of the trachea, acquiring a 3D scan of the airway lumen. As shown in the axial view of Fig. 3, the *a*OCT scan enabled quantification of the lumen diameters at the time of the bronchoscopy.

A strong correlation was observed between CT and *a*OCT estimates of airway lumen diameters. A representative site in the proximal left main bronchus was selected for the purposes of illustration, with the same anatomical site visually identified for comparison. Using CT, the airway diameter was estimated to be 17.8mm x 14.1mm (Fig. 2). In the *a*OCT scan, the diameter was measured as 17.3mm x 13.9mm. Note that with the CT scan, we have used the oblique (not axial) view, so as to orient the measurement perpendicular to the central axis of the airway.

Fig. 2. Patient A. Chest CT depicting the lower airway (View 1). Top row (L-R): 3D view; Axial slice at the level of the main carina. Middle row (L-R): Coronal view; Sagittal view. Bottom row (L-R): Lightbox view; Oblique view measuring airway diameter.

Figure 4. An interactive 2D and 3D dataset derived from ISP software[21]**.**

Dryad's Repository --- Bio-diversity

Dryad is an international data repository created by the National Evolutionary Synthesis Center (collaborative effort by Duke, North Carolina, and North Carolina State Universities, and the National Science Foundation) and the University of North Carolina Metadata Research Center. The data underlies peer-reviewed articles in basic and applied biosciences and enables scientists to achieve multiple objectives, including validation of published findings, repurposing data for research in new and innovative ways unanticipated by the original authors, and performing synthetic studies, to name a few. Dryad is governed by a consortium of journals that promote data archiving and will ensure the sustainability of the repository. Figure 5 below provides an illustration of the general concept and architecture of Dryad[22]:

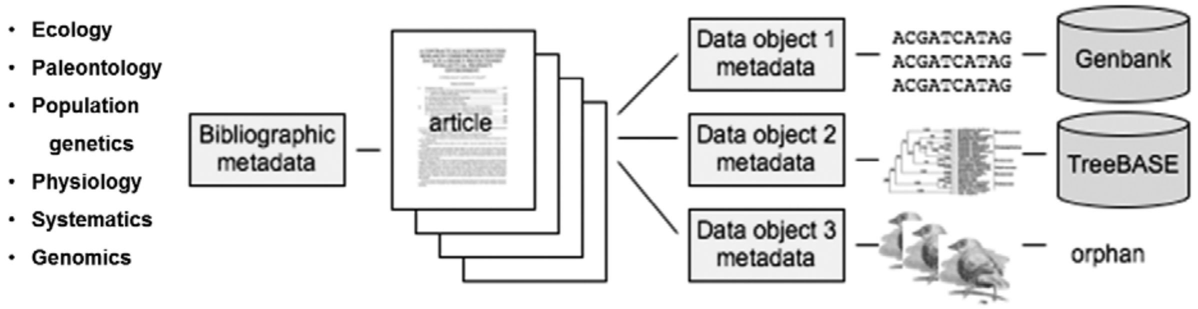


Figure 5: The Dryad repository conceptual flow diagram and general architecture**.**

As of Nov 23, 2010, Dryad contains 354 data packages and 868 data files, published in 50 journals.

Oak Ridge National Laboratory (ORNL) --- Earth Sciences

ORNL is a preeminent center for environmental scientific data management responsible for archiving, managing, and distributing data, and for enabling the distribution, use, and analysis of this data[23]. Three major data repositories (Atmospheric Radiation Measurement Archive, Carbon Dioxide Information and Analysis Center (CDIAC) [24], and the ORNL Distributed Active Archive Center[25]), comprise the central resources in environmental data. At the core of this enterprise is the Mercury Metadata Clearinghouse[26] architecture that harvests metadata from multiple nodes and compiles information in a Metadata Index. The end user is able to access the data through a single portal with numerous search capabilities, including RSS feeds and other web alert services.

Opportunities/Look to the Future

Collaborative efforts to address the complex issues of making research data more available have been a major focus in recent years, and results are undeniably being achieved as a result. Recent survey results of data management practitioners[27] (see Figure 6 below) highlighted in a Scientific Data Management (SDM) Workshop reveal that 20% of respondents claimed "yes" that they had sufficient knowledge of datasets when planning projects and research programs, while only 14.3% responded no to the same question. However, over 65% responded to the same question with "sometimes" identifying that considerable improvement can still be obtained[28]. The case studies addressed in this paper identified a trend in increasing transparency in their respective disciplines, and comments from STI managers indicated that "people are getting the message that data needs to be accessible" and that organizations have truly adopted a "netcentric" way of doing things. There will be an increasing sense that data belongs to the publication not just to the object. We will be looking at enriched publications that are supported by data and all tied together as bits in a web of science.

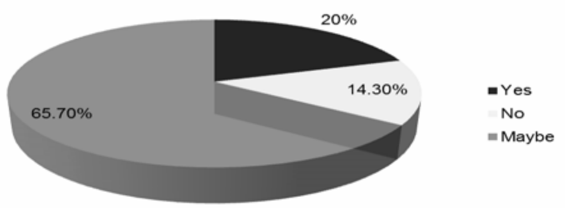


Figure 6: The percentage of survey responses that indicated sufficient knowledge of datasets when planning projects and research programs[29].

Conclusions

This thought paper briefly addressed the issue of scientific data as grey literature. It suggests that making scientific data more transparent and accessible is a major focal point of a lot of discussion at the national data policy level as well as from the communities of practice who are taking steps to gain better management control and provide better accessibility to scientific data.

In just the past few months since the paper was given at the Twelfth International Conference on Grey Literature (GL12) in Prague, Czech Republic[29], laws and Presidential directives in the US have focused on the need for better accessibility to scientific data[30]. In the UK a new Data Corporation has been proposed which will is a new but powerful "publisher" in the data landscape.

Although grey today under most definitions, the management and accessibility of scientific datasets are well on their way to more transparent access and control.

References

[1] National Science and Technology Council (NSTC), Office of Science and Technology Policy. Networking and Information Technology Research and Development (NITRD) Program. 2009. Harnessing the power of digital data for science and society. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. Available: http://www.nitrd.gov/about/harnessing_power_web.pdf [2] CENDI: Scientific Data Management Workshop, Washington DC, June 29-July 1, 2010. Accessed on October 15, 2010. Available: http://www.cendi.gov/activities/06_29_10_SDM_workshop_agenda.html [3]
Grey Literature definition from GreyNet International. Accessed on November, 15, 2010. Available: http://www.greynet.org

[4] Szalay, A., & Gray, J. 2006. Science in an exponential world. *Nature, (440),* 413-414.

[5] Gray, J. (2005). Scientific Data Management in the Coming Decade. SIGMOD Record, Vol. 34, No. 4, 34-41.

[6] Gray, J. (2005). Scientific Data Management in the Coming Decade. SIGMOD Record, Vol. 34, No. 4, 34-41.

[7] European Organization for Nuclear Research. The Large Hadron Collider. 2008. Accessed on January 19, 2011. Available: http://public.web.cern.ch/public/en/LHC/LHC-en.html

[8] National Science Board. National Science Foundation. 2005. Long-lived digital data collections: enabling research and education in the 21st century. Available: http://www.nsf.gov/nsb

[9] ProQuest. Cambridge Information Group. 2011. Accessed on January 18, 2011. Available: http://www.csa.com/e_products/pqdeepindex.php

[10] Ecological Society of America. Ecological Archives. 2011. Accessed on January 18, 2011. Available: http://esapubs.org/archive/

[11] Dryad Repository. 2010. Accessed on November 17, 2010. Available: http://www.dryad.org

[12] DataNet information. National Foundation. 2010. Accessed on November 17, 2010. Available: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141&org=OCI

[13] The Weinberg Report to the President's Science Advisory Committee (PSAC), 1963. Science, government and Information.

[14.] Carbon Dioxide Information Analysis Center (CDIAC). Oak Ridge National Laboratory (ORNL). 2011. Accessed on November 17, 2010. Available: http://cdiac.ornl.gov/

[15] Data.gov. 2011. Accessed on January 18, 2011. Available: http://data.gov

[16] National Biological Information Infrastructure Clearinghouse. Unites States Geological Society. 2011. Accessed on January 18, 2011. Available: http://metadata.nbii.gov/clearinghouse/

[17] Electronic Cultural Atlas Initiative (ECAI) Clearinghouse. The University of Sydney and Time Map™. 2010. Accessed on January 18, 2011. Available: http://ecaimaps.berkeley.edu/clearinghouse/

[18] Mercury Metadata Clearinghouse. Oak Ridge National Laboratory (ORNL). 2008. Accessed on January 18, 2011. Available: http://mercury.ornl.gov/

[19] CENDI: Scientific Data Management Workshop, Washington DC, June 29-July 1, 2010. Accessed on October 15, 2010. Available: http://www.cendi.gov/activities/06_29_10_SDM_workshop_agenda.html

[20] International Council for Science: Committee on Data for Science and Technology (CODATA). 2011. Accessed on January 18, 2011. Available: http://www.codata.org/index.html

[21] Interactive Science Publishing Optics InfoBase. The Optical Society. 2011. Accessed on January 19, 2011. Available: http://www.opticsinfobase.org/isp.cfm

[21] Childs, J. 2010. Interactive science publishing: a joint OSA-NLM project. Presented on January 12, 2010 at the National Library of Medicine (NLM). Available: http://www.cendi.gov/presentations/01-12-10_Ackerman_Michael_ISP.pdf

[22] Dryad Repository. 2010. Accessed on November 17, 2010. Available: http://www.dryad.org

[23] Oak Ridge National Laboratory (ORNL). 2011. Accessed on January 18, 2011. Available: http://ornl.gov

[24] Carbon Dioxide Information Analysis Center (CDIAC). Oak Ridge National Laboratory (ORNL). 2011. Accessed on November 17, 2010. Available: http://cdiac.ornl.gov/

[25] Distributed Active Archive Center for Biogeochemical Dynamics. Oak Ridge National Laboratory. 2010. Accessed on November 17, 2010. Available: http://daac.ornl.gov/

[26] Mercury Metadata Clearinghouse. Oak Ridge National Laboratory (ORNL). 2008. Accessed on January 18, 2011. Available: http://mercury.ornl.gov/

[27] CENDI: Scientific Data Management Workshop, Washington DC, June 29-July 1, 2010. Accessed on October 15, 2010. Available: http://www.cendi.gov/activities/06_29_10_SDM_workshop_agenda.html Survey results available: http://www.cendi.gov/presentations/06-29-10_IWGDD_SDM_Survey_Carroll.pdf

[28] CENDI: Scientific Data Management Workshop, Washington DC, June 29-July 1, 2010. Accessed on October 15, 2010. Available: http://www.cendi.gov/activities/06_29_10_SDM_workshop_agenda.html Survey results available: http://www.cendi.gov/presentations/06-29-10_IWGDD_SDM_Survey_Carroll.pdf

[29] Twelfth International Conference on Grey Literature (GL12), Prague, Czech Republic. December 6-7, 2010. Available: http://www.textrelease.com/gl12program.html

[30] PCAST Report, Report, Designing A Digital Future: Generally Funded Research and Development in Networking and Information Technology, Executive Office of the President, December 2010, Section 6.3 Large-Scale Data Management and Analysis. Accessed January 19, 2010. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nitrd-report-2010.pdf; America Competes Reauthorization Act of 2010, H.R.5116, Section 103(b). Accessed January 19, 2010. Available: http://democrats.science.house.gov/Media/file/Commdocs/H.R.%205116_bill_text.pdf; and the Scientific Integrity Memo from the Office of Management and Budget (OMB). Office of Science and Technology Policy (OSTP). December 17, 2010. Accessed January 19, 2010. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/scientific-integrity-memo-12172010.pdf

EBSCO
HOST

Invenio: A Modern Digital Library System for Grey Literature

Jérôme Caffaro and Samuele Kaplun,
European Organization for Nuclear Research, CERN, Switzerland

Abstract
Grey literature has historically played a key role for researchers in the field of High-Energy Physics (HEP). Consequently CERN (European Organization for Nuclear Research) as the world's largest particle physics laboratory has always been facing the challenge of distributing and archiving grey material. Invenio, an open-source repository software, has been developed as part of CERN's institutional repository strategy to answer these needs.

In this document we describe how the particular context of grey literature within the HEP community shaped the development of Invenio. We focus on the strategies that have been established in order to process grey material within the software and we analyse how it is used in a real production environment, the CERN Document Server (CDS).

1 Introduction

1.1 Background

CERN The European Organization for Nuclear Research in Geneva is the world's largest particle physics laboratory. Originally founded in 1954 by 12 European countries, CERN has established a solid reputation in scientific research throughout history. CERN is currently run by 20 European member states with over 40 additional participating observers (states and organizations). About 8,000 scientists from around 580 universities come to CERN to work on their research. The main current research program at CERN is the LHC (Large Hadron Collider), the largest (a 27km ring of superconducting magnets) and most powerful accelerator, smashing particles together to understand the basic constituent of matter. Over a thousand publications are published yearly by CERN scientists in established journals.

HEP Community The High Energy Physics (HEP) community is estimated to have about 20,000 scientists. It essentially comprises researchers working in the major particle physics laboratories around the world such as CERN, Desy (Germany) Fermilab (USA), SLAC (USA) and KEK (Japan).

1.2 Invenio

Invenio is an integrated digital library system [1] originally developed at CERN to run the CERN Document Server (CDS). It is currently one of the largest institutional repositories worldwide. It was started over 15 years ago and has matured through many release cycles. Invenio is a GPL2 Open Source project based on an Apache/WSGI+Python+MySQL architecture. Its modular design enables it to serve a wide variety of requirements, from a multimedia digital object repository, to a web journal, to a fully functional digital library. The development strategy used to implement Invenio ensures that it is flexible in every layer. Being based on open standards such as MARCXML and OAI-PMH 2.0 its interoperability with other digital libraries is guaranteed. Having been originally designed to cope with the CERN requirements for digital object management, Invenio is suitable for middle-to-large scale digital repositories (100K~10M records).

2 Grey Material at CERN

Invenio software was born in a rich grey literature producing environment [2]. One early major impact on the HEP community, and by consequence on the development of Invenio was the definition of a strong policy towards the dissemination of the work done at CERN. Indeed the convention that established CERN in 1954 states that "*[...] the results of its experimental and theoretical work shall be published or otherwise made generally available*" [3]. The implied openness of this mission forged a strong idea of responsibility for the community to give access not only to published documents, but also to additional material produced by the organization. CERN being an international organization, involving collaborations with a large number of universities and institutions, efficient sharing of information was a primary concern not only for scientific results, but also for all the material necessary for good coordination and proper running of the experiments: engineering drawings, technical reports, notes, etc. containing important scientific or technical data, but not suitable for publication in journals.

An important trend that took off among HEP researchers more than 50 years ago was the habit of mailing to their peers printed copies of their work at the time of submission to journals [4], reducing by several months the access to results of a possibly major importance in the context of the laboratory: machines and tools built for the CERN experiments being giant and technically advanced prototypes, they require several iterations of optimizations which can be performed through the analysis of early experimental

results. Reducing the duration of these cycles was necessary to help keeping the cost of the experiments as low as possible.
Early experimental results would also be used as input by theoretical physicists in order to refine existing theories or build new ones, and suggest new areas of study for experimental physics [5].

Another early concern expressed by the researchers at CERN was simply related to the physical access to grey literature. It is common for several thousands of scientists and engineers to work on the same experiment due to its complexity and size. A geographical distribution of the experts is inevitable, considering both the country of origin of these experts and location on the experiment(s) site(s) for practical reasons. Coordination of such large projects is only possible through formal, written forms. That particular need of giving access to a large amount of information in an electronic way resulted in the creation of the World Wide Web in 1989, as a proposal by Tim Berners Lee [6].

In the last two decades, HEP continued to pioneer solutions in scholarly communications. SPIRES, the SLAC (Stanford Linear Accelerator) literature database became in 1991 the first web server in the USA and the first online database in the world [7]. The same year arXiv.org (named at the time "LANL preprint archive") opened its web frontend, first as repository of physics preprints before expanding to other fields of science. In 1993 CERN released its preprints database on the web as an early version of the Invenio software, with the initial goal of fulfilling CERN's needs of access to grey literature.

3 Invenio for Grey Literature

In this section we review the strategies adopted in Invenio to support the management of grey literature. Examples of applications to real production systems running Invenio are given. In particular numbers given in the following paragraphs are updated statistics of the CERN Document Server (powered by Invenio) for November 2010.

MARCXML as core metadata format

MARCXML is the core bibliographic metadata format of Invenio. It offers all the required flexibility to model a great variety of digital assets.
Being a library standard, MARCXML offers the advantage of being well-known by professional librarians, giving a unique chance for grey material to be curated by the institutional library team. The very same tools used to manage published material can be used to process grey material, ensuring higher quality data and helping grey literature find its way more to standard institutional processes more easily.

Flexible metadata-formatting layer A flexible metadata formatting layer in combination with the MARCXML format allows the visualization of practically any digital asset within Invenio. An accessible HTML-like markup offers the opportunity for librarians to define the display of the managed records. Additional support for XSLT at the level of the formatting layer enables easy conversion from MARCXML to other XML flavor formats in a standard way.
CDS uses 122 different formatting templates, approximately half of them being used to prepare search results output, and the other half providing detailed information about the records. The templates deal with standard preprint objects, as well as video, audio or photo content. Combined with a customizable collection tree, it is possible to offer subject-based "portals" regardless of the actual type of contained material.

Customizable workflow engine

The submission system of Invenio lets administrators configure their own customized workflows. The framework offers the tools to create web front-ends for users to submit data (metadata and files), and an extensible set of functions to process the collected data.
Typical workflows result in the creation of a new record in the repository. Thanks to the mapping of the collected data to MARCXML, the flexibility offered by the submission system of Invenio regarding the type of supported data can be extended to the archival of this data. Submissions of Invenio can also be the starting point for subject specific jobs such as OCR (Optical Character Recognition) for scanned documents, and image downsizing for the creation of web versions.
91 different workflows are currently maintained on the CERN Document Server, some very similar to the basic workflow described above and other implementing complex reviewing and approval systems implicating thousands of different users. An average of more than 30 documents are submitted per day (less during week-ends) through these web-based workflows (300 documents per day when considering alternative input methods).

Collaborative tools

Invenio supports the creation of user groups (local or derived from the institution identity management system) and "baskets" letting users share information in a controlled, targeted manner. This feature is particularly useful in order to provide a community-based selection of unpublished documents at a quality that a ranking algorithm cannot match. Shared baskets can then offer a fast changing community-based hierarchical structure of data that Invenio main navigable collection tree cannot provide. In November 2010, CDS counted 6,245 non-empty baskets set up by 4,575 distinct users, covering about 10% of the whole archive. 6% of the baskets were shared among several users.

Invenio also features basic commenting and reviewing capabilities enabling a better understanding of the quality of the material. Consequently it also archives information about the documents which in the past was wrongly regarded as transient information only. The most active collection on CDS, regarding the number of comments, gets an average of about 30 comments per day through its built-in commenting system, usually on recent work being under peer-review at CERN.

Access control also plays an important role in helping the ingestion of grey material: providing an adequate, secure and restricted collaborative workspace for draft documents is an incentive for users to move their workflow to a central server, hence giving more opportunity for the final document to be made publicly (or not) available. Indeed Invenio can accompany documents through their life cycle, thanks to the integration of an advanced role-based access control system into the flexible workflow engine.

To bring a better awareness of the quality of the material to users and help with the discovery of documents of possible interest, Invenio display recommendation based on document usage statistics (*"People who viewed this page also viewed..."*) and features download/citation history graphs.

Search engine A major concern for large repositories of grey material is to provide an efficient way to retrieve new and archived material. Invenio includes a very fast search engine optimized for large repositories (millions of documents) on simple infrastructures, combining metadata and fulltext search in a simple Google-like query language. Advanced users are also given the opportunity to perform advanced queries, such as *find document written by Ellis from years 2000 to 2010, mentioning "higgs boson" in the fulltext, referring to documents written by Randall, and cited more than 50 times*. The CERN Document Server serves about 25,000 queries per day, for an archive of about 1 million records. Retrieved documents can be ranked according to several techniques, such as "word similarity" ranking or "citation-graph" [8] based ranking etc. in order to accommodate to the type of searched material: for example a researcher new to some subject is more likely to search for general reference documents while an expert might rather be looking for all new material in his field of interest.

The Invenio search engine technology is at the core of many functionalities offered by the software. For example combined with the flexible metadata-formatting layer, it can provide personalized search-based RSS feeds or email alerts: new results to some specific search queries such as the sample one mentioned above can be sent periodically to the subscribers. This has proven to be an essential functionality for CERN physicists in need of the latest information on some very specific topics. CDS counts more than 12,000 RSS subscriptions set up by 3000 distinct users (IP-based). 2417 emails alerts have also been set up by 1615 users.

Other use cases of the search engine include the suggestion of documents similar to a given one, or the creation of the bibliography (BibTEX) of a given author, personalized podcasts, etc.

Interoperability Invenio implements standard protocols to help the ingestion and dissemination of documents. Though these protocols are usually independent of document- type, they can still suffer from the conversion process usually occurring to ensure that repositories use a common language. For example OAI-PMH is able to support any document type, but most repositories only support the Dublin Core schema, hence narrowing down the possibility to use this protocol in some scenarios. Invenio is able to export and import any metadata format in OAI-PMH thanks to the underlying layers supporting custom conversion templates. For example OAI-PMH is used at CERN to feed an installation of Invenio with conference and meeting objects coming from the institutional conference management system Indico. Another example of usage of OAI-PMH in Invenio is in the context of the OPENAIRE project, which is planning to exchange usage statistics among participating repositories through this protocol.

Integrated digital library Invenio is a multi-purpose repository software: not exclusively designed for grey material, it offers the advantage of being a solution for the common needs of a library. It results in lower infrastructure maintenance costs by grouping several library services and processes on a single server. Reusing the same technology and concepts for these different services is also reducing the learning curve to master the necessary tools. It becomes much easier to justify the cost of supporting grey material within the institution.

4 Conclusions

Invenio is a well-established open source repository software. The context in which the software was conceived and then further developed has played an important role in defining a core set of features suitable for the ingestion, processing and distribution of grey material.

The performance and flexibility of the software has led to its adoption in a variety of scenarios, strengthening the will to drive the development efforts towards an increased support for grey material.

References

[1] Invenio official website.
http://invenio-software.org/
(Last visited on 19 November 2010)

[2] L. Goldschmidt-Clermont, Communication Patterns in High-Energy Physics; High Energy Physics Libraries Webzine, issue 6, March 2002.
http://library.web.cern.ch/library/Webzine/6/papers/1/
(Last visited on 18 November 2010)

[3] Convention for the Establishment of a European Organization for Nuclear Research, Paris, 1st July, 1953
http://council.web.cern.ch/council/en/Governance/Convention.html
(Last visited on 18 November 2010)

[4] Anne-Gentil Beccot et al., Information Resources in High-Energy Physics: Surveying the Present Landscape and Charting the Future Course, J.Am.Soc.Inf.Sci.60, 150-160, (2009)

[5] Anne-Gentil Beccot, How do High Energy Physics scholars search their information?, Grey J. 4, 1 (2008).

[6] Tim Berners-Lee Information Management: A Proposal CERN-DD-89-001-OC, 1989
http://cdsweb.cern.ch/record/369245

[7] L. Addis, Brief and Biased History of Preprint and Database Activities at the SLAC Library, 1962-1994.
http://www.slac.stanford.edu/spires/papers/history.html
(Last visited on 19 November 2010)

[8] Ludmila Marian et al., Citation graph based ranking in Invenio. LNCS (Research and Advanced Technology for Digital Libraries) 6273 (2010)

GL Transparency: Through a Glass[1] Clearly[2]

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

GL (Grey literature, interpreted here as grey objects) is very heterogeneous in content, form and quality. Most GL objects evolve through a workflow. Some of these phases involve some form of evaluation or peer review, commonly internal within the management structure of an organisation and possibly involving external advice, including from 'friendly peers' via an e-preprint mechanism. Unlike white literature the evaluation process commonly is unrecorded and undocumented. This leads to accusations that grey literature lacks quality and transparency. This paper proposes how the GL community can overcome this – generally unfounded – accusation, building on our previous work.

A GL repository records the intellectual property of that organisation (2004). We have demonstrated that effective use of this resource requires that the metadata is formalised (1999, 2004) – more precisely in a CERIF-CRIS (Common European Research Information Format – Current Research Information System) (2005). The GL is then available in the context of the work of the organisation and/or its stakeholders managing strategy, evaluation, funding and cost-accounting, innovation and knowledge transfer and public information (2005). This provides user-evaluated assurance on the quality and relevance of the grey object. CERIF provides temporally-based relationships between grey objects (and white objects) thus recording evolution of the object during the workflow – hence provenance. This concept was further refined as 'Greyscape' (2007) and the technologies for interoperation – in order to provide the underpinning homogeneous access to the heterogeneous repositories - surveyed (2008). Efficiency of using CERIF was outlined in (2009). Using advanced hyperactive objects (2006) is postponed until the requirement is realised by the community.

CERIF-CRIS provides the capability for greater quality and transparency through novel methods of evaluating quality, provenance and review including Web2.0 recommender-type systems as well as conventional review mechanisms. CERIF-CRIS provides the way to overcome criticism of GL.

The key messages are:

1. formal metadata associated with grey literature repositories improves relevance and quality;
2. transparency requires recording the workflow phases of a grey object within the context of a research information system;
3. a solution – CERIF – exists already which covers these requirements.

Background

Previous Work

For more than two decades, the authors have worked on research information in the widest sense comprising information not only about grey literature (grey objects) but also all the outputs of research (products, patents, publications) and the context within which the research was done including projects, organizations, funding, persons, facilities, equipment, events. Within the GL community we have highlighted the issues as we see them:

1. the need for formal metadata to allow machine understanding and therefore scalable operations (Jeffery 1999);
2. the enhancement of repositories of grey (and other) e-publications by linking with CRIS (Current Research Information Systems) (Jeffery and Asserson 2004);
3. the use of the research process to collect metadata incrementally reducing the threshold barrier for end-users and improving quality in an ambient GRIDs environment (Jeffery and Asserson 2005);
4. an architectural model for scaleable, highly distributed, workflowed repositories of grey literature based on hyperactive 'intelligent' documents (Jeffery and Asserson 2006).
5. A 'from 10,000 metres altitude' view of the grey information landscape 'Greyscape' based on the hypothesis that grey literature is the foundation for the knowledge economy (Jeffery and Asserson 2007).
6. An analysis of interoperation architectures among research information systems 'INTEREST' (Jeffery and Asserson 2008).

[1] Grey Literature Architecture for Sustainable Systems

[2] "For now we see through a glass, darkly". The Bible: 1 Corinthians xiii, 12

7. A proposal that Grey Literature should be seen within the context of e-Science supported by a CERIF-CRIS (Jeffery and Asserson 2009).

The Requirement

Our work has convinced us of the need in the Grey literature community for two key technologies:

a) Metadata with formal syntax and declared semantics to allow reliable, scalable management and interoperation of grey resources;
b) Workflow within a research process context to minimize effort for the researcher, research manager, librarian or other knowledge worker, to record the provenance of a grey object and thus to increase accuracy, relevance and contextual awareness of the research information;

These technologies are required for many purposes in GL including – but not limited to – transparency. Transparency is defined in physics as the property of allowing light to pass through a material while more generally it implies openness, communication, and accountability. The latter meaning is used in this paper.

We contend that the currently widely-accepted metadata standards for GL – namely Dublin Core (DC) and (MARC) – are insufficient for the purposes of:

a) Discovery (relying on multilingual semantics over multicharactersets);
b) Management (utilising especially the formal syntax);
c) Utilisation (including security and privacy);
d) Understanding (relying on semantics);
e) Re-purposing (relying on both syntax and semantics);
f) Contextualising (utilising contextual metadata such as project, organisation);
g) Provenance (the stages through which the material has been);
h) Preservation/curation (for later re-use by future researchers);
i) Quality assessment (utilising the recorded workflow steps (provenance));

Without appropriate metadata transparency is lost (and the impact of the work recorded in the GL object is much reduced).

Further, we contend that unless GL material is collected in the context of a research workflow of services acting on the grey objects:

a) the threshold barrier to collection is high and discourages those producing the GL from providing the metadata (or even the source material);
b) associated contextual information is lost including any quality controls or peer review, or information allowing reputational judgement – thus transparency, so essential for confidence and trust in the information, is also lost;

We propose that both of these problems are overcome by use of a CERIF-CRIS.

The Hypothesis

The hypothesis is in three assertions:

1. formal metadata associated with grey literature repositories improves relevance and quality;
2. recording the workflow phases of a grey object within the context of a research information system provides provenance;
3. a solution – CERIF – exists already which covers these requirements.

Utilisation of this technology provides a GLASS (Grey Literature Architecture for Sustainable Systems) enabling GL users previously "seeing through a glass darkly" to see clearly.

Proposed Architecture

Introduction

The proposed GLASS – to achieve all the required aspects of a GL environment including transparency - consists of grey objects, metadata and services operating over a virtualised e-infrastructure based on GRIDs (as proposed in (Jeffery and Asserson 2009) or CLOUD technology (for a survey and analysis see (Schubert, Jeffery, Neidecker-Lutz 2010)) thus in the same domain as that in which researchers do their other work. In this way activities associated with GL are not divorced from observation, experimentation, simulation or project management.

Grey Objects

It is expected that the grey objects will be heterogeneous (either in the local collection or the virtual collection obtained by accessing across heterogeneous distributed GL repositories) and of various (multi)media types. The only architectural problems concerning the objects are to ensure that appropriate services are available to utilise them (see list of functions in section 1.2). This implies rich metadata related to the objects to characterise the way in which they are utilised.

Metadata for Grey Objects

There are several classifications of metadata and multiple standards across many domains of scholarly research. The classification of metadata should relate to the purposes for which it will be utilised (through services available to the user) related to the object. For example schema metadata is used to assure integrity whereas descriptive metadata is used – among other purposes – for discovery.

Table 1: Metadata Kinds for Grey Objects Related to Services

SERVICE	METADATA	COMMENTS
Discovery	Descriptive	Multicharacterset, multilingual
Management	Schema Descriptive Restrictive Navigational Provenance Curation/Preservation	Depending on the management process different kinds of metadata are utilised
Utilisation	Schema Descriptive Restrictive Navigational	The schema metadata connects the object to the service assuring integrity, the descriptive metadata assures relevance and the restrictive metadata assures rights, security, privacy compliance
Understanding	Descriptive Provenance	The descriptive metadata assures relevance. The provenance metadata illuminates the evolution of the object
Re-Purposing	Schema Descriptive Restrictive Navigational Contextual Provenance Curation/Preservation	In order to re-use a grey object as much information about it as possible is required to assure that the re-use is valid.
Contextualising	Contextual	For example placing the object in the context of a research project, or related to a research facility
Provenance	Provenance	One aspect of quality
Preserving / Curating	Schema Descriptive Restrictive Navigational Contextual Provenance Curation/Preservation	All metadata is needed to allow re-purposing at a later time when the grey object creator may be unavailable
Quality Assessment	Schema Contextual Provenance	The schema metadata provides integrity, provenance metadata describes the object evolution and contextual metadata covers the research context

Clearly there are advantages if the metadata for grey objects is stored within one standard structural environment. CERIF provides such an environment covering all the kinds of metadata outlined above, except schema which – by definition – relates to the conceptual, logical and physical representation of the object within the hosting environment. The availability of this rich metadata for grey data objects assures transparency.

Services

Services are executed to fulfil the requirements of the end-user. Services, themselves, require metadata in the same way as grey objects. Services need to be discovered, managed etc. Services can be :

(a) object-independent i.e. generic processes that act on any data or
(b) object-dependent i.e. including and enclosing the object(s) together with the processes.

Examples of (a) are the relational algebra operators (select, project, difference, union, join) which then act on any relational table(s) whatever the data stored in those tables. Examples of (b) include currency conversion services where the current exchange rate table is incorporated within the service and its associated atomic processes. There is an argument that composed services should themselves be described by metadata and contain both processes and objects, each of which is also described by metadata.

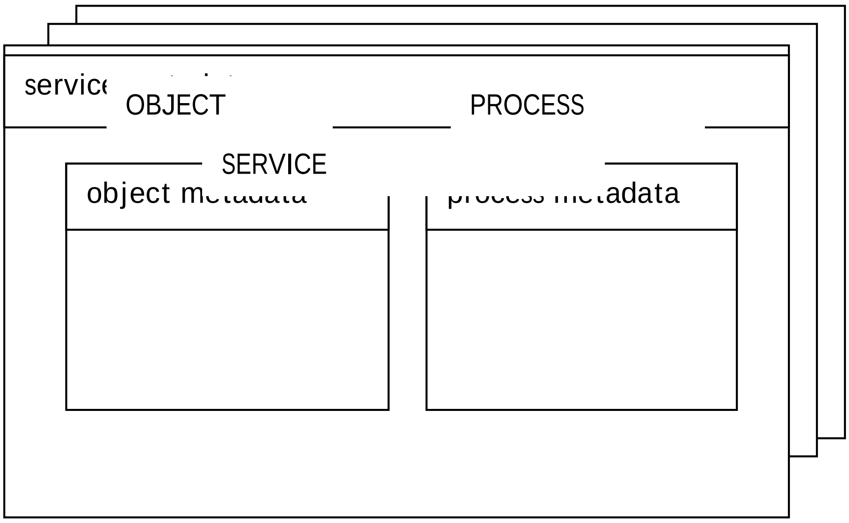


Figure 1: Service, Object, Process Metadata

However, whereas objects can be collected together in collections (usually as a set of similar objects based on some parameters e.g. all grey objects generated in 2010 relating to global warming and climate change) services also may be composed – that is groups of atomic services are linked together in some ordered fashion to execute the requirements of the end-user. A simple sequence of services (e.g. select, count) is a simple workflow. A more complex workflow has decision points and branches. However, for reasons of performance and resilience – especially in a heterogeneous distributed environment – the composition can include multiple parallel tracks of workflow with replicated services. It is necessary that each service can be executed anywhere – depending on requirements of performance, security etc – which demands that the services are mobile – that is the program code can be moved to the locus of execution. This leads to the requirement for self-organising (composing, managing, optimizing) services because the complexity of matching the requirement (including non-functional aspects such as performance, security, adherence to a service level agreement or quality of service) to the execution environment (distributed, heterogenous, parallel, multi-tenanted) is too great and too dynamic for human management.

Within the architecture task group of euroCRIS, a set of services for any CERIF-CRIS is being discussed.

Metadata for Services

The services provided require metadata in order for them to be utilised and – more importantly – for them to be utilised correctly.

1. Schema metadata: analogous to the schema controlling integrity in a data object is used to assure integrity in the service particularly in the parameters and input/output declarations;
2. Navigational metadata: analogous to that for data objects is used to locate the service:
3. Descriptive metadata: analogous to that for data objects is used to discover the service and then (together with the schema and restrictive metadata) to assure fitness for purpose;
4. Restrictive metadata: analogous to that for data objects assures enforcement fo non-functional properties of the service such as performance, security, privacy, rights management, (micro-)payment for usage;
5. Provenance metadata: analogous to that for data objects records the transition states of the service as it evolves;
6. Curation/Preservation metadata: analogous to that for data objects records the additional information require to assure (a) preservation of the service code and documentation (specification) over time e.g. through media conversion and evolution; (b) curation such that the purpose and characteristics may be understood in future time;

7. Contextual metadata: analogous to that for data objects describing how the service fits within a context of other research information such as projects, organizations, facilities, equipment etc. CERIF is the preferred standard for use here.

The availability of this rich metadata for services in the GL domain assures transparency.

Conclusion

The proposed GLASS architecture achieves transparency through several mechanisms:

1. encouraging the provision of full metadata using CERIF to cover all aspects of the grey data object thus maximizing the potential utilisation and providing information relating to integrity and quality;
2. encouraging the provision of full metadata using CERIF to cover all aspects of services thus maximizing the potential utilization (including in composed services) and providing information relating to integrity and quality;
3. through CERIF defining metadata with formal syntax (for reliable computer processing) and declared semantics (for computer or human understanding);
4. through CERIF providing a data model which records the date/time interval associated with any relationship between two base entities. This provides automatically a provenance track and also can be used for non-functional aspects such as security, privacy, rights restrictions;

References

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

(Asserson and Jeffery 2005) Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' The Grey Journal volume 1 number 1: Spring 2005 TextRelease/Greynet ISSN 1574-1796 pp5-8

(CERIF) www.eurocris.org/cerif

(DC) http://www.oclc.org:5046/research/dublin_core/

(DSpace) DSpace repository product homepage http://www.dspace.org/

(ePubs) ePubs repository product homepage http://epubs.cclrc.ac.uk/

(eprints) ePrints repository product homepage http://www.eprints.org/

(Fedora) Fedora repository product homepage http://www.fedora.info/

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

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

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

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

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

(Jeffery 2005) K G Jeffery CRISs, Architectures and CERIF CCLRC-RAL Technical Report RAL-TR-2005-003 (2005)

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

(Jeffery and Asserson 2005) K G Jeffery, A G S Asserson 'Grey in the R&D Process'; Proceedings Grey Literature 7 Conference, Nancy, December 2005; TextRelease; ISBN 90-77484-06-X ISSN 1386-2316

(Jeffery and Asserson 2006a) Keith G Jeffery, Anne Asserson: 'CRIS Central Relating Information System' in Anne Gams Steine Asserson, Eduard J Simons (Eds) 'Enabling Interaction and Quality: Beyond he Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp109-120 Leuven University Press ISBN 978 90 5867 536 1

(Jeffery and Asserson 2006b) Keith G Jeffery, Anne Asserson: 'Supporting the Research Process with a CRIS' in Anne Gams Steine Asserson, Eduard J Simons (Eds) 'Enabling Interaction and Quality: Beyond the Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp 121-130 Leuven University Press ISBN 978 90 5867 536 1

(Jeffery and Asserson 2006c) Keith G Jeffery, Anne Asserson: 'Hyperactive Grey Objects' Proceedings Grey Literature 8 Conference (GL8), New Orleans, December 2006; TextRelease; ISBN 90-77484-08-6. ISSN 1386-2316 ; No. 8-06-X

(Jeffery and Asserson 2007) Keith G Jeffery, Anne Asserson: 'Greyscape' Opening Paper in Proceedings Grey Literature 9 Conference Antwerp (GL9) 10-11 December 2007 pp9-14; Textrelease, Amsterdam; ISSN 1386-2316

(Jeffery and Asserson 2008) Keith G Jeffery, Anne Asserson: 'INTEREST' Proceedings Grey Literature Conference Amsterdam 8-9 December 2008 in Tenth International Conference on Grey Literature : Designing the Grey Grid for Information Society, 8-9 December 2008, Science Park Amsterdam, The Netherlands ed. by Dominic J. Farace and Jerry Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, February 2009. GL-Conference series, ISSN 1386-2316; No. 10. - ISBN 978-90-77484-11-1.

(MARC) http:// http://www.loc.gov/marc/

(PURE) http:// http://www.atira.dk/en/pure/

(RDF) http://www.w3.org/RDF/

Schubert, Jeffery, Neidecker-Lutz 2010 'The Future of Cloud Computing' http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf .

(XML) http://www.w3.org/XML/

(Z39.50) http:// http://www.loc.gov/z3950/agency/

Grey Literature in Engineering Sciences and Technology and its Use Pattern in the Research Institutions in India: The Case Study of Karnataka State

N. Chowdappa, BMS College of Engineering, and
C.P. Ramasesh, University of Mysore, India

Abstract
The paper is based on the survey research undertaken during 2008-2010. The availability of GL collections and its use pattern in the Recognised Research Centres of Visvesvaraya Technological University (VTU) have been analysed on the basis of the responses of the librarians and the personal observation of the situation in the libraries. A detailed attempt is also made to capture the responses of the users on the use pattern of GL. The users constitute the faculty members engaged in supervising the research programmes, and the research scholars who have registered for doctoral research. The extent of the use of the 6 types of conventional form of GL: theses, dissertations, institutional publications, trade literature, technical reports and proceedings of conferences, seminars and workshops have been studied.

Further, the research covers aspects involved in building need based GL collections. Habits of the faculty in visualizing the recent releases of GL and the trends in exploiting the GL collections of the local and regional libraries have been depicted. The research suggests for the creation of databases of GL, resource sharing and development of digital repositories of theses and proceedings of conferences and seminars, for which increased demand is noticed from the user community. The study also captures the training needs of the working librarians and the user community for better access and use of GL .

Keywords: Use Pattern; Resource Sharing; Engineering Sciences; Orientation Programme

1. Introduction: The grey literature has a great significance in supporting ongoing academic and research initiatives especially, in the disciplines of engineering sciences and technology. It is produced mainly by researchers and practitioners, in fact more quickly, more detailed with greater flexibility than white literature. Grey literature includes standards, specifications, technical notes, conference proceedings, patents, trade literature, work manuals, work flow-charts, institutional reports, plans and lay-outs and the term "Grey" in the professional environment is "unconventional" "informal" "informally published" "fugitive" and even "invisible". Auger (1998, p.3) includes in the purview, reports, technical notes, theses, trade literature, preprints etc. In his definition all of which are characterized by: poor bibliographic control and information; low print runs and a non-professional layout and format; difficult to obtain, as they are not available through normal book selling channels.

2. Scope of the Study: The literature study reveals that till now no study has been done on grey literature available in engineering colleges in India. The way they are organized to facilitate easy access, and to what extent users are using these resources, though significant, are yet to be understood. The study also unfolds the use pattern of GL which will help to find new approaches for facilitating better access to the grey materials. The study focuses on the six categories of printed grey literature available in the holdings of these Recognised Research Centres in the State of Karnataka : theses, dissertations, institutional publications, trade literature, technical reports and proceedings of conferences, seminars and workshops.

Theses: Researchers and scholars in the academic institutions and universities take up research programmes and finally come out with the research reports / theses for the award of doctoral degrees. These reports constitute unpublished sources of primary information and become property of the universities. Theses are considered to be vital as they carry in depth studies indicating objectives, methodology, hypotheses, and findings of research encompassing up-to-date statistical data, citations and information pertaining to the narrow field of study. Theses are valuable because they are original studies, and subject to strict review.

Dissertations: By and large, master's degree students take-up minor research projects as part fulfillment of their degree programmes and submit the dissertations to the academic institutions. Studies in dissertations cover small area, involving limited population samples, and the research work gets completed within the stipulated semester period. Dissertations cover all issues such as methodology, scope and findings of research, including up-to-date reading list, usually useful in case of similar project works.

Institution Publications: Every institution, whether academic or research brings out number of in-house publications for internal circulation and reference. Annual reports, budget reports, academic progress reports, quality assessment reports, audit reports, reports covering academic and research programmes, including extension programmes and publications covering infrastructure development, human resource development, and training facility fall into this category. The institution publications form vital source of information giving details on the background/historical development of the institution. Usually, scholars working in the same institution access them for facts and figures.

Trade literature: Manufacturing and service sectors bring-out regularly the publications in the form of book-lets, catalogues, brochures and guides. They cover various products and services of industries, laboratories and manufacturers giving specific details pertaining to models, components, prices, warranty, mode of delivery and so forth. Users refer to these literature to know the latest products, composition, models and prices.

Technical Reports: Industrial and research institutions regularly bring-out report literature which carry experimental procedure, production data, operating plans, flow-charts, drawings, mainly dealing with technological specifications useful for engineers and technologists. By and large, research organizations prepare technical reports as their outcome of research activities, addressing a particular problem. Technical reports are usually very rich in experimental detail according to the MAGiC Project (Needham, 2002, p.20). Sometimes, institutions apply colour coding for easy identification and circulation and immediate reference. Technical reports are the most valuable source of grey literature useful for research and development in any given field.

Proceedings of Conferences, Seminars and Workshops: Academicians and scientists belonging to a particular discipline come together on a common platform to discuss and deliberate on a particular subject and come-out with recommendations. Before/after the conferences or seminars usually the professional bodies/societies/associations bring out the proceedings. The proceedings carry research articles covering up-to-date data and information, contributed by many scholars and hence, form very useful tool to know the current trends and developments (Ramadevi, 2006,pp.53-54). Proceedings of the workshops gives practical details and step-by-step guidelines usually circulated among the participants before the conduct of practical orientation or training sessions, emphasizing mainly on hands-on experience.

3. Objectives :

- To know the availability of grey literature in the engineering institutions.
- To assess the use pattern of grey literature by the faculty supervisors and research scholars and further, to suggest methods to be adopted for facilitating easy access.
- To know the extent of sharing of the GL collections of regional libraries.To comprehend the training requirements of the librarians and users for better access and use of grey literature

4. Methodology : The study began with a detailed and extensive literature search. Considering the availability of the infrastructure facility and the qualified faculty, 65 engineering colleges have been identified for research programmes out of 171 engineering colleges. The primary data relating to the availability and organization of grey literature in the libraries of the Recognised Research Centres have been collected through questionnaire from the librarians. The researcher also visited all these centres and observed the grey collection and its organisation. After having received the filled in questionnaires, the data was tabulated and analysed. The second questionnaire was designed and administered to all the users, the faculty members who are engaged in supervising the doctoral programmes and the research scholars who have registered for Ph.D programme. The data thus obtained through the questionnaires have been analysed in the light of the objectives. The entire population covered in the study is 930 research scholars and 340 faculty guides. The percentage of total users covered in the study is 84.6.

In India, the All India Council for Technical Education (AICTE) has taken initiative in providing access to scholarly publications, including the digital information resources to the engineering colleges and research institutions. The Indian National Digital Library in Engineering Sciences and Technology (INDEST), popularly known as INDEST-AICTE Consortium, in association with the University Grants Commission (UGC) is entrusted with the task of meeting the information requirements of the research community. In the State of Karnataka, there are 171 engineering colleges. Out of which, excluding industrial establishments, 65 colleges are recognized as research centres.

5. Collection of GL and its Use Pattern.

Below table furnishes statistics of GL collection. As submission of dissertations is a part of master’s degree programme and mandatory for the post-graduate students, the total holding of dissertations in the libraries is 79845. Many libraries have got a good collection of dissertations, proceedings of CSW and technical reports.

Table 1: Total collection of GL in the libraries

Sl.No.	Types of GL	Total Collection
1	Theses	793
2	Dissertations	79845
3	Institutional publications	9499
4	Trade literature	3021
5	Technical reports	22181
6	Proceedings of CSW	16456
	Total	**131795**

The trade literature collection is quite small. As the research programme has been initiated recently by VTU, there are just 793 theses volumes in the holdings of these libraries. Many of the theses volumes before 2003 were submitted to various universities of the state. Further, some of the libraries do not have complete collection of their own college. Engineering colleges have to make it mandatory to send one or more copies of their publications to their libraries so that comprehensive collection can be accessed in the libraries without any gaps. College publications include annual reports, budget reports, progress reports, calendar of events, development review reports, quality monitoring reports, details of facilities and services, college handbooks, guides, brochures and manuals. Acquisition policy, collection development policy in the engineering colleges has to be evolved with action plan to build-up GL collection. A survey of users has to be conducted on the use of GL to identify the area and specific documents required for various academic and research programmes.

It is clear from the below table that the GL in all the govt. College libraries is easily accessible without any difficulty. In the Aided college, 88.9% of the libraries have kept GL at vantage points facilitating easy access.

Table 2: GL kept at vantage points and easily accessible

Sl. No.	Extent of Easy Access	Govt. Colleges		Aided Colleges		Private Colleges		Total	
1	Most easily accessible (100%)	5	83.3%	3	33.3%	26	52.0%	34	52.3
2	Easily accessible (75%)	1	16.7%	5	55.6%	18	36.0%	24	36.92
3	Moderately easy to access (50%)	Nil	Nil	1	11.1%	5	10.0%	6	9.2
4	Difficult to access (25%)	Nil	Nil	Nil	Nil	1	2.0%	1	1.5
	Total	**6**	**100**	**9**	**100**	**50**	**100**	**65**	**100**

As regards private college libraries, 88% of the libraries facilitate easy access to the GL. Considering the 65 colleges recognized for research, a large majority of the libraries representing 89.2% have kept GL at vantage points so that the users can access easily

Table 3 : Extent of dependency of librarians on other libraries 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 that the dependency of librarians on regional libraries is to a lesser extent in the range of 25% to 50%. To a moderate extent of 50%, the librarians depend on other library collection especially for technical reports and the proceedings of CSW. Not much of dependency on other libraries can be noticed as far as institution publications and theses volumes are concerned. From this, it can be deduced that a large percentage of librarians depend on regional libraries to the lesser extent of 25% to 50%. The librarians need to be educated to make use of the resources and services of regional libraries too for meeting the needs of the users beyond expectation, and thereby creating the 'delight' among the library users. This has to be treated as the quality library services.

Table 4: User- wise frequency of access and use of GL

Sl. No.	Frequency	Category of Users					
		Faculty guides	Percentage	Research scholars	Percentage	Total	Percentage
1	Most frequently	20	5.9	67	7.2	87	6.8
2	Frequently	189	55.6	519	55.8	708	55.8
3	Moderately	98	28.8	279	30.0	377	29.7
4	Occasionally	31	9.1	64	6.9	95	7.5
5	Not at all	2	6	1	0.1	3	0.2
	Total	**340**	**100**	**930**	**100**	**1270**	**100**

A large segment of users, 209 faculty guides and 586 research scholars representing 61.5% and 63% respectively, access and use GL frequently or most frequently for academic and research needs. The negligible number of users, 2 faculty guides and 1 researcher have offered negative opinion. From among the entire population of users, it is clearly evident that 62.6% of them access GL quite often. Hence, it is true that the GL collection available in the holdings of the engineering college libraries is of great value for the faculty and research scholars.

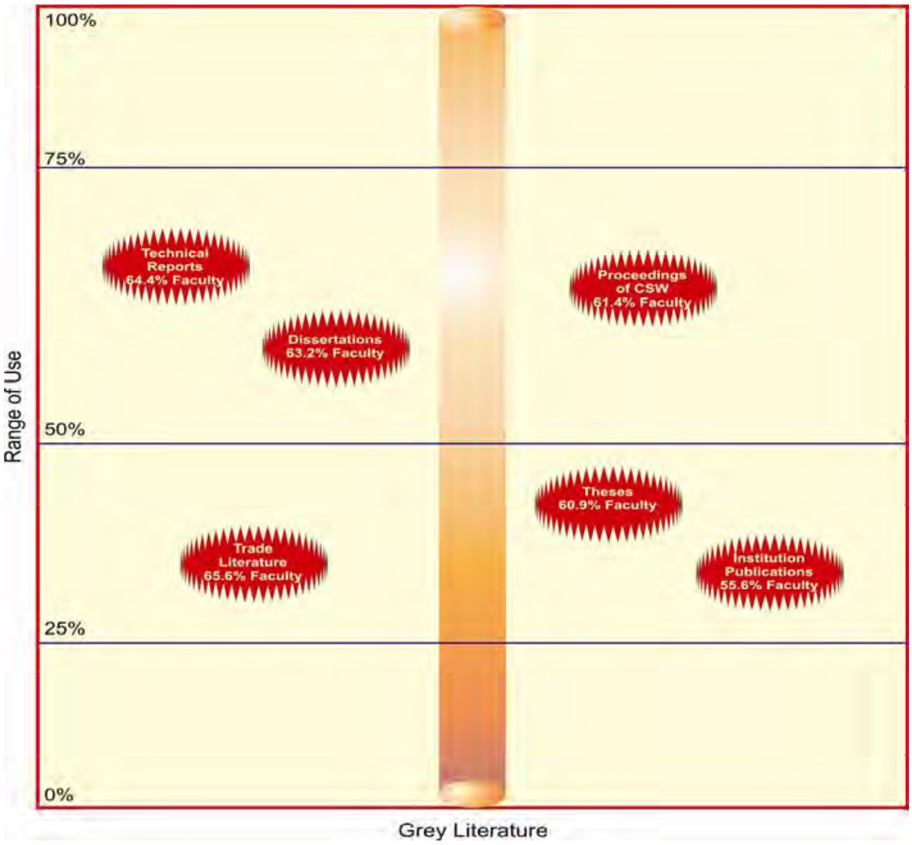
Table 5: Response on the method of locating the GL in the libraries

Sl. No.	Means of Locating GL	Faculty Guides		Research scholars		Total	
		No.	%	No	%	No	%
1	Use of catalogues	128	37.6	331	35.6	459	36.1
2	Use of new arrival lists	97	28.5	250	26.9	347	27.3
3	Web notices/alerts	17	5.0	50	5.4	67	5.3
4	Browsing at the shelf	98	28.8	299	32.2	397	31.3
	Total	**340**	**100**	**930**	**100**	**1270**	**100**

It is clear that 459 respondents representing 36.1% use the library catalogs. Further, it is also important to note that 347 respondents comprising 27.3% use new arrivals list compiled by the libraries. Another segment of respondents representing just 5.3% depends upon web alerts, web notices. It is worth observing here that 397 respondents representing 31.3% follow the method of browsing at the shelves. Among the patterns of locating grey literature, use of library catalog and browsing at the shelves are being followed to a greater extent.

It is evident from chart 1 that the large majority of the faculty guides have indicated that the technical reports, proceedings of CSW and dissertations are useful in the higher range of 50% to 75%. However, the doctoral theses, trade literature and institutional publications are useful in the lower range of 25% to 50%. From this it is clear that the entire collection of GL is moderately useful for faculty for teaching purpose.

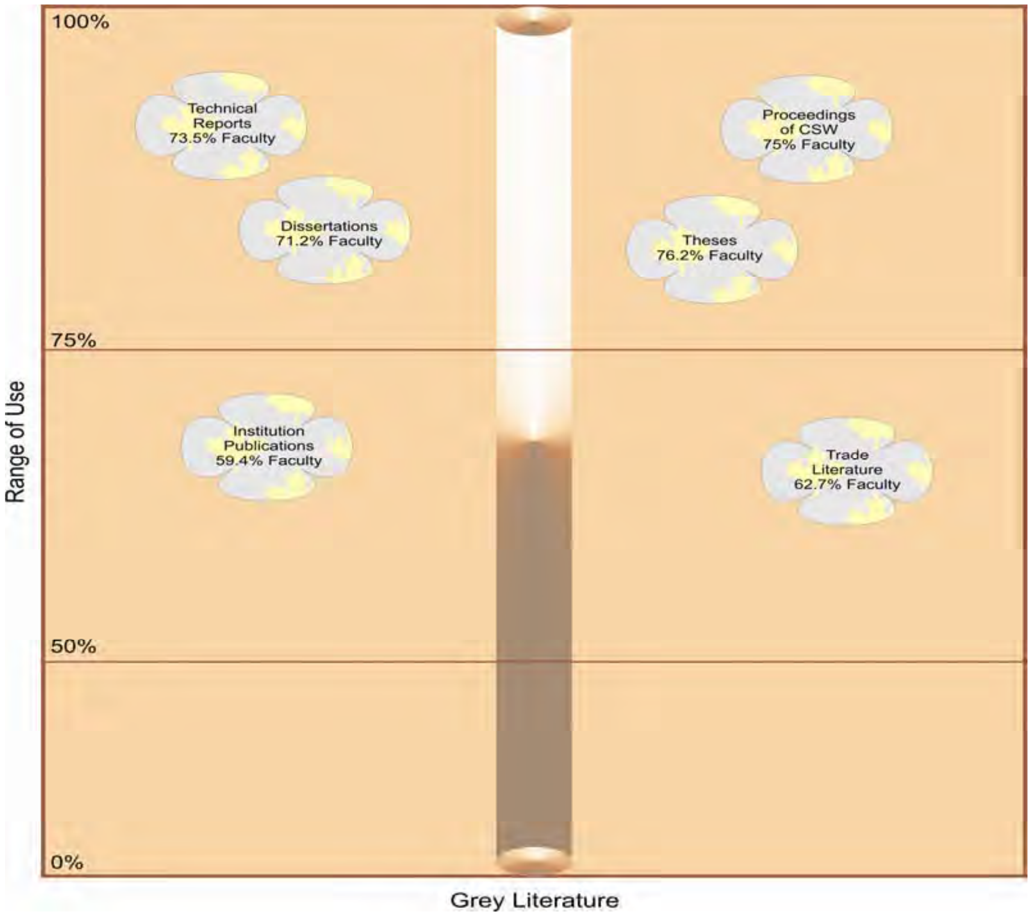
Chart 1 :Range of use of GL by majority of faculty for teaching



The extent of the use of GL for teaching is low when compared to the use for research purpose. Invariably, scholars need current literature and up-to-date statistics for preparation of research articles for scholarly journals, research papers to be presented in the seminars and conferences, bringing out research monographs, delivering special lectures and keynote addresses, and preparation of project works and research reports/theses. All these constitute the task of research work, for which scholars rely upon library services on modern lines.

Chart 2 depicts that large majority of the faculty members have opined that GL comprising theses, dissertations, technical reports, and the proceedings of the CSW are useful in the highest range of 75% to 100%. However, trade literature and institutional publications are found useful in the higher range of 50% to 75%. From this it is clear that all type of GL collections form vital sources of information for faculty for research

Chart 2: Range of the use of GL by the majority of the faculty for research



The below bar diagram clearly indicates that a large percentage of professors and associate professors rely upon GL for research. However, comparatively lower percentage (59.6%) of lecturers does depend on GL collections for research tasks. Professors possess long research experience when compared to lecturers.

Chart 3:Faculty-wise frequency of access to GL for research

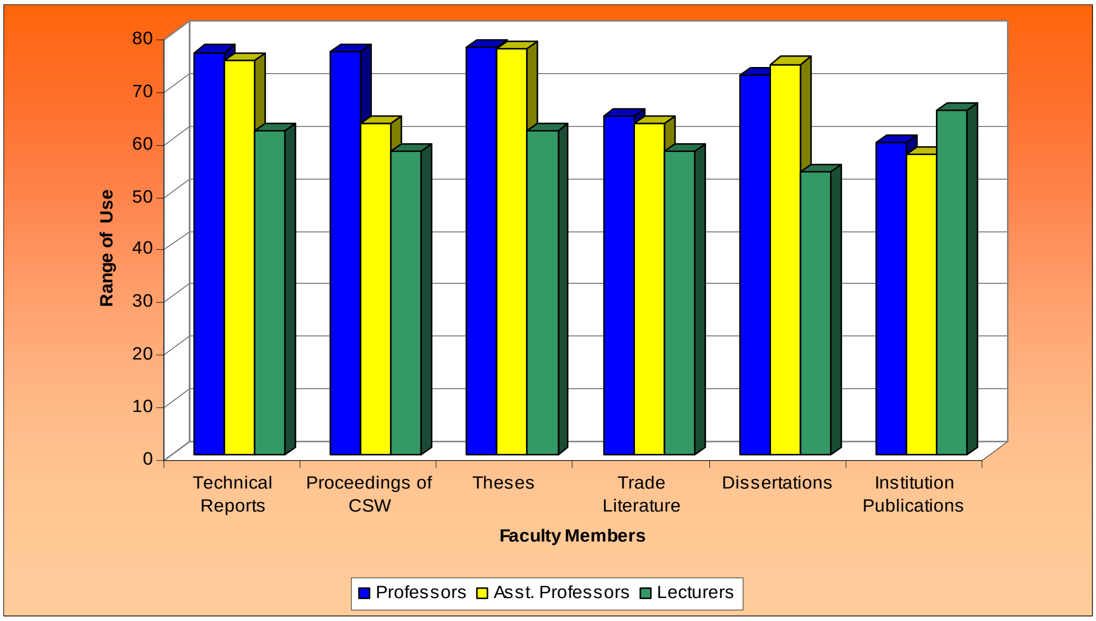
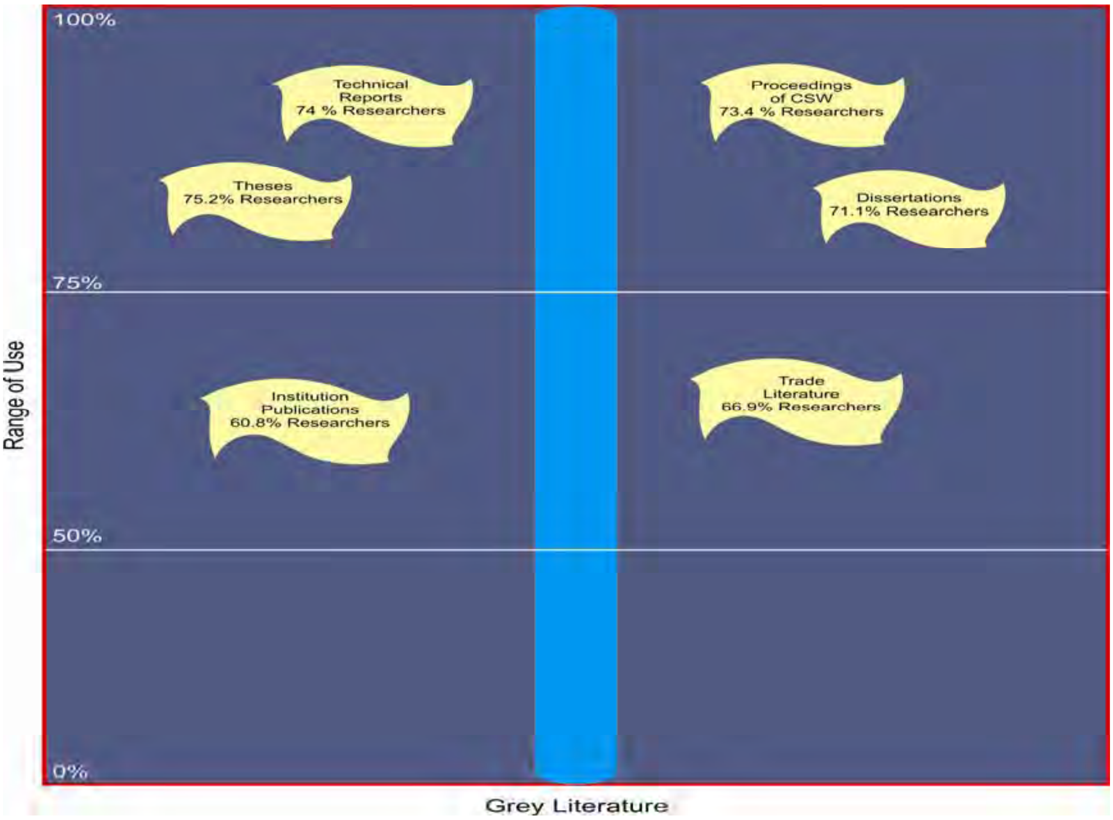


Chart 4: Range of the use of GL by the majority of the research scholars

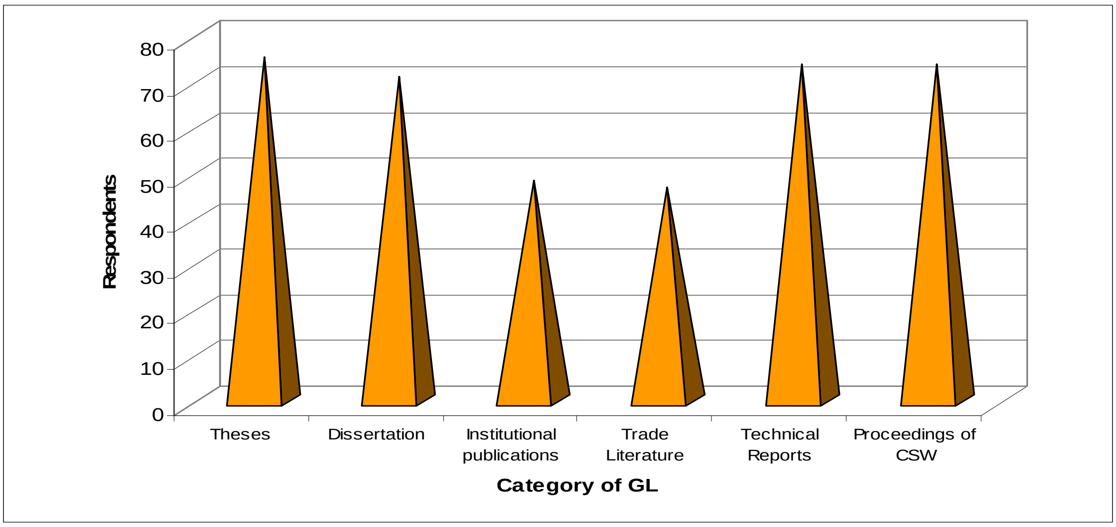


A large majority of the researchers rely upon theses, dissertations, technical reports and proceedings of CSW in the highest range of 75% to 100%. However, the use institutional publications and trade literature is noticed in the range of 50% to 75%. From this it is clear that the research scholars use GL collections of all types to a great extent for research and publications.

Table 6: Use of GL in the highest range for research and publication

S. No.	Category of GL	Use in the highest range of 75% to 100%		
		Faculty Guides	Research scholars	Total
01	Theses	259 76.2%	699 75.2%	958 75.4%
02	Dissertations	242 71.2%	661 71.1%	903 71.1%
03	Institutional publications	159 46.7%	456 49.1%	615 48.4%
04	Trade Literature	147 43.3%	449 48.3%	596 46.9%
05	Technical Reports	250 73.5%	688 74.0%	938 73.8%
06	Procedures of CSW	255 75.0%	683 73.4%	938 73.8%
	Average use in Percentage	64.3%	65.2%	64.9%

Chart 5: Use of GL in the highest range for research



As regards the six types of GL, an overwhelming majority, representing 71% to 75% of the users (faculty and researchers) opine that the theses, dissertations, technical reports and the proceedings of CSW are useful in the highest range for research purposes. On the other hand, slightly less than half of the users, 46% to 48% feel that the institutional publications and trade literature are useful in the highest range. The MAGIC project undertaken by Paul A. Needham and others in the year 2002(p.7), pointed out that the engineers often rely upon, technical reports, for research purposes.

Table 7: Faculty guides on the method of keeping track of recently released GL

Sl. No.	Method of keeping track of GL	Designation wise faculty guides			
		Professors	Asst/Associate Professors	Lecturers	Total
1	Regularly attend CSW	158 64.8%	5 7.1%	19 73.1%	222 65%
2	Browse the web	64 26.2%	45 64.3%	7 26.9%	86 25.3%
3	Visit other/local libraries	16 6.6%	15 21.4%	Nil	21 6.2%
4	Don't keep track of GL	6 2.5%	5 7.1%	Nil	11 3.2%
	Total	**244** **100%**	**70** **100%**	**26** **100%**	**340** **100%**

Here, 64.8% of the professors and 73.1% of the lecturers keep track of the release of the proceedings of the conferences, seminars and workshops as they regularly attend such programmes. 64.3% of the assistant and associate professors browse the web for information on grey literature. Only a small percentage of faculty guides go to local libraries in search of grey literature acquired recently. 6.6% of professors and 21.4% of assistant professors have the habit of relying upon the local library collections. A large majority of faculty guides to the extent of 65% keep track of recently released conferences, seminars and workshop proceedings by regularly attending the programmes.

Table 8: Response on the weeding-out of outdated GL collection

Sl. No.	Response on weeding out of the old GL	Faculty guides and Research scholars			
		Government colleges	Aided colleges	Private colleges	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%)	85 (67.%)
	Total	**29** **(100%)**	**371** **(100%)**	**870** **(100%)**	**1270** **(100%)**

67% of the respondents opine that the old grey literature should not be weeded out from the libraries. They prefer to see that the old grey literature collection still maintained in the library for easy access and use. In the present study, 33% of the respondents argue in favour of regular weeding out process which facilitates easy access. It is important to note that research centres of engineering colleges need to maintain old collection of GL systematically organised for use.

Now-a-days, library cannot meet the needs of its users from its own collection. To meet various needs of users and inter-disciplinary approaches, libraries need to depend on the information sources of regional libraries in a network environment. The information resource sharing programmes have come up on account of this trend.

Table 9: Gender-wise response on knowledge of GL of regional libraries

Sl. No.	Knowledge of GL available in other libraries	Gender-wise response of users					
		Male Respondents		Female Respondents		Total	
		No.	%	No.	%	No.	%
01	Yes, I have the knowledge	455	47.3	101	32.8	566	43.8
02	No, I have no knowledge	507	52.7	207	67.2	714	56.2
	Total	**962**	**100**	**308**	**100**	**1270**	**100**

It is evident from the above table that 43.8% of the users are familiar with the GL collections of regional libraries. A major segment of respondents express that they have no knowledge of GL collection of regional libraries. This segment of the respondents account for 56.2%.While considering gender-wise familiarity, 67.2% of the ladies who are faculty guides and researchers have no knowledge about the GL collection of regional libraries. Only 32.8% of the ladies are familiar as they make use of regional libraries. However, 46.8% of the male respondents are familiar with the GL collection of regional libraries. From this it can be deducted that a large segment of female respondents have no knowledge of GL collection of regional libraries. They find it difficult to move out of their colleges and travel for long distance to locate and refer GL of regional libraries. Therefore, there is need to compile the union list of important GL available in local libraries, as well as, engineering college libraries. Further, the bibliographic details of GL available in these local and engineering college libraries may be made accessible on-line.

6. Training Needs Projected by the Users and Librarians.

Table 10: Orientation / Training needs projected by the users

Sl. No.	Category of Response	Faculty Guides		Research Scholars		Total	
		No.	%	No.	%	No.	%
01	Strongly agree	159	46.8	345	37.1	504	42.0
02	Agree	158	46.5	499	53.7	657	50.1
03	Agree to a little extent	18	5.2	79	8.5	97	6.8
04	Disagree	5	1.5	7	0.8	12	1.1
1+2	Agree & strongly agree	317	93.2	844	90.7	1161	91.4

Training has become essential for academic growth. The necessity of training is noticed for updating professional skills. This is especially more so in case of the staff working in the field of Engineering Sciences sand Technology. Laloo (2002, pp. 140-148) describes in his work on "Information needs, information seeking behaviour and users" various methods of user education, especially the teacher users. The methods and steps explained here are useful. Considering the responses of both the categories of users, an overwhelming majority of the respondents representing 91.4% argue in favour of conducting orientation/training programmes for better access and use of GL collection. However, only 12 respondents representing just 1.1% hold negative view; such a kind of orientation or training is not necessary for them.

Table 11: **Training requirement projected by the librarians**

S. No.	Response on Training Requirement	Government Colleges	Aided Colleges	Private Colleges	Total
01	Positive response	5 8.33%	9 100%	43 89.2%	57 87.7%
02	Negative response	1 16.7%	Nil	7 10.8%	8 12.3%
	Total	**6** **100%**	**9** **100%**	**50** **100%**	**65** **100%**

The highlighting fact is that 57 librarians representing 87.7% express positive response stating that they need to attend training / orientation programmes in order to up-date themselves with latest trends and developments. The study conducted by Paul A.S. Needham, et.al.(2002, pp113-116), in the MAGiC Project Report submitted in the year 2002, furnishes conduct of seminar for working librarians. Such a seminar was organized at the British Library, London for the participants on managing effective access to GL collection (technical reports) available in the regional libraries.

Table 12: Extent of training needed by the librarians

S. No.	Extent of training requirement	Government Colleges	Aided Colleges	Private Colleges	Total
01	Higher extent	4 80.0%	7 77.8%	30 69.8%	41 71.9%
02	Moderate extent	1 20.0%	2 22.2%	11 25.6%	14 24.6%
03	Slight extent	Nil	Nil	2 4.6%	2 3.5%
	Total	**5** **100%**	**9** **100%**	**43** **100%**	**57** **100%**

Large segment of library staff representing 71.9% feels like undergoing intensive training in order to develop and handle GL collection on modern lines. Here, it is recommended to organize orientation programme or workshop by the Professional Associations for 3 to 4 days covering the themes of;

- Organization and building of GL collection
- Rendering information services using GL collection.
- Resource sharing in the network environment
- Creation of digital repositories

The workshop may also cover extension programmes such as display of important and current GL in engineering sciences, and invited talks by the resource persons.

7. CONCLUSIONS

The faculty and researchers who are actively engaged in research have felt the importance of GL. Having noticed the value of GL mainly for research, a large majority of them is found accessing and using GL collections existing in the holdings of the engineering college libraries in the higher range of 50% to 75%. The working librarians in the recognised research centres have been experiencing increased demand for GL.

The holdings of GL lack up-to-date collection in 43.7% of the libraries. However, it is encouraging noting that the majority of the libraries have systematically organised their GL collections with classification and cataloguing schemes. On account of this, the collections in most of the libraries are easy to access and use. The institutional publications and trade literature are hereby recommended to maintain in coloured box files and arranged in the chronological order so that, identification and access becomes easy for the users on account of the uniformity and consistency in colour coding.

The librarians, faculty members and the library committee have to come together for developing strong need based GL collection in every library. Irrespective of the cadre, every faculty and researcher has to be encouraged to recommend GL for procurement which they consider as vital for research. In this way, the collection can be strengthened in every engineer discipline. Educate the faculty and researches through orientation so as to identify, recommend and uninterruptedly access GL collections. Secondly, orientation is expected to develop knowledge and skills among working librarians for tracking, procuring, organising and facilitating better access to GL collections in the networked environment.

What is most important is augmenting the programme of sharing the resource of GL in the networked environment. The GL holdings of engineering college libraries and the technical libraries of the region have to be freely accessed and shared for mutual and reciprocal benefits. For the purpose of bringing the collection of GL together, there is need for developing the database of the holdings and facilitating on-line access. This venture in fact has to be encouraged and supported by all the concerned as this programme has to be developed and nurtured on co-operative basis, settling the issue of copyright.

While considering the GL by the category, VTU must insist research scholars to submit their doctoral theses and dissertations in soft copy so that, full-text of the digital version can be made accessible on LAN / web site. The Vidyanidhi Project of Digital Theses which has already digitized a part of the theses available in the holdings of university libraries in India may also consider taking up the theses available at VTU. Further, the proceedings of the conferences and seminars form valuable sources, projecting vast amount of data, including the trends in the field, useful for scholars. Creation of an open repository by the VTU will be of immense help as the large segment of researchers has inclination to refer to these proceedings, time and again. The materials for digitization may be decided upon considering the below indicated alternatives as well as, on the feed-back of the user community.

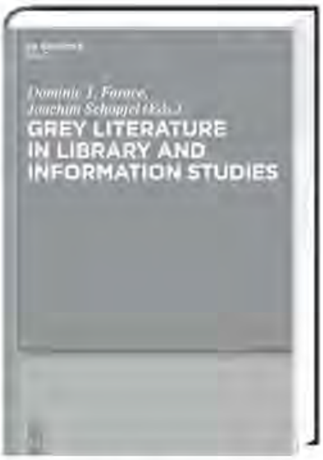
- Digitise the entire collection of the proceedings.
- Digitise only the recent ones of 3 to 5 years
- Dgitise only the active collection/those that are in demand.
- Digitise, considering the emerging new disciplines/thrust areas of research

The faculty and researchers to a considerable extent are found relying upon the local libraries, where the GL collections are comprehensive and up-to-date. However, it is highlighting that a large segment of female users though they have felt-need for grey literature, are ignorant of the collection of regional libraries on one hand, and face practical difficulties to freely move out of their colleges and travel a long distance and get into the environment of other libraries. Hence, there is need for creating on-line database in order to show the availability of various GL collections of engineering college libraries. Further, the libraries have to take initiative for getting the required GL on inter-library co-operation. It is also a good practice if individual libraries host new additions of GL on the web sites/blogs. Of late, open digital repositories facilitating on-line access is desired. The VTU has to venture in this direction to augment quality research and output of publications.

8. REFERENCES

1. Auger, C. P. (1998). *Information sources in grey literature*, 3rd edition. London: Bowker Saur.
2. Farace, Dominic J (2004). Grey literature: Grey matters in the world of networked information. *Publishing Research Quarterly,* 20 (1):3-91.
3. Farace, Dominic J, et.al.(2009). Open SIGLE, Home of GreyNet's Research Community and its Grey Literature Collections: Initial results and a project proposal. *The Grey Journal,* 5 (1): 48-52.
4. Gelfand, J. M. (2000). Grey literature poses new challenges for research libraries. *Collection Management,* 24 (1-2):137-147.
5. Gelfand, Julia (2007).Grey Literature: Taxonomies and structures for collection development. *The Grey Journal*, 3 (1): 7-16.
6. 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.
7. *Grey Net International (2009)* (http://www.greynet.org)
8. Laloo, Bikika T (2002). *Information needs, information seeking behavior and users*. New Delhi : Ess Ess Publications.
9. Lee, Lafleur and Rupp, Nathan (2005). Making grey literature available through institutional repositories. *The Grey Journal*, 1 (2): 67-72.
10. 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.
11. Needham, Paul A, et.al (2002). *The MAGIC Project : Managing Access to Grey literature Collection,* Final report. London ; The British Library.
12. Open SIGLE Repository (2010), http://opensigle.inist.fr
13. Ramadevi, Tella (2006). *Access and use of grey literature in social sciences :Problems and prospects.* New Delhi: Ess Ess Publications.
14. 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.
15. Siegel, Gretta E (2003): Capturing academic grey literature: Starting at home. "Perspectives on the design and STI". Proceedings of the Third. International Conference on Grey Literature, Luxembourg, Nov.13-14,1999. (http:// www.greynet.org / images / GL5.Siegel / preprint).
16. 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.

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

www.degruyter.com

Circulation Improvement of Articles in Journals written in Non-English Languages: A Special Journal Titles Translation List of Journals written in Japanese for the International Bibliographical Database

Mayuki Gonda, Katsuhiko Kunii, Hidemitsu Nakajima, Kiyoshi Ikeda, Keizo Itabashi
JAEA, Japan Atomic Energy Agency, Japan

Akemi Koike, Ayumi Igarashi
TOSS, Total Support Systems Corporation, Japan

Abstract
Non-English articles are still "Grey Literature" due to language barriers even though material circulation has improved like English articles with the expansion of the Internet era. In the INIS Database, bibliographic information such as titles and abstracts etc. is written in English. This feature of the INIS Database contributes to improvement of international circulation of scientific information from the nuclear field. However, titles of journals written in non-English languages were described in transliterated Roman alphabet which means that non-native users cannot understand the nature or subject of those journals. Consequently, we developed a special journal titles translation list written in the Japanese language for the INIS Database, as an attempt to improve the circulation of articles in journals written in non-English languages.

1. Introduction

1.1. The one of "Grey Literature" in the INIS Database

Today, the circulation of scientific information has improved internationally, and English which is the "international language" of choice takes the lead in scientific communication. On the other hand, non-English articles are still "Grey Literature" due to language barriers although physical circulation improves like English articles. In other words, non-English articles stand out from "Grey Literature" in the Internet era. However, we can find examples of non-English articles leading to Nobel Prizes which cannot be ignored. In this paper, we will introduce one of the Japan Atomic Energy Agency (JAEA) Library's attempts to improve the circulation of articles written in non-English languages.

The International Nuclear Information System (INIS) is the open access database for published scientific literature on the peaceful uses of nuclear science and technology. INIS was established in 1970 by the International Atomic Energy Agency (IAEA). The INIS Database now contains over 3 million bibliographic records. INIS is operated by the IAEA in collaboration with 123 member states and 24 international organizations.

The INIS Database contains 25% bibliographic information of articles written in non-English languages (Fig. 1). As of November 2010, 3% of the articles in the INIS Database are written in Japanese which is the 4th ranked language. This is not so high but not too low. We cannot ignore these articles written in non-English languages.

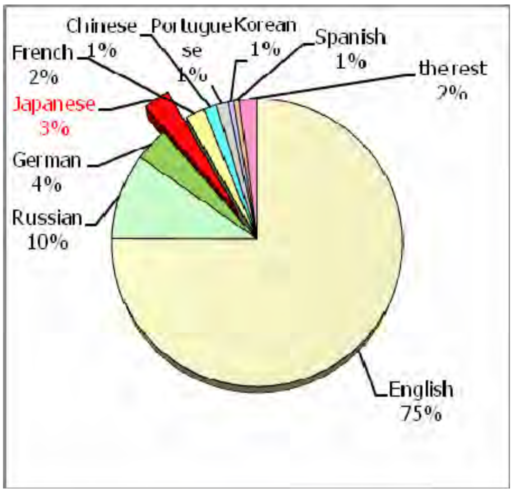


Fig. 1 Languages of articles in the INIS Database (As of Nov 2010)

1.2. Journals regularly scanned at the INIS National Center of Japan

JAEA was established in October 2005 as the sole research institute in Japan dedicated to comprehensive research and development in the field of nuclear energy, integrating the former Japan Atomic Energy Research Institute (JAERI) and the former Japan Nuclear Cycle Development Institute (JNC).

The JAEA Library is one of the largest nuclear information centers in Japan, and we support the scientific research and development activities throughout Japan. The JAEA Library maintains a circulating collection of books in physics, chemistry, computer science and engineering. The JAEA Library also collects the information on papers contributed to journals by JAEA staff, and publishes JAEA Reports for the dissemination of JAEA R&D results via the Internet. Another main role of the JAEA Library is working as the INIS National Center of Japan.

JAEA is designated by the Japanese government as the INIS National Center of Japan. We prepare bibliographic records about nuclear related literature published in Japan, and submit them to the INIS Database. Furthermore, we endeavor to promote the utilization of the INIS database in Japan. The INIS National Center of Japan sends 5,000 items to IAEA every year which is the 3rd most numerous among member countries in 2009 (Fig. 2).

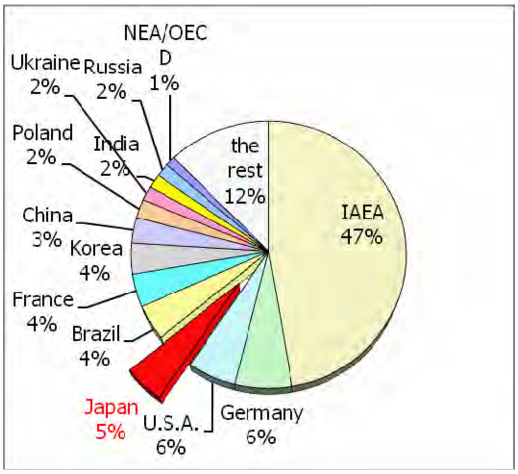


Fig. 2 Input from Member States (2009)

Since it is impossible for the JAEA Library to cover all journals published in Japan, JAEA plays a key role to integrate input prepared by the Japan Science and Technology Agency (JST) and the International Medical Information Center (IMIC). There are 491 regularly scanned journals in Japan. (Table 1)

Table 1 Journals regularly scanned in Japan

	Titles	Journals regularly scanned
JAEA	328	Core journals in nuclear field and open access journals
JST	33	Non-core journals (Environmental sciences, life sciences etc.)
IMIC	130	Non-core journals (Medical sciences etc.)

2. Backgrounds

2.1. The INIS bibliographic description for journal title

In the INIS Database, bibliographic information such as titles and abstracts etc. are written in English. This feature of the INIS Database contributes to improvement of international circulation of scientific information in the nuclear field. However, most titles of journals written in non-English languages were described in transliterated Roman alphabet due to the first edition of "INIS: Guide to Bibliographic description" (IAEA-INIS-1) which states that "A Journal title is entered in the original language". It is means "A journal title in non-English languages should be described in transliterated Roman alphabet." At that time, this rule was based on the cataloging rule of International Organization for Standardization (ISO). Since September 1992, this rule has changed slightly, in IAEA-INIS-1 (Rev.8) and English titles of journals written in non-English languages are available. (Table 2)

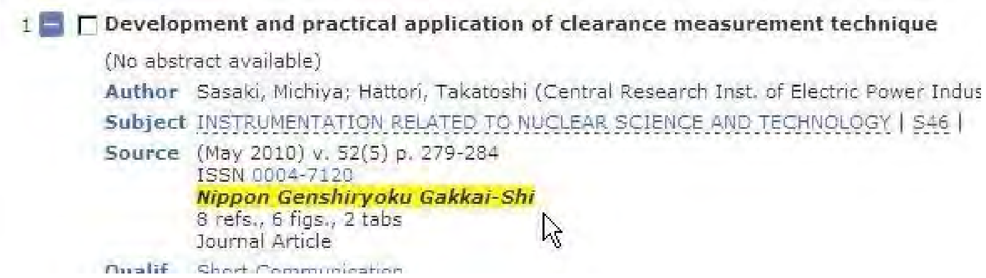
Table 2 INIS bibliographic description for Journal title

IAEA-INIS-1 (Rev. 7)
6.4.7 Serial/Journal Title (Tag 230, Bibliographic Level S) *A Journal title* is entered in the original language, translated if Necessary.
IAEA-INIS-1 (Rev. 8) (1992.9-)
3.8.4. Full Journal Title (Tag 229, Level S) This field is only used for Type of Record J. It is a mandatory field and contains a standardized version of the original title of the journal. The title may be in English or in a non-English language.

As in Table 2, English titles of journals written in non-English languages are available in the INIS Database now. However, INIS national Center of Japan is not using English titles of journals written in

Japanese for bibliographic descriptions of the INIS Database. Some English titles of journals written in non-English languages are listed in "INIS: AUTHORITY LIST FOR JOURNAL TITLES" (IAEA-INIS-11), but they are not perfect.

Figure 3 is a sample of the INIS search result for a journal written in Japanese. This bibliographic information is described in English. However, the journal title is described in transliterated Roman alphabet. The English title for the "Nippon Genshiryoku Gakkai-Shi" is the "Journal of the Atomic Energy Society of Japan". Non-Japanese-native users cannot understand the nature or subject of journals with this title and therefore do cannot obtain useful articles. This language barrier should be broken as much as possible for improved circulation of journals written in non-English languages.



1 □ Development and practical application of clearance measurement technique

(No abstract available)

Author Sasaki, Michiya; Hattori, Takatoshi (Central Research Inst. of Electric Power Indus

Subject INSTRUMENTATION RELATED TO NUCLEAR SCIENCE AND TECHNOLOGY | S46 |

Source (May 2010) v. 52(5) p. 279-284
ISSN 0004-7120
Nippon Genshiryoku Gakkai-Shi
8 refs., 6 figs., 2 tabs
Journal Article

Qualif Short Communication

Fig. 3 Search result of the INIS Database

2.2. Purpose of this study

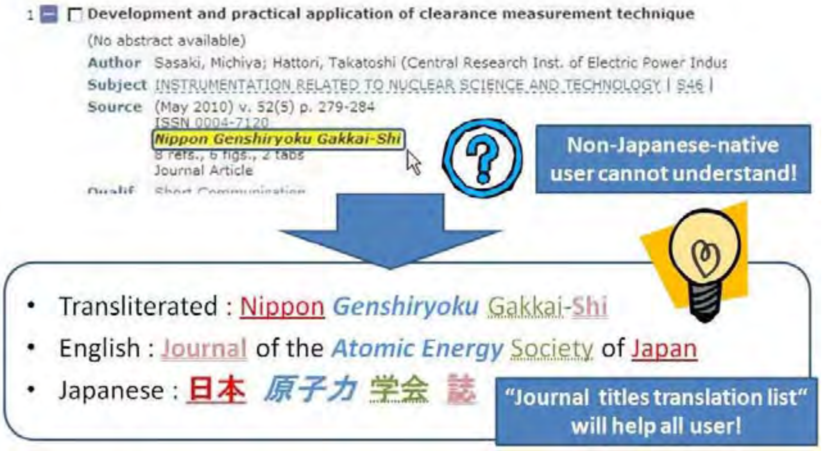
As described in a previous section, in the INIS Database, for only journal titles in transliterated Roman alphabet means that non-Japanese-native users are unable to understand meanings. On the other hand, for only English title, non-Japanese-native users can locate journals; however, it is difficult for Japanese-native users to identify the journal since they recognize journals with Japanese title. By including both the transliterated Roman alphabet title and the English title, all users can locate and identify the journal. Additionally, Japanese titles described in Chinese characters are more convenient for users who can understand the meanings like Japanese and Chinese people. Such feature of different language titles is shown in Table 3.

Table 3 Feature of each title

	for Japanese-native	for non-Japanese-native
in Transliterated Roman alphabet (the INIS Database)	can locate.	difficult to understand meanings...
in English	can locate, but difficult to identify…	can locate.
in Japanese	can easily locate!	difficult to understand meanings...

In other words, the journal titles translation list including titles in transliterated Roman alphabet, English and Japanese is helpful for all users (Fig. 4). The list will also improve the circulation of articles in journals written in Japanese. Consequently, we have developed a special journal titles translation list of journals written in Japanese for the INIS Database.

The purpose of developing the list is for non-Japanese-native INIS users' convenience. Another purpose is to improve the circulation of articles in journals written in Japanese. Additionally, this list is effective to manage regularly scanned journals in the INIS National Center of Japan as basic data of Japanese journals in the nuclear field.



1 □ Development and practical application of clearance measurement technique
(No abstract available)
Author Sasaki, Michiya; Hattori, Takatoshi (Central Research Inst. of Electric Power Indus
Subject INSTRUMENTATION RELATED TO NUCLEAR SCIENCE AND TECHNOLOGY | S46 |
Source (May 2010) v. 52(5) p. 279-284
ISSN 0004-7120
Nippon Genshiryoku Gakkai-Shi
8 refs., 6 figs., 2 tabs
Journal Article
Qualif Short Communication

Non-Japanese-native user cannot understand!

- Transliterated : Nippon Genshiryoku Gakkai-Shi
- English : Journal of the Atomic Energy Society of Japan
- Japanese : 日本 原子力 学会 誌

"Journal titles translation list" will help all user!

Fig. 4 Purpose of developing the list

3 The journal titles translation list of journals written in Japanese

3.1 Outline of the list

We use Microsoft Excel for making the list. As a first step, we checked the JAEA Central Library catalog and examined titles of regularly scanned journals written in Japanese. These 154 titles are core journals in the nuclear field collected in Japan. Data are flags of key journals, titles (transliterated Roman alphabet/Japanese/English), ISSN-L, ISSN (print, online, digital media), CODEN, media type (print, online, digital media), frequency of publication, publisher (Japanese/English), URL (online journal, publisher), type of journal (academic, letters, membership, technical etc.), language and the call number of the JAEA Library. We surveyed not only journal titles in each language but also other information aimed at users' convenience and our management of the list.

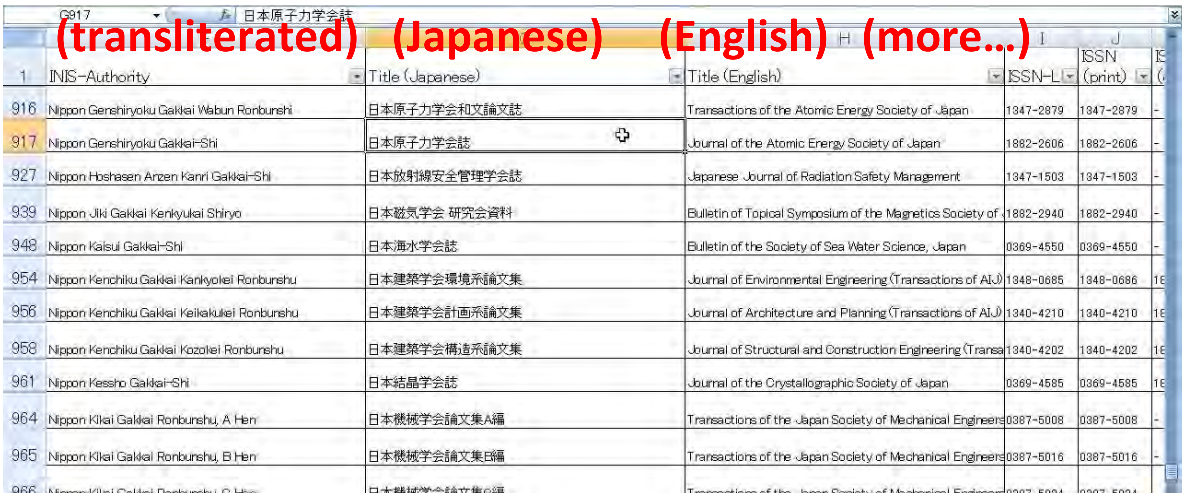
We surveyed journal titles and items in printed media, online journal, journal website, publisher website, NDL-OPAC (Japan's National Diet Library OPAC), NACSIS Webcat and INIS: AUTHORITY LIST FOR JOURNAL TITLES (IAEA-INIS-11).

3.2 Advantages

Figure 5 is the list made up through this work which contains journal titles with other information such as ISSN, CODEN and publisher etc. The transliterated Roman alphabet titles in the INIS Database are identified and are necessary not only for the INIS users but for its bibliographic agencies consisting of the INIS secretariat, other INIS member countries and international organizations like us. The Japanese titles written in Chinese characters are easily identified visually due to feature specific to Chinese characters, each of which has a meaning, making Japanese titles very helpful to not only Japanese but Chinese people as well. Then the English titles are, of course, most useful for most people who use English commonly.

The point of this list is the greater contribution to the circulation of journal articles written in Japanese. That is, all people such as users, bibliographic agencies etc. can identify and locate on websites, a specific journal by referring to this list even if their language is different from each other, i.e., English, Japanese and Chinese. This list is also very useful for people who would like to refer to, or access, Japanese information, knowledge and skills etc., for example, in-situ specifics, in the field of nuclear science and technology.

Fig. 5 Sample of the list



1	INIS-Authority	Title (Japanese)	Title (English)	ISSN-L	ISSN (print)
916	Nippon Genshiryoku Gakkai Wabun Ronbunshi	日本原子力学会和文論文誌	Transactions of the Atomic Energy Society of Japan	1347-2879	1347-2879
917	Nippon Genshiryoku Gakkai-Shi	日本原子力学会誌	Journal of the Atomic Energy Society of Japan	1882-2606	1882-2606
927	Nippon Hoshasen Anzen Kanri Gakkai-Shi	日本放射線安全管理学会誌	Japanese Journal of Radiation Safety Management	1347-1503	1347-1503
939	Nippon Jiki Gakkai Kenkyukai Shiryo	日本磁気学会 研究会資料	Bulletin of Topical Symposium of the Magnetics Society of J	1882-2940	1882-2940
948	Nippon Kaisui Gakkai-Shi	日本海水学会誌	Bulletin of the Society of Sea Water Science, Japan	0369-4550	0369-4550
954	Nippon Kenchiku Gakkai Kankyokei Ronbunshu	日本建築学会環境系論文集	Journal of Environmental Engineering (Transactions of AIJ)	1348-0685	1348-0686
956	Nippon Kenchiku Gakkai Keikakukei Ronbunshu	日本建築学会計画系論文集	Journal of Architecture and Planning (Transactions of AIJ)	1340-4210	1340-4210
958	Nippon Kenchiku Gakkai Kozokei Ronbunshu	日本建築学会構造系論文集	Journal of Structural and Construction Engineering (Transa	1340-4202	1340-4202
961	Nippon Kessho Gakkai-Shi	日本結晶学会誌	Journal of the Crystallographic Society of Japan	0369-4585	0369-4585
964	Nippon Kikai Gakkai Ronbunshu, A Hen	日本機械学会論文集A編	Transactions of the Japan Society of Mechanical Engineers	0387-5008	0387-5008
965	Nippon Kikai Gakkai Ronbunshu, B Hen	日本機械学会論文集B編	Transactions of the Japan Society of Mechanical Engineers	0387-5016	0387-5016

3.3. Issues occurred

Some journals have no official English title. These titles were filled in transliterated Roman alphabet of official Japanese title and will be identified by "publisher" and "type of journal". For example, "BUTSURI" ("Physics" in Japanese) is identified by "Physical Society of Japan (publisher)" and "Membership journal (type of journal)".

Meanwhile, some journals have those English titles as a cover design or web design and it is difficult to determine the official English title. In this list, such titles are contained for users' convenience.

4. Conclusion

We are now checking and evaluating the first step of the list. After that, we will complete the list (Non-core, titles of journal written in English) and release the list on the INIS Database and JAEA Library website. Subsequently, we will distribute the printed list to foreign researchers at the JAEA Library. In the future, we would like to incorporate the English titles into the INIS Database in cooperation with the INIS Secretariat and promote the idea of the list in other non-English languages.

In this paper, we introduce one of our continuous efforts to improve international circulation of articles written in non-English languages. The circulation of scientific information improves spectacularly with the development of computer and globalization today. However the international circulation of academic information is still done in English which is the "international language" and the circulation of non-English languages still has many issues to be solved. These issues are common for all non-English languages, especially, for non-alphabetical languages. These are very basic issues, but very important. And these are old issues, but still relevant to today.

References

Ikeda, Kiyoshi. Current status and short history of grey literature: Focusing on the International Conference on Grey Literature. Joho Kanri. 2009, 53(8), p. 428-440. (Japanese).

Nagayama, Taisuke; Narui, Keiko; Habara, Takako. "A study of bibliographic description of journals written in Japanese on international circulation". Proceedings of the 5th National Convention for the Study on Information and Documentation, Tokyo, Japan, 14-15 November 1968, The Japan Information Center of Science and Technology, 1968, p. 111-116. (Japanese).

ISO Standards for Quality Education and Training for Facilitating Access to Grey Literature

C.P. Ramasesh, University of Mysore, and
N. Chowdappa, BMS College of Engineering, India

Abstract

To increase the level of the user satisfaction and to continually improve upon the performance of service in libraries, imparting education training and is found a prerequisite condition for librarians who are involved in visualizing, acquiring, organizing and facilitating access to GL collection. Paper mainly focuses on the issues concerned with the performance of the professionals and the modus operandi of training at the Centre for Information Science and Technology (CIST), University of Mysore, under ISO 9001 standards of quality monitoring system. Paper highlights the eight principles of quality management system and describes issues concerned with quality audit and corrective measures and preventive measures to overcome defective procedures. Paper also depicts improvement in quality performance at the library of the CIST based on the feed-back of students during the years 2005 to 2009. Attempt is also made to highlight the aspects related with the proof of quality education.]

Keywords: ISO Standards, Library services; Grey literature, User education; Grey literature, Orientation programme; QMS, Grey literature; Grey literature, Customer delight.

Standards are primary source of information which gives specification for uniformity and consistency in practice and production. Standards are evolved by national / international bodies. The International Organization for Standardization (ISO) is a world-wide federation established in Geneva. It has the responsibility of evolving standards depending upon the need across the world. Specifically speaking, the ISO 9001 is a generic Standard for building, operating and documenting quality management system. The mission of ISO is to provide international standardization to facilitate worldwide exchange of goods and services.

In India, the Ministry of Human Resources Development (Education) has been allocated Rs.445 billion for the plan outlay. This in fact is a real big boost. On account of this several reforms have been brought.

1. Encourage ‘Quality Research’
2. Encourage ‘Brain Gain Policy’
3. ‘Quality Assessment’ and ‘Accreditation’ has become mandatory
4. Conduct of ‘Teacher Capacity Building’ programmes.
5. Conduct of ‘Employable Skill Development programmes.

While coming to the realm of education and training for library and information services, it has become more and more technology oriented due to the application of information and communication technology. There is really a special emphasis for hands-on experience and learning skills. This in fact necessitates the appropriate orientation, as well as, quality education and training. One of the means of achieving the quality and excellence in education and training is to go for ISO certification. To be specific, the needs for ISO are;

- Day-by-day the needs of students are becoming more and more stringent
- To continually improve the quality of training and education, including the curriculum and the lesson plan
- Uniformity is insisted for global marketing of products and services
- Provide confidence to top management regarding quality being achieved and sustained.
- Provide confidence to the customers (students) that the intended quality is being delivered in education and training programmes.
- To reduce the cost of operation and waste to the maximum extent for the benefit of students/trainees.
- To get international recognition as top class / world class organization as far as quality education is concerned. Above all, it adds in gaining public opinion.
- To meet the needs of students/trainees constantly beyond expectations through periodical evaluation. Above all helps to achieve the ‘Customer Delight’.

PRINCIPLES OF Q.M.S : The quality management system (QMS) necessitates the need for application of certain principles. There are a set of eight principles which need to be adopted for achieving quality in performance. They are:

1) Student focused organization
2) Development of leadership qualities
3) Involvement of people
4) Process approach
5) Systems approach
6) Continual improvement
7) Factual approach to decision making
8) Mutually beneficial supplier relationship (relationship with service providers).

IMPLEMENTATION OF ISO STANDARD: Noticing the need and importance of quality in education, the management has to decide the steps to be taken for preparation. There are several issues which are to be considered at the preparatory stage.

1. **Defining the quality policy** - The organization has to define its policy in broader terms. It is on this statement all functions and activities will be evolved or determined. Policy statement will be usually idealistic yet achievable. Everyone in the organization should have commitment to the policy stated by the organisation.
2. **Stating the objectives** - The organization has to state all its objectives in the light of the above stated policy statement. As far as possible, the objective will have to be quantifiable or measurable. Measurable objectives help the organization to demonstrate continual improvement.
3. **Preparation of quality manual** - The manual giving the details of the organization, its functions and activities, responsibilities of the staff, the curriculum and the relevant divisions has to be prepared.
4. **Preparation of procedure manual** - The procedure manual states the lesson plans, the instructions for theory classes and practical sessions, including the tests, assignments, project work / field work and so forth, indicating the number of hours per topic to be covered. Accordingly, the entire programme has to be executed. All necessary instructions for conduct of theory and practical sessions will have to be indicated in the procedure manual. The procedure for the management of library services ,duties and responsibilities of each staff, the list of quality records to be maintained by every staff will also be stated.
5. **Curriculum design** - In the educational institutions, the design of the curriculum plays a major role. It is the main indication of the quality / excellence. The curriculum needs to be improved every year noticing the trends and developments on one hand and the ways and means of facilitating access to grey resources on the other. The procedure also include the search strategies and on-line information resources available. The design indicates number of hours of theory classes, practical sessions, tests, project work, field work etc.
6. **Feedback** - For effective implementation of the education and training programme, feedback from students/trainees as to the clarity, expression, subject coverage, interaction, knowledge in the subject and effectiveness of teaching of each staff need to be collected. The data be tabulated and interpreted for demonstration of improvements in the system.
7. **Training matrix** - Training is an important ingredient and therefore periodical training needs to be conducted for staff to keep them abreast of recent trends and also to develop skills in handling IT based curriculum. The ISO version released in the year 2008 gives emphasis for training. Every year a training matrix be prepared seeking the area of training needed by every staff member and the training has to conducted well in time to ensure what they teach and train in the class and lab is up-to-date. This way there is sufficient room to increase the competency of the personnel.
8. **Continual improvement** - ISO recommends the Kaisen Approach - the continual improvement. Every individual in the organization has a significant role to play, cooperate and contribute to the improvement in the performance of the system. The improvement / performance according to ISO be quantified and the organization should demonstrate continual improvement or step by step improvement.

Each staff in the organisation has his/her own assigned duties and responsibilities. He/she had to concentrate for creativity, cooperate for support, contribute for performance. Work of every staff member is significant and his/her contribution through his own creativity is expected and appreciated. It is on this basis that the organization will demonstrate its improvement, sustenance and quality in service. To achieve the quality at every step, the activities are to be planned and checked for expected results and further improvement. This is popularly called as PDCA method (Plan-Do-Check-Act). The plans that are going to be implemented should have certain characteristics like:

- To be implemented in short time
- With least effort, and
- With minimum resources

CUSTOMER DELIGHT : The concept of improvement in performance has gained momentum on account of the Japanese technique of: Khaisen Approach", which aims at day-to-day improvement in the organization. Whatever may be the situation today; there should be commitment, cooperation, and contribution from every staff for improvement of the system. Team work and leadership both are required and emphasized here for sustenance of quality education. By quality they mean the totality of characteristics and features of service that bears its ability to satisfy the stated or implied needs of the customers. Achieving customer delight means serving the customer beyond his expectations Here, all those activities of the organization which intend to gain appreciation of its customers beyond their expectations is the aspect of 'delight'.

Quality Education and Training at CIST: Centre for Information Science and Technology (CIST) is a self sustaining educational institution offering five major information technology courses at post-graduate level.

- Multimedia Production
- Software Development
- E-commerce
- Bio-informatics and
- Geographic Information System

CIST has been rendering educational services under the administrative jurisdiction of the University of Mysore. Of late, to maintain quality performance, the Centre is relying upon the quality manual and feedback of the customers.

To project the performance of CIST during 2005 and 2009, the feedback analysis reports were taken as the source material for the projecting quality performance. The CIST has been conducting user education or orientation programmes for students and librarians on the following aspects.

- Library facilities and services
- Access to information sources
- Participation in collection development
- Use of regional library facilities and services
- Library extension programmes
- Digital library systems

CIST library has a collection of 6130 sources, out of which project works, manuals, and course-ware and report literature form 48.4% of the collection of the library. The students from all the five courses are also trained to make use of the facilities and services of the University Library where the collection exceeds 6,08,400 information resources. Out of the entire collection of the University Library, GL collection constitutes 11.9%. The GL collection mainly comprises of technical reports, theses and dissertations, trade literature, proceedings of the conferences, seminars and workshops and the university publications.

GL Collection at the University and CIST

		Total GL Collection at the	
Sl. No.	**CATEGORY OF GL**	**University Library**	**CIST Library**
1	Project works, Theses and Dissertations	58%	58%
2	Institutional Publications	15%	23%
3	Technical Reports	21%	12%
4	Proceedings of CSW	6%	7%
	TOTAL GL COLLECTION	**72400 (11.9%)**	**2965 (48.4%)**

As regards user training / orientation, the CIST librarians as per the guidelines indicated in the quality manual, conduct classes for a period of 3-4 days and obtain feedback from the students. On the other hand, the LIS faculty members handle training classes for librarians for a period of one week. The responses were obtained every year from 190 - 275 students who were studying in CIST.

FEED BACK FROM STUDENTS (QUESTIONNAIRE)

Name …………………… Course ………………. Year ……………………..

E = Excellent (81-100%), G = Good (61-80%), S = Satisfactory (41-60%), N = Need to Improve (up to 40%)

01	Teaching Performance	□
02	Teaching Methodology	□
03	Library Resources and Services	□
04	IT Infrastructure Facilities	□
05	Extension Programmes	□
06	Counseling and Security	□

07	Performance of Individual Faculty	
Name ……………	Paper ………………………………	□

Based on the feedback rating, the below table depicts the improvement in the performance in library services, starting from 2005 to 2009. The feedback rating was obtained as to the performance the library on the three aspects.

a) Rating on organization and browsing facility.
b) Rating on the information sources available in the library.
c) Performance of library staff in rendering information services.

Continual improvement in library resources and services

S.L	RATING ON THE CRITERIA	2005	2006	2007	2008	2009
1	Information Resources/GL Available in the Library	65%	66%	66%	68%	70%
2	Organisation of GL/sources and Browsing Facility	62%	65%	67%	67%	68%
3	Performance of the Library Staff in rendering Service	62%	62%	64%	66%	69%

The statistics in the table shows the continual improvement in the performance over the past five years. The ratings given by the students as excellent performance above 80%; and good performance above 60% are considered here.

CONTINUAL IMPROVEMENT IN LIS (E & G RATINGS)

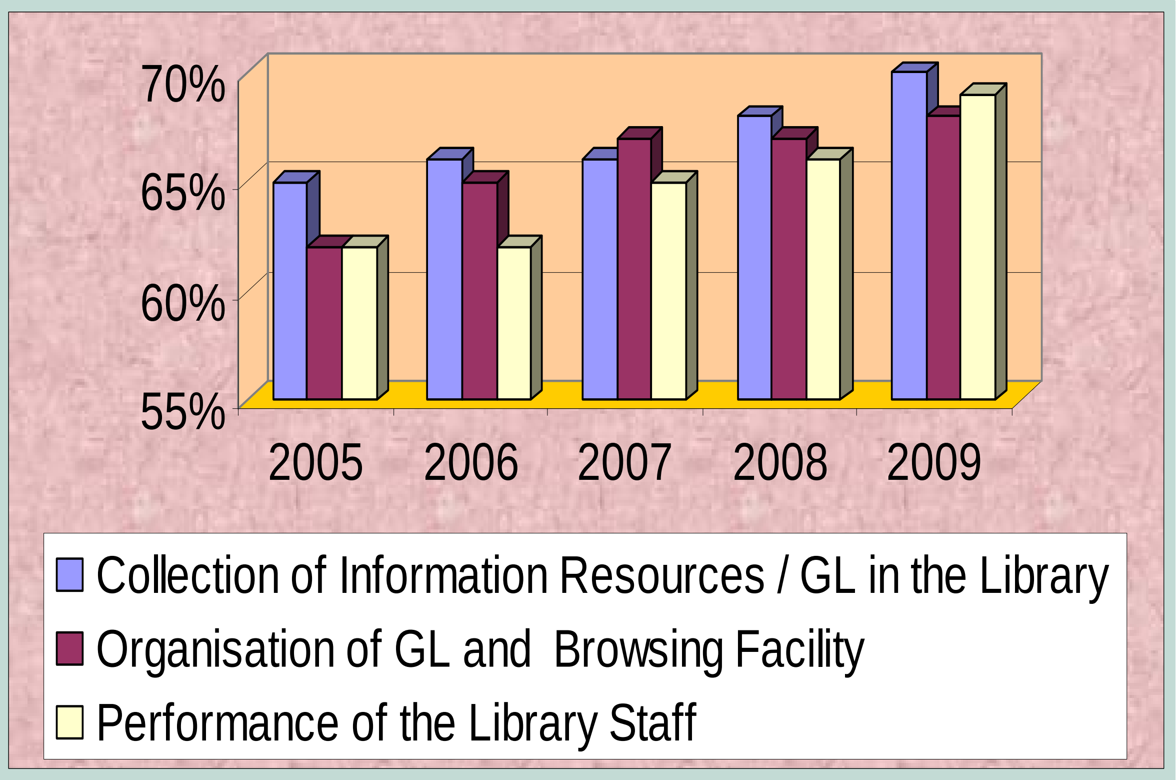
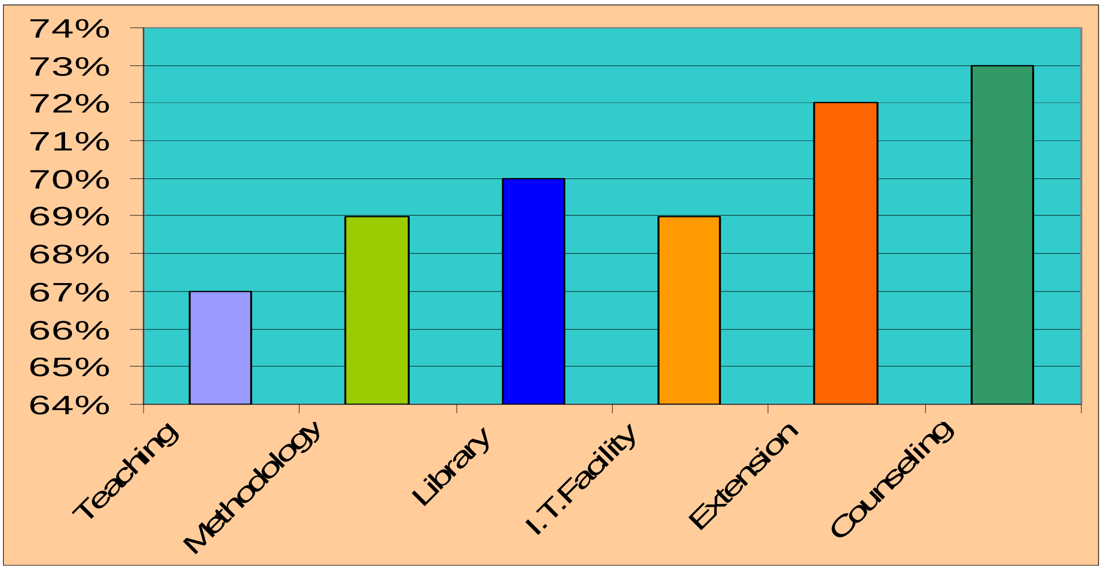


Table projects the continual improvement in library management and readers services. The diagram with the statistics helps the institution to track the record of continual improvement. This vindicates how the ISO standard 9001- 2008 is useful in extending quality services and further helps in improving the performance of the library on continual basis. The below chart depicts performance of various criteria.

CRITERIA-WISE PERFORMANCE OF CIST CIST(2009)



The following records are placed as a proof for having rendered quality teaching and training.

PROOF OF QUALITY SERVICES (Quality Records / Best Practices)

1) Orientation Programme – Attendance signed by the students
2) Certification of lesson plan
3) Circulation and reference statistics
4) Distribution of courseware / library guide

5) Conduct of extension programmes
6) Inter-library lending statistics
7) Feedback analysis report
8) Circulation of new arrival lists
9) Sources recommended and actual procurement
10) Library websites projecting holdings & services
11) Handling customer grievances

PERFORMANCE AUDIT: The management should plan and execute assessment / audit to verify whether the activities comply with planned arrangements to determine the effectiveness of the quality system. The quality assessors or the performance auditors are concerned with verification of conformances. In the course of seeking conformances, they may come across non-conformances, between what you say and what you do.. Noticing the non-conformities, institution has to initiate corrective measures and preventive measures against the defects, verify the result and close the issue of non-conformances.

In fact, it is the management which is going to define policy and objectives, including the duties and responsibilities. It is the management on the other side which is going to give assurance to its customers. ISO auditors are rather concerned with verification and cross checking for "non-conformances", between what the management has said or assured, and what it has practiced. Auditing is not a fault finding mission, it is a fact-finding exercise focuses on;

- System deficiencies
- Human slip-ups (unidentified errors)

Non-conformances are classified, into three categories. Critical, Major and Minor. There will be observations in addition to these three which can be set right in a day or two. Those lapses which may cause injury, and where there is risk to the life of an individual staff or customer, they are noted as critical non-conformances. Those non-conformances that affect objectives to a greater extent are classified as major ones Ex: Lapse in coverage of the lesson plans, lack of internet facility. The minor ones are usually lack of systematic arrangement of volumes in the library , lapse in conducting sufficient classes of orientation lectures for students, lack of reading facilities in the libraries. The observations will be with reference to wrong filing of records, delay in the tractability of records, lack of cleanliness and so forth, which can be corrected immediately. The management has to initiate corrective and preventive measures to rectify the deficiencies indicated by the auditors from time to time.

CONCLUSION

By opting for ISO certification, the management and the students are assured of the quality of education and training as planned and documented. Therefore, ISO is the one of the best means to control and improve the performance in libraries, where IT based training plays a major role. The process gives opportunity for everyone to involve fully, develop leadership qualities, and contribute creatively. Everyone in the organization will get recognition for his creativity, competency and contribution. By these ways and means, more than anything, there is going to be "work culture" in the organization and appreciation for the quality education and training from the student community. By adopting for ISO procedures or in the process of ISO movement, the LIS schools become more and more 1. Student oriented, 2.Technology oriented and 3. Revenue oriented. Further, QMS

- Provides written procedures that define authority responsibility and interfaces
- ISO provides documented system which controls the activity / education / service, systematically.
- Developing systematic steps for maintenance of gadgets and operations
- Ensures all non-conformities (errors, deficiencies, objections, and complaints) through constant feedback
- Opportunity for the evaluation of the customer satisfaction.
- Opportunity for the monitoring of continual improvement in the system’s overall performance.
- Promotes efficiency, competency and cost effectiveness.

In the process of following the quality procedures as indicated in the quality manual, the customer / student becomes the hub of all planning and programming of activities of the organization. The customers are treated as the main focus in the organization. The procedure also supports the management because there is definite way of fixing the individual responsibility and time target to accomplish the expected task. The quality procedures guide the administrators that ensures continual improvement in performance and thereby meet the expectations of customers. The regular or the periodical quality audit in the system helps to overcome human slip-ups and system deficiencies through corrective and preventive measures. Above all, the library resources, including grey resources are exploited to the maximum extent through constant quality auditing procedures.

ISO helps the libraries to easily identify and clearly study the system weaknesses and inefficiency of the staff. The aim of extending quality service and achieving customer delight brings revenue to the library and work culture in the organization and finally creates public image which brings appreciation for the staff, who also feel ultimately delighted. The fixing of duties and responsibilities eliminates anxiety from the individual librarian/faculty or management whether or not the part work will be attended to in time or not. Above all, there will be continuous feed-back from the students which helps the management to set the things in right order applying corrective and preventive actions.

BIBLIOGRAPHY AND WEBLIOGRAPHY

1. Beyond ISO 9000 & TQM, Mona Associates, 2002.
2. Bhat, N. S. : Auditing the Quality System. Special Lecture, KSSIDC, Mysore, 2001.
3. Chandrashekara,M and Ramasesh, C.P: Application of standards for quality education and training in LIS. A-LIEP Conference, Japan, 2009.
4. Daleda, Suresh : ISO 9000 Manual for Total Quality Management, 1999.
5. .Implementation and upgradation guide to ISO 9001-2000,Mona Associates,2003
6. Ingle, Sud. : Quality Circles : Master Guide, 2000.
7. ISO 9001 : 2000 : Quality Management Systems Requirements, 2001
8. ISO 9004 : QMS – Guidelines for Performance Improvements, 2001.
9. ISO Products and Servicer – http://www.iso-9000-2000.com
10. ISO Information Center /ISO Tools – http://www.isopro.co-uk
11. ISO News Letter – http://www.iso-9000-2000.com
12. ISO 9000 – RAQ Ball Tek Ltd – Informative Site – http://www.balltek.com
13. QPR Process Guide – http://www.qpronine.com/iso9001
14. Jayanna, M. Information Technology components in LIS curriculum in universities in India, Mysore, University of Mysore (Doctoral Theses), 2009.
15. Quality Assurance Improvement, Training – http://www.balltek.com
16. Radice, Ronald : ISO 9001 Interpreted for Software Organizations, 2001.

Developing small worlds of e-science: using quantum mechanics, biological science, and oceanography for education and outreach strategies for engaging research communities within a university

Plato L. Smith II
Florida State University Libraries, United States

Abstract

The paper provides an introductory and exploratory research method for developing a data management plan strategy through the experimental approach of developing and preserving small worlds of e-science of grey literature from the physics, biological, and oceanographic research disciplines at Florida State University. The paper opted for an experimental approach of digitization, cataloging, and resource discovery in a digital content management system (DigiTool), online public access catalog (OPAC) in Aleph and OCLC WorldCat, and preservation in Florida Digital Archive (FDA), and MetaArchive while introducing connections to information science, organization, and information system theories for the development of digital curation theory through the theory building strategy of Metatriangulation. The paper provides descriptive research insight into exploring the application of information science theories, organizational theory concepts, and information systems models to the emerging field of digital curation in the development of a data management plan that includes defining digital curation, data curation, and digital preservation. Because of the chosen research approach, the research results may lack generalisability. Therefore, researchers are encouraged to further develop research. The paper aims to contribute to the body of knowledge of digital curation through OAIS Reference Model, The DCC Curation Lifecycle Model, and Data Management Model digital curation theory building through the theory building strategy of Metatriangulation. Originality/value – This paper fulfills an identified need to study how digital curation theory building via Metatriangulation can be enabled and mapped to Boyer's Model of Scholarship.

Project Background and Development

The idea for this project started as a result from the development and submission of a poster session proposal for the 31st Annual IATUL (International Association of Scientific and Technology University Libraries) Conference http://blogs.lib.purdue.edu/iatul2010/at Purdue University (West Lafayette, IN) and Chicago, IL, United States from June 20-24, 2010. The theme for the 2010 IATUL conference was "The Evolving World of e-Science: Impact and Implications for Science and Technology Libraries" and the titled of the poster session proposal was "Developing small worlds of e-science: using quantum mechanics, biological science, and oceanography for education and outreach strategies for engaging research communities within a university." The call for papers was received via the ACRL Science & Technology Section Discussion List http://lists.ala.org/wws/info/sts-l on 2009-11-23, poster session proposal submitted on 2009-11-30, and accepted on 2010-01-14.

The IATUL 2010 poster session highlighted the developing digitization work of quantum physicist Dr. Paul A. M. Dirac which included select Dirac digitized works as his 1926 dissertation, relativistic quantum mechanical wave equation formulated by Dirac in 1928 written on blackboard of Dirac's former office in the Keen Building on Florida State University's campus, and mp3 of Dirac's 1979 Lecture "Why We Believe in Einstein Theory" given at Southern Illinois University Carbondale for "Albert Einstein Centennial Celebration" on March 2, 1979 to name a few. The Paul A.M. Dirac material selected and digitized for digital collection development for the IATUL 2010 poster was further developed and extended to this GL 12 article publication to include Dirac's Nobel Prize in Physics, black & white photos of Dirac teaching classes at Florida State University, principles of quantum mechanics, 1933 Solvay Congress lecture "Theory of the Positron", list of recipients for the Lucasian Chair of Mathematics at Cambridge University, and the Paul A.M. Dirac International Scholarly Certificates digital collection which includes certificates from around the world including Great Britain, India, France, Paris, Germany, China, Hungary, Ireland, and USA. In addition to the previously unpublished Dirac materials which is now available online, images of biological silica from FSU Biological Scientist, Dr. A.K.S.K. Prasad biological science research, and faculty research data from several FSU faculty from the department of earth, ocean, and atmospheric are also available online via http://digitool.fcla.edu/R/?LOCAL_BASE=GEN01-FSU01&pds_handle=GUEST FSU Paul A.M. Dirac Science Library and FSU D-Scholarship digital collections, respectively. The content and development of the online digital collections developed for the IATUL 2010 poster and GL12 presentationshttp://2007.ispace.ci.fsu.edu/~psmithii/GL12-2010/GL12_2010-11-05_FSU.pdfwere made possible through the interdepartmental/campus partnerships, cross-disciplines research digital projects,and inter-institutional collaborations between FSU Libraries Special Collections, Dirac Science Library, FSU Libraries Digital Library Center , FSU Department of Biological Science, FSU Department of Earth, Ocean, and Atmospheric Science FSU Cataloging in

Technical Services, Florida Digital Archive (Florida Center for Library Automation), and MetaArchive (Educopia).

Acknowledgements

Special thanks to Giesele Towels, FSU Libraries DLC Library Operations Supervisor, Willa Patterson, FSU Libraries Library Associate, Sharon Schwerzel, former FSU Dirac Science Library Department Head (retired Fall 2010), Dr. Lucy Patrick, former FSU Special Collections Department Head (retired Fall 2009), Dr. A.K.S.K. Prasad, Tamara Weatherholt, FSU Serials/Metadata Librarian, Lydia Motka, Manager of Florida Digital Archive, and FSU College of Communication and Information and science discipline graduate students. The IATUL 2010 poster session that preceded and developed in to the GL12 presentation that led to the ultimate development of this article publication is the fruition of contribution of many individuals.

This article is dedicated to Sharon Schwerzel who presented the IATUL 2010 poster at Purdue University on behalf of FSU and who strongly supported and encouraged GL12 proposal submission that led to the development of this article.

Foreword

"A record if it to be useful to science, must be continuously extended, it must be stored, and above all it must be consulted" (Bush, 1945).

As We May Think (Bush, 1945) concepts of record extension, storage, and consultation are just as significant, if not more, now in 2011 as they were prophetically prolific in 1945. "Science is becoming data-intensive and collaborative" (Seidel, 2010) and the need for "a more comprehensive approach to data policy" (Marrett, 2010) is tantamount for consultation (Otlet, 1903) for maintaining, preserving, and adding value to digital research data throughout its lifecycle (DCC, 2010) not only for scientists seeking NSF funding but for all institutions and organizations concerned with emphasis on community access to data in support of more open sharing of research data (NSF, 2010). NSF's Press Release 10-077 titled "Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans" dated May 10, 2010 following a May 5, 2010 meeting of the National Science Boardhttp://www.nsf.gov/news/news_summ.jsp?cntn_id=116928,is an excellent example of a government funding agency taking an active role in addressing some of the data management issues inherent in the digital universe as addressed by Kroll & Forsman and Zverina, respectively.

> *"Researchers report that they struggle unsuccessfully with storage and management of their burgeoning volume of documents and data sets that they need and that result from their work. While some universities have devised new services to better manage data and other information derived from research, many researchers flounder in a disorganized and rising accumulation of useful findings that may be lost or unavailable when conducting future research" (Kroll & Forsman, 2010).*
>
> *"A recent study by the International Data Corporation (IDC) said that in 2007, the amount of digital data began to exceed the amount of storage to retain it, and will continue to grow faster than storage capacity from here on. The IDC study predicts that by 2011, our "digital universe" – consisting of digitally-based text, video, images, music, etc. – will be 10 times the size it was in 2006." (Zverina, 2008).*

The NSF's 2010 Press Release 10-77 was pivotal in pushing forward the data management plan agenda across institutions, organizations, and research disciplines. The NSF funding data management plan requirement underscores dedication to research, publication, open access, and education (GreyNet, 2011) while also complimenting the need for the management of research data ensuing from the NIH funded research projects impacted from introduction and implementation of The NIH Public Access Policy, Division G, Title II Section 218 of PL 110-161 (Consolidated Appropriations Act, 2008)http://publicaccess.nih.gov/policy.htm.Both of these government-wide initiatives contribute to the concept of curation of data and the emerging field of digital curation. The DCC Curation Lifecycle Model "provides a graphical, high-level overview of the stages required for successful curation and preservation of data" (DCC, 2010).

Introduction

While visiting FSU Dirac Science Library Department Head at the FSU Dirac Science Library on Florida State University campus in Tallahassee, Florida, the discussion of digitizing high-use Dirac primary resource was discussed leading to the digitization of 20 international scholarly certificates and development of the Paul A.M. Dirac International Scholarly Certificates open-access, online digital collections in Fall 2010. The Dirac Science Library Department Head suggested the digitization and development of this collection due to frequent visits and photographs taken of the collection from researchers and scholars from around the world. The Science Library Department noted that science researchers andscholars particularly enjoyed viewing international scholarly certificates that represented their native countries.

Prior to, during, and after the process of digitization of select Dirac materials, the data curation process took shape, developed, and continues today as a result of the application of some of the stages required for curation from The DCC Curation Lifecycle Model. The DCC Curation Lifecycle Model was introduced to the emerging field of digital curation at The 3rd International Digital Curation Conference (IDCC) in 2007 in Washington, DC. However, Clifford Lynch's closing remarks at The 1st International Digital Conference in 2005 in Bath on curation permeates curation, The DCC Curation Lifecycle Model, and data management while also inclusive of the concept of consultation (Otlet, 1903, Bush, 1945).Lynch (2005) describes curation as:

- Curation as a finite process, with handover to preservation as its end point
- Curation as a whole life process, with evolving objects, and
- Curation as managing a growing, living collection

Curation has also been described as including "stewardship, resource management, access, presentation, active care, involves long time and preservation" (Rusbridge, 2006: 2007). Both Lynch and Rusbridge concepts of curation guided the development of the quantum mechanics, biological science, and earth, ocean, and atmospheric digital collections contained in this article. Due to varying degrees of copyright management, faculty contribution participation, access to and processing of "born digital" and "digitized" research data, each one of these three digital collections are in various stages required for successful curation and preservation. Throughout the research process of developing these digital collections, Dr. Kathleen Burnett, Florida State University College of Communication and Information (CCI) faculty, introduced Boyer's Model of Scholarship in LIS 6289 Doctoral Seminar Education Information Studies Spring 2009, and Fall 2010 purchase of the Digital Curation Centre DCC Digital Curation 101 training manual introduced The DCC Curation Lifecycle Model mapping to the OAIS reference model. However, in order to attempt to apply LIS theories to The DCC Curation Lifecycle Model and glean, process, and apply multiple concepts, models, and theories throughout the research process while exploring the emerging field of digital curation in the development of the three digital collections for this project explored the theory-building process of Metatriangulation from the field of organizational theory first explored serendipitously during a doctoral class at UNC-Chapel Hill in 2002. The remainder of this article is separated into three topical sections taken from the 31st IATUL 2010 conference, followed by bibliography and appendixes.

Section I - Understanding the needs, challenges, and opportunities of computational scientific research and the implications for the scientific and technological library community

Collection #1: Quantum Mechanics – Paul A. M. Dirac Collection – Paul A.M. Dirac is recognized by most authorities as one of the greatest theoretical physicists of the 20th century and the father of the field of quantum mechanics. Prior to this project and paper, none of the current digitized Dirac lectures, notes, and 1926 dissertation existed online and/or in digital format. The Dirac is a current, work-in-process digitization project that includes access via FSU Libraries OPAC (Online Public Access Catalog), OCLC WorldCat, and digital content management system via DigiTool. Selected Dirac digitized materials that was once inaccessible to the non-campus community is now open-access available.

- Current plans – continue to digitize Dirac works 2000+ objects in which FSU hold copyrights and secure copyright permissions from the estate, family, publishers, and copyright holders on Dirac materials to create a more diverse, robust, and critical mass Dirac collection.
- Future plans – create Florida Digital Archive (FDA) Metadata Encoding Transmission Standard (METS) eXtensible Markup Language (XML) Submission Information Package (SIP) using the open-source METS creation tool created by the University Florida of all Dirac archival master files for ingestion and preservation in FDA. May also consider MetaArchive as a redundant distributed digital preservation strategy.

Collection # 2: Biological Science - Dr. A.K.S.K Prasad – Dr. A.K.S.K. Prasad is a biological science courtesy faculty member with 29 years experience in the Systematics of micro algae (Botany), Diatom systematics, phytoplankton systematics, Algae Blooms in freshwater and coastal waters. Dr. Prasad continues to work with the FSU Libraries to develop, promote, and preserve select images of biological silica. The images of biological silica digital collection is a growing collection available online in FSU Libraries OPAC, OCLC WorldCat, digital content management system via DigiTool, preserved via FDA and MetaArchive. This collection was developed as demo preservation digital collection to include senior management for FSU Libraries to join FDA (free) and was the first successful FSU Libraries collection preserved via MetaArchive in 2009 even though FSU has been a member of MetaArchive since 2004.

- Accomplishments – assisted faculty member with the completion of recent DEP grant funded project for manual publication. Faculty contributed success to assistance from the Digital Library Center is organizing some of his research data for publication report. First faculty digital collection to be preserved via FDA and MetaArchive. First digital collection to received written copyright permission from faculty granting FSU Libraries Digital Library Center permission to make select research data open access via DigiTool. This digital collection has been highlighted and presented at several national and international conferences whereas the collection once existed on CDs in the faculty member's office.

- Current plans – awaiting return of faculty member from India, increase faculty contribution participation (over 6000+ unprocessed images of biological silica), and active participation of faculty to provide more research data, field notes, and geographical technical, metadata information.
- Future plans – work with faculty with research data from 2year BP funded research of the oil spill on diatoms and biological silica in the Gulf this semester. Extend and expand scope of digital curation of images of biological silica.

Collection # 3: Earth, Ocean, and Atmospheric Science – The Earth, Ocean, and Atmospheric Science digital collection contains select digitized oceanographic technical reports contributed by several faculty member for digitization and institutional repository population. This collection once existed offline in faculty members' offices inaccessible to the public now exists online accessible via FSU Libraries OPAC, OCLC WorldCat, and digital content management system.

- Current plans – reestablish faculty contribution participation with newly merged departments as a recent of last year's layoff of FSU faculty and merging of select departments including oceanography, anthropology, and communication to name a few.
- Future plans – acquire new and more content while applying stages of The Digital Curation Lifecycle Model to the development of newer digital collections.

Section II - Tools for e-science initiatives, including the development of new frameworks, platforms, and systems for discovery, access, management, and preservation of data; strategies for metadata management, repository structures; and digital data curation

- All three digital collections that once resided offline now exist online in a digital content management system with access via OPAC and OCLC WorldCat.
- FDA digital preservation and MetaArchive distributed digital preservation strategy have been applied to collection # 2 with future application to collections #1, #2, and future collections.
- FSU Libraries purchased BePress Digital Commons in Fall 2010 as part of new and improved IR solution to the management of scholarly output including faculty pre-prints, ETDs, grey literature, and research data.
- Fall 2010 DLC begin exploring and using the METS creation tool to create METS XML SIP for FDA ingestion for FDA preservation. This new workflow includes data curation and digital curation in the development of data management of research data, born digital, and digitized content.
- Boyer's Model of Scholarship is being applied to all collections in tandem with Spring 2011 Digital Library graduate course being taught by CCI faculty, Dr. Sanghee Oh. Boyer's Model of Scholarship includes:
    1. Discovery – Build new knowledge through traditional research
    2. Integration – Interpret the use of knowledge across disciplines
    3. Application – Aid society and professionals in addressing problems
    4. Teaching – Study teaching models and practices to achieve optimal learning
- Applied aspects of the Conceptualise, Create or Receive, Appraise & Select, Ingest, Preservation Action, Store, Access, Use, & Reuse, Transform, Curate, Preserve, Preservation Planning, Representation Information, and Description stages from The DCC Curation Lifecycle Model.
- The Submission Information Package (SIP), Archival Information Package (AIP), and Dissemination Information Package (DIP) part of The Open Archival Information System (OAIS) reference model international standard (ISO) 14721:2003 preservation planning &administration via http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/aboutoais.jspmaps to respective required stages in The DCC Curation Lifecycle Model and Boyer's Model of Scholarship whenever scientists/curators/users interface in managing research data from producer (scientist) to data management (curator) to access (discovery) to consumer (teaching).

The development of a digital curation theory building framework is necessary to address the varied approaches in cost/benefit studies, tools, and methodologies focusing on long-lived data (Beagrie, 2010) in keeping research data safe in the emerging field of digital curation. However, despite the continual advancement in research and development within the emerging field of digital curation more work needs to be done on developing a theoretical framework with which to further build and develop the emerging field of digital curation. Cursory review of digital curation literature reveals digital curation and digital preservation terms sometimes being used synonymously and interchangeably. "To use them as synonymous is to ab-use them. This [is] theoretically incorrect and pragmatically dangerous" (Callaos & Callaos, 2002). Schutt (2006) triangulation and Lewis & Grimes (1999) metatriangulation are two research method approaches, which could be employed to explore and contribute towards, work in the development of a theoretical correct and pragmatically sound framework within the emerging field of digital curation. Future research will explore the concept of digital curation metatriangulation (DCM) for the identification and analysis of digital curation exemplars.

Section III - Local, regional, national and international collaborations related to data management

- Local collaborations include FSU Dirac Science Library, Special Collections, Department of Biological Science, Earth, Ocean, and Atmospheric Science
- Regional collaborations include Florida Center for Library Automation (FCLA), Florida Digital Archive (FDA), University of Florida (UF)
- National collaborations include Educopia/MetaArchive
- International collaboration potential with KRDS in the UK initiated 2011-01-05

In conclusion, all the research work, collaborations, digital collections development, digital curation activities, and data management work involved beginning from SPARC 2008, IDCC 2008, ACRL 2009, OR 2009, GL11 conference poster, 31st IATUL 2010 conference poster presentations to the GL12 presentation were used to in one way or another to engage scholars and faculty from 2008 to present as part of outreach strategies that finally contributed to decision for FSU Libraries to purchase BePress Digital Commons in Fall 2010, specifically to increase the visibility, discoverability, and accessibility of scholarly research at FSU. BePress Digital Commons was selected asan improved institutional repository (IR) solution for FSU Libraries for implementation in 2011. With the purchase of this product, senior management recommended the development of an IR implementation charged with developing work flows, policies, procedures, campus partners, and guiding IR content contribution. The IR implementation team in addition to currently working to acquire faculty participation and representation from The Graduate School, Humanities, English, and Science research disciplines has already acquired representation and senior management participation from the following campus partners for 1st quarter of 2011.

- College of Communication and Information (CCI) – includes tenured faculty
- College of Law
- College of Music
- College of Medicine
- Biological Science
- Earth, Ocean, and Atmospheric Sciences
- Undergraduate Honors in the Major
- Dirac Science Library and Goldstein Library

The IR implementation team will help to develop and implement a quality IR program for the display of research data and information. This project paper will be extended to further explore and develop digital curation metatriangulation and the difference between data and information as represented by Bjorn Langefors infological equation of "I = I (D, S, T), where I stands for information, D data, S the recipient prior knowledge as result of the individual's life experience" (Johannesson & Söderström, 2008).

Appendixes

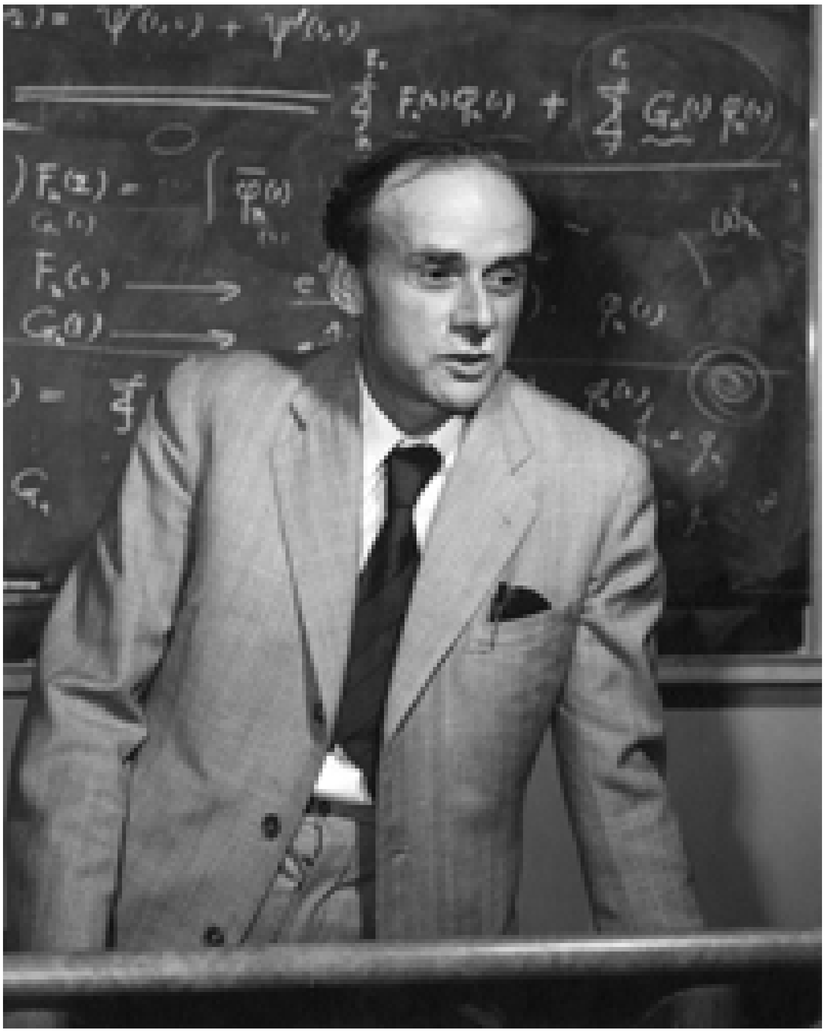
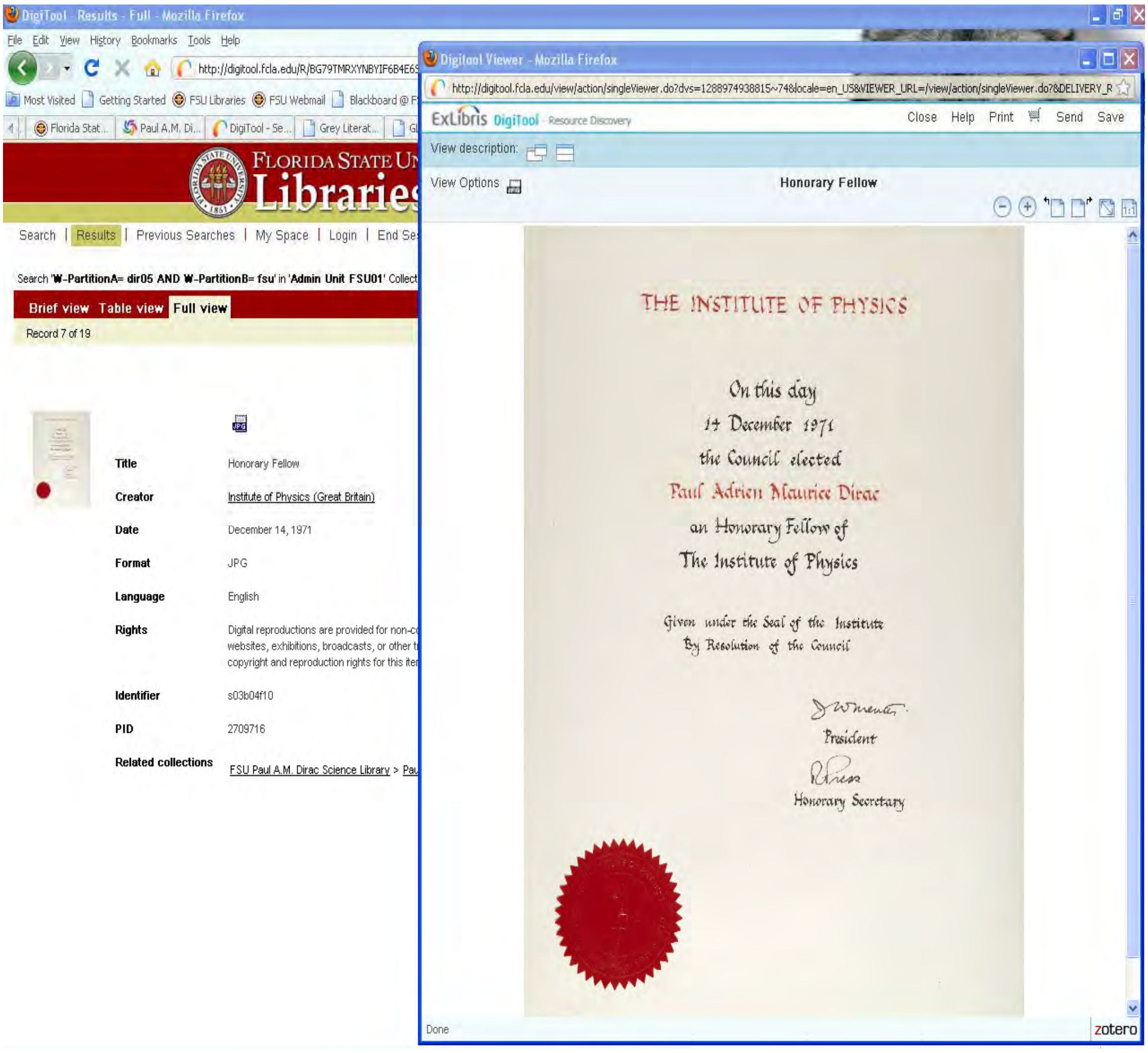


Figure 0: Dr. Paul A.M. Dirac – 1955 Lecture, Ottawa, Canada


DigiTool - Results - Full - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://digitool.fcla.edu/R/BG79TMRXYNBYIF6B4E6
Most Visited Getting Started FSU Libraries FSU Webmail Blackboard @ F
Florida Stat... Paul A.M. Di... DigiTool - Se... Grey Literat...
FLORIDA STATE UN
Libraries
Search | Results | Previous Searches | My Space | Login | End Se
Brief view Table view Full view
Digitool Viewer - Mozilla Firefox
http://digitool.fcla.edu/view/action/singleViewer.do?dvs=1288974938815~74&locale=en_US&VIEWER_URL=/view/action/singleViewer.do?&DELIVERY_R
ExLibris DigiTool - Resource Discovery
Close Help Print Send Save
View description:
View Options
Done
zotero

Search '**W-PartitionA= dir05 AND W-PartitionB= fsu**' in '**Admin Unit FSU01**' Collect

Record 7 of 19

Title	Honorary Fellow
Creator	Institute of Physics (Great Britain)
Date	December 14, 1971
Format	JPG
Language	English
Rights	Digital reproductions are provided for non-c websites, exhibitions, broadcasts, or other t copyright and reproduction rights for this ite
Identifier	s03b04f10
PID	2709716
Related collections	FSU Paul A.M. Dirac Science Library > Pau

Honorary Fellow

THE INSTITUTE OF PHYSICS

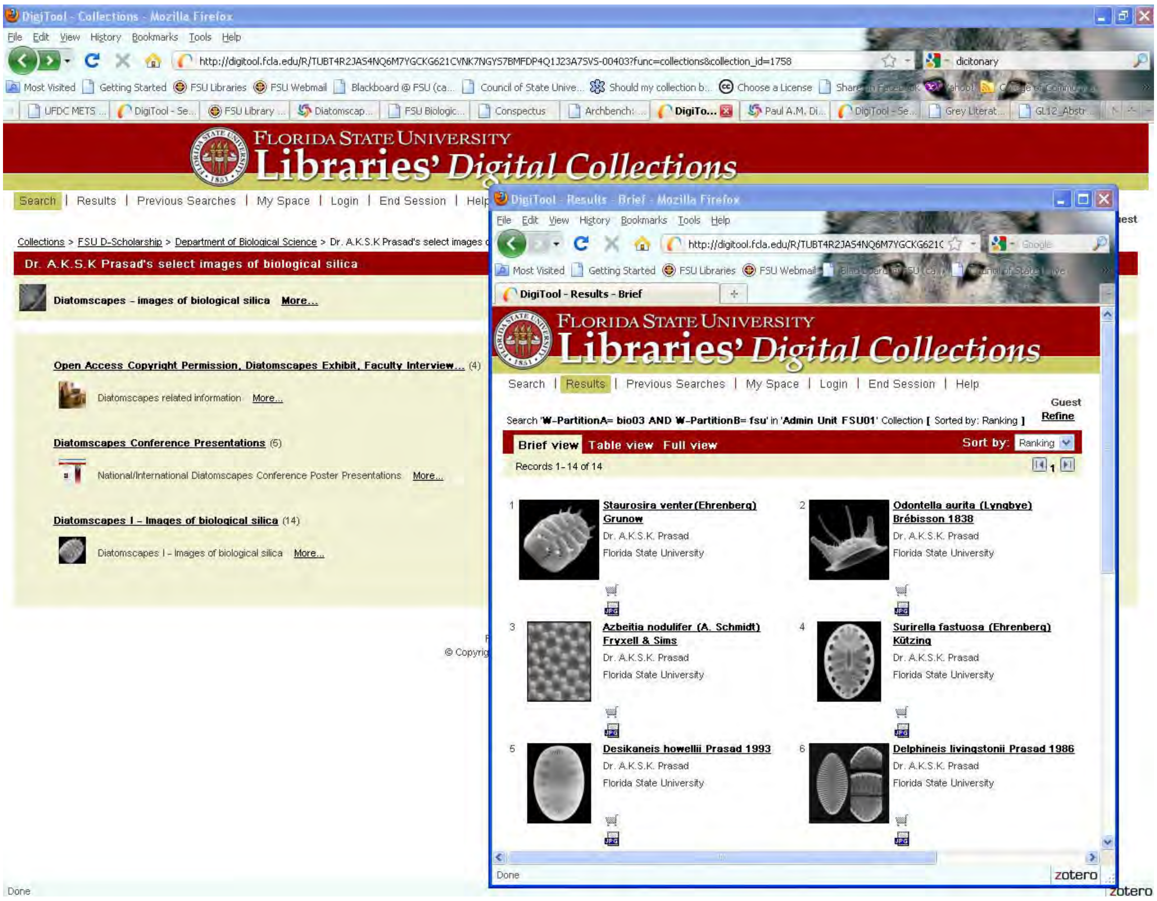
On this day
14 December 1971
the Council elected
Paul Adrien Maurice Dirac
an Honorary Fellow of
The Institute of Physics

Given under the Seal of the Institute
By Resolution of the Council

President

Honorary Secretary

Figure 1: Paul A.M. Dirac International Scholarly Certificates
Institute of Physics (Great Britain) – December 14, 1971


DigiTool - Collections - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Most Visited Getting Started FSU Libraries FSU Webmail Blackboard @ FSU (ca... Council of State Unive... Should my collection b... Choose a License
UFDC METS ... DigiTool - Se... FSU Library ... Diatomscap... FSU Biologic... Conspectus Archbench: DigiTo... Paul A.M. Di... DigiTool - Se... Grey Literat... GL12_Abstr
Search | Results | Previous Searches | My Space | Login | End Session | Help
Done
DigiTool - Results - Brief - Mozilla Firefox
DigiTool - Results - Brief
Brief view Table view Full view
Sort by: Ranking
Done
zotero

http://digitool.fcla.edu/R/TUBT4R2JAS4NQ6M7YGCKG621CVNK7NGYS7BMFDP4Q1J23A7SVS-00403?func=collections&collection_id=1758

FLORIDA STATE UNIVERSITY Libraries' Digital Collections

Collections > FSU D-Scholarship > Department of Biological Science > Dr. A.K.S.K Prasad's select images o

Dr. A.K.S.K Prasad's select images of biological silica

Diatomscapes – images of biological silica More...

Open Access Copyright Permission, Diatomscapes Exhibit, Faculty Interview... (4)

Diatomscapes related information More...

Diatomscapes Conference Presentations (5)

National/International Diatomscapes Conference Poster Presentations More...

Diatomscapes I – Images of biological silica (14)

Diatomscapes I – Images of biological silica More...

http://digitool.fcla.edu/R/TUBT4R2JAS4NQ6M7YGCKG621C

FLORIDA STATE UNIVERSITY Libraries' Digital Collections

Guest Refine

Search 'W-PartitionA= bio03 AND W-PartitionB= fsu' in 'Admin Unit FSU01' Collection [Sorted by: Ranking]

Records 1-14 of 14

1 Staurosira venter(Ehrenberg) Grunow
Dr. A.K.S.K. Prasad
Florida State University

2 Odontella aurita (Lyngbye) Brébisson 1838
Dr. A.K.S.K. Prasad
Florida State University

3 Azbeitia nodulifer (A. Schmidt) Fryxell & Sims
Dr. A.K.S.K. Prasad
Florida State University

4 Surirella fastuosa (Ehrenberg) Kützing
Dr. A.K.S.K. Prasad
Florida State University

5 Desikaneis howellii Prasad 1993
Dr. A.K.S.K. Prasad
Florida State University

6 Delphineis livingstonii Prasad 1986
Dr. A.K.S.K. Prasad
Florida State University

Figure 2: Dr. A.K.S.K. Prasad's select images of biological silica
Diatomscapes I – Images of biological silica

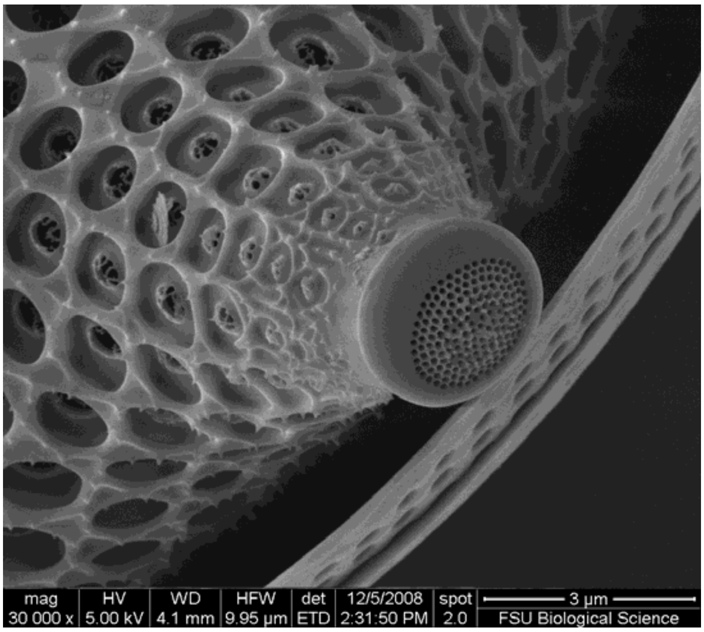


Figure 2 a: Dr. A.K.S.K. Prasad research data image ex # 1

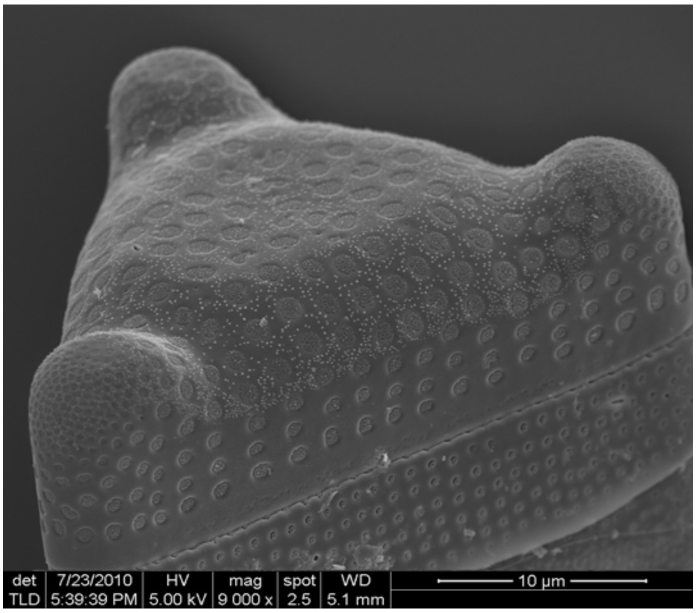
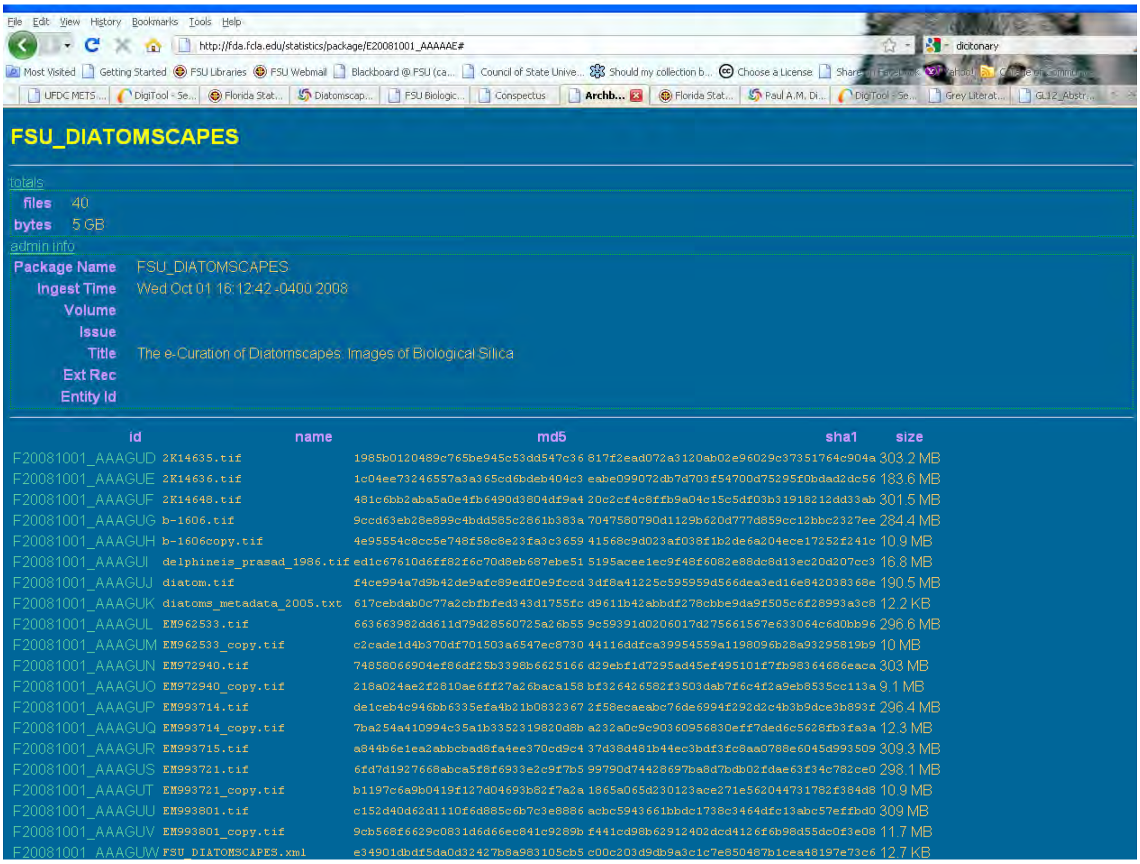


Figure 2 b: Dr. A.K.S.K. research data image ex # 2



FSU_DIATOMSCAPES

totals

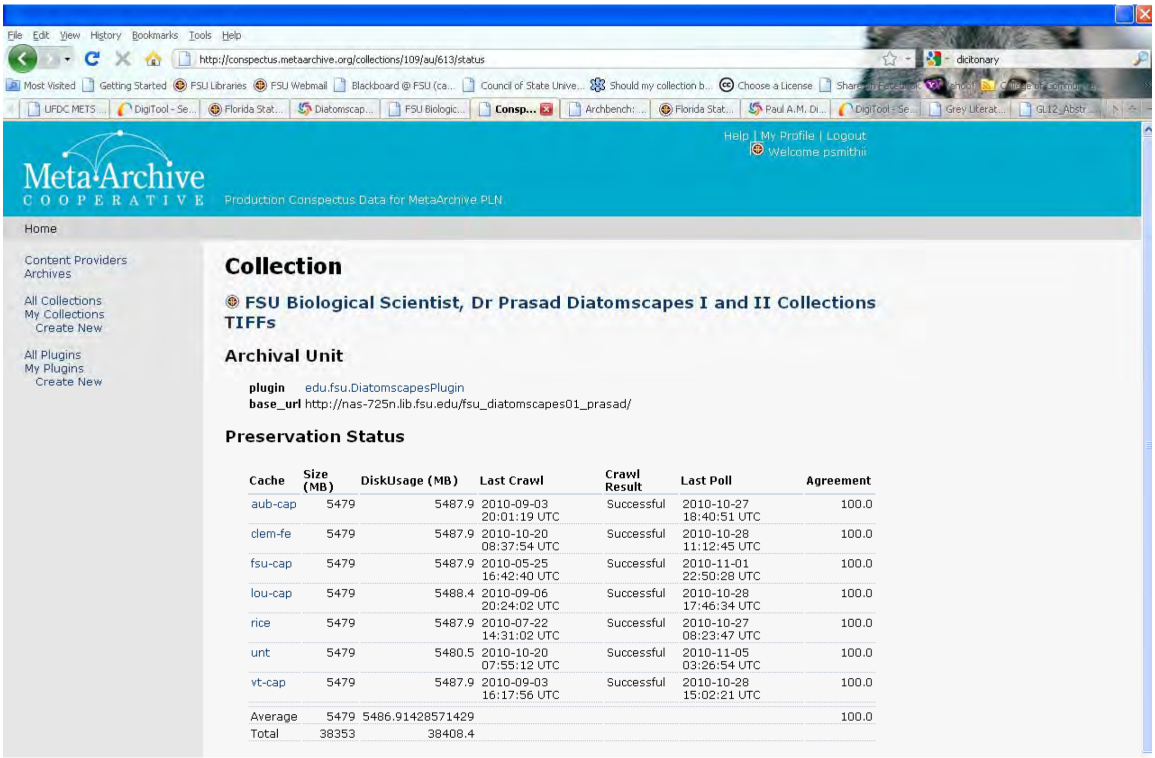
files	40
bytes	5 GB

admin info

Package Name	FSU_DIATOMSCAPES
Ingest Time	Wed Oct 01 16:12:42 -0400 2008
Volume	
Issue	
Title	The e-Curation of Diatomscapes: Images of Biological Silica
Ext Rec	
Entity Id	

id	name	md5	sha1	size
F20081001_AAAGUD	2K14635.tif	1985b0120489c765be945c53dd547c36	817f2ead072a3120ab02e96029c37351764c904a	303.2 MB
F20081001_AAAGUE	2K14636.tif	1c04ee73246557a3a365cd6bdeb404c3	eabe099072db7d703f54700d75295f0bdad2dc56	183.6 MB
F20081001_AAAGUF	2K14648.tif	481c6bb2aba5a0e4fb6490d3804df9a4	20c2cf4c8ffb9a04c15c5df03b31918212dd33ab	301.5 MB
F20081001_AAAGUG	b-1606.tif	9ccd63eb28e899c4bdd585c2861b383a	7047580790d1129b620d777d859cc12bbc2327ee	284.4 MB
F20081001_AAAGUH	b-1606copy.tif	4e95554c8cc5e748f58c8e23fa3c3659	41568c9d023af038f1b2de6a204ece17252f241c	10.9 MB
F20081001_AAAGUI	delphineis_prasad_1986.tif	ed1c67610d6ff82f6c70d8eb687ebe51	5195acee1ec9f48f6082e88dc8d13ec20d207cc3	16.8 MB
F20081001_AAAGUJ	diatom.tif	f4ce994a7d9b42de9afc89edf0e9fccd	3df8a41225c595959d566dea3ed16e842038368e	190.5 MB
F20081001_AAAGUK	diatoms_metadata_2005.txt	617cebdab0c77a2cbfbfed343d1755fc	d9611b42abbdf278cbbe9da9f505c6f28993a3c8	12.2 KB
F20081001_AAAGUL	EM962533.tif	663663982dd611d79d28560725a26b55	9c59391d0206017d275661567e633064c6d0bb96	296.6 MB
F20081001_AAAGUM	EM962533_copy.tif	c2cade1d4b370df701503a6547ec8730	44116ddfca39954559a1198096b28a93295819b9	10 MB
F20081001_AAAGUN	EM972940.tif	74858066904ef86df25b3398b6625166	d29ebf1d7295ad45ef495101f7fb98364686eaca	303 MB
F20081001_AAAGUO	EM972940_copy.tif	218a024ae2f2810ae6ff27a26baca158	bf326426582f3503dab7f6c4f2a9eb8535cc113a	9.1 MB
F20081001_AAAGUP	EM993714.tif	de1ceb4c946bb6335efa4b21b0832367	2f58ecaeabc76de6994f292d2c4b3b9dce3b893f	296.4 MB
F20081001_AAAGUQ	EM993714_copy.tif	7ba254a410994c35a1b3352319820d8b	a232a0c9c90360956830eff7ded6c5628fb3fa3a	12.3 MB
F20081001_AAAGUR	EM993715.tif	a844b6e1ea2abbcbad8fa4ee370cd9c4	37d38d481b44ec3bdf3fc8aa0788e6045d993509	309.3 MB
F20081001_AAAGUS	EM993721.tif	6fd7d1927668abca5f8f6933e2c9f7b5	99790d74428697ba8d7bdb02fdae63f34c782ce0	298.1 MB
F20081001_AAAGUT	EM993721_copy.tif	b1197c6a9b0419f127d04693b82f7a2a	1865a065d230123ace271e562044731782f384d8	10.9 MB
F20081001_AAAGUU	EM993801.tif	c152d40d62d1110f6d885c6b7c3e8886	acbc5943661bbdc1738c3464dfc13abc57effbd0	309 MB
F20081001_AAAGUV	EM993801_copy.tif	9cb568f6629c0831d6d66ec841c9289b	f441cd98b62912402dcd4126f6b98d55dc0f3e08	11.7 MB
F20081001_AAAGUW	FSU_DIATOMSCAPES.xml	e34901dbdf5da0d32427b8a983105cb5	c00c203d9db9a3c1c7e850487b1cea48197e73c6	12.7 KB

Figure 3: Florida Digital Archive (FDA) preservation
FSU_DIATOMSCAPES preserved via FDA with online preservation reporting



http://conspectus.metaarchive.org/collections/109/au/613/status

Production Conspectus Data for MetaArchive PLN

Collection

FSU Biological Scientist, Dr Prasad Diatomscapes I and II Collections TIFFs

Archival Unit

plugin edu.fsu.DiatomscapesPlugin
base_url http://nas-725n.lib.fsu.edu/fsu_diatomscapes01_prasad/

Preservation Status

Cache	Size (MB)	DiskUsage (MB)	Last Crawl	Crawl Result	Last Poll	Agreement
aub-cap	5479	5487.9	2010-09-03 20:01:19 UTC	Successful	2010-10-27 18:40:51 UTC	100.0
clem-fe	5479	5487.9	2010-10-20 08:37:54 UTC	Successful	2010-10-28 11:12:45 UTC	100.0
fsu-cap	5479	5487.9	2010-05-25 16:42:40 UTC	Successful	2010-11-01 22:50:28 UTC	100.0
lou-cap	5479	5488.4	2010-09-06 20:24:02 UTC	Successful	2010-10-28 17:46:34 UTC	100.0
rice	5479	5487.9	2010-07-22 14:31:02 UTC	Successful	2010-10-27 08:23:47 UTC	100.0
unt	5479	5480.5	2010-10-20 07:55:12 UTC	Successful	2010-11-05 03:26:54 UTC	100.0
vt-cap	5479	5487.9	2010-09-03 16:17:56 UTC	Successful	2010-10-28 15:02:21 UTC	100.0
Average	5479	5486.91428571429				100.0
Total	38353	38408.4				

Figure 4: MetaArchive preservation
FSU Biological Scientist, Dr. Prasad Diatomscapes I & II Collections (TIFFs) being preserved by Auburn University, Clemson University, Florida State University, Rice University, University of North Texas, and Virginia Tech LOCKSS caches

Bibliography

Beagrie, C. (2010). Keeping research data safe. Cost/benefit studies, tools, and methodologiesfocusing on long-lived data. Retrieved December 21, 2010 via http://www.beagrie.com/krds.php.

Bush, V. (1945). As we may think. The Atlantic (July), 101-108. Retrieved November 10, 2010 viahttp://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881/.

Callaos, N. & Callaos, B. (2002). Toward a systematic notion of information: practical consequences. Informing Science. 5 (1).

DCC. (2010). What is digital curation? Retrieved January 14, 2011 via http://www.dcc.ac.uk/digital-curation/what-digital-curation.

Johannesson, P., & Söderström, E. (2008). Information systems engineering: from data analysis to process networks. IGI Publishing, p. vi.

Kroll, S. & Forsman, R. (2010). A slice of research life: information support for research in the United States. Retrieved October 28, 2010 from http://www.oclc.org/research/publications/library/2010/2010-15.pdf.

NSF, Seidel, E., &Marrett, C. (2010). Scientists seeking NSF funding will soon be required to submit data management plans. Retrieved January 14, 2011 from http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928.

Otlet, P. (1903). The science of bibliography and documentation. In W. B. Rayward (Ed. and Trans.) (1990). The international organization and dissemination of knowledge: Selected essays of Paul Otlet. Amsterdam: Elsevier.

*Rusbridge. (2007). Tomorrow, and tomorrow, and tomorrow: poor players on the digital curation stage. Retrieved January 14, 2011 via http://www.era.lib.ed.ac.uk/handle/1842/2150. (error in online pub date)

Smith, P. Schwerzel, S., & Weatherholt, T. (2010). Conference poster presentation. 31st Annual IATUL 2010 Conference. Retrieved January 14, 2011 via http://2007.ispace.ci.fsu.edu/~psmithii/iatul2010/IATUL-FSU_2010-06-17.pdf.

Zverina, J. (2008). The economics of data preservation: international blue ribbon taskforce issues interim report on economic issues brought on by "data deluge" in the information age. Retrieved July 22, 2009 from http://ucsdnews.ucsd.edu/newsrel/general/12-08BRTF.asp.

Appendices

Author Information

Asserson, Anne 95
Anne Asserson holds a Cand. polit. with a Masters in Information Science from the University of Bergen, UiB. She has been working with Research Documentation, and has participated in substantial parts of CRIS developmental work, locally and nationally. Anne Asserson has been part of the establishing and implementing of a Research Documentation system, Fdok http://www.ub.uib.no/fdok/sok/, at the UiB. For several years she was the chairwoman of the Steering Group of the national CRIS system and project secretary of a National system for academic administration. Anne Asserson is presently representing UiB in the national group that is implementing a new national research documentation system, FRIDA. She has also participated in The CORDIS funded European-wide project on " Best Practice" 1996 . She was a member of the working group set up 1997 that produced the report CERIF2000 Guidelines (1999) www.cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS is now the custodian of the CERIF model www.eurocris.org. Anne Asserson is a member of the Best Practice Task Group. anne.asserson@fa.uib.no

Candlish, J.R. 83
J.R. Candlish is a Technical Assistant and GIS Analyst at Information International Associates, Inc. (IIa). He received his M.S. in Biosystems Engineering Technology from the University of Tennessee, Knoxville and his B.S. in Environmental Studies: Natural Resources from Sewanee: The University of the South. At IIa he provides secretariat services to CENDI agencies, the federal scientific information manager's group and also fulfills a mapping role for the USGS National Biological Information Infrastructure's Southern Appalachian Information Node (SAIN). jcandlish@iiaweb.com

Carroll, Bonnie C. 83
Bonnie C. Carroll, President of Information International Associates, Inc. (IIa), supports government and industry in managing information as a strategic resource. She is Secretariat Director of CENDI, the federal scientific and technical information (STI) managers' group and Consultant to USGS, supporting the development of the National Biological Information Infrastructure and the OSTP Biodiversity Informatics Working Group to promote interagency coordination of national and international biodiversity informatics initiatives. Ms. Carroll is currently the US National Representation to International CODATA under the auspices of the International Committee for Science (ICSU) and is on the editorial board for the journal Information Services and Use. She has an M.S. from Columbia University and a B.A. from Cornell University. Ms. Carroll has been active in grey literature research and management dating back to the early 1990s.
Email: bcarroll@iiaweb.com

Chowdappa, N. 101, 118
N. Chowdappa Obtained a Post-Graduate Degree in Geo-science and Master's Degree in Library and Information Science from the Bangalore University, Bangalore with a distinction. Submitted Doctoral Thesis on "organization and use patterns of grey literature in engineering research institutions" to the University of Mysore, Mysore for the award of Doctoral Degree in Library and Information Science. Serving as Chief Librarian at BMS College of Engineering, Bangalore for the last 26 years. Special interest in promoting the use of grey literature among faculty and researchers in engineering discipline. Responsible for the organization of 22 professional and extension programmes for Teachers and Librarians in Science and Technology. Served as resource person in library and information science, and Academic Counselor for Indira Gandhi National Open University, New Delhi. Presently holding the position of President , AKELPA - All Karnataka Engineering College Library Professionals Association, Bangalore. Areas of professional interest are : Scholarly Communications, Facilitating Research Programmes and Reference Service. Email: ncbmsce@yahoo.co.in

Crowe, June 83
June Crowe is the Technical Director, Intelligence Division at Information International Associates, Inc. (IIa). She received her AMLS from the University of Michigan, Ann Arbor and her M.Ed. in geographic education from the University of Georgia, Athens. She has extensive experience in the management and operations of library services across government, public, academic, and special libraries. At IIa she manages the open source research division which focuses on medical, socio-cultural, science and technology and business research. Her primary interests are open source information in Grey Literature, digital repositories, and open source intelligence. Email: jcrowe@iiaweb.com

Di Cesare, Rosa 72
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). r.dicesare@irpps.cnr.it

Farace, Dominic J. 32
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 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. info@textrelease.com

Gonda, Mayuki 113
Mayuki Gonda works as librarian at the Central Library of JAEA (Japan Atomic Energy Agency). He joined JAERI (former JAEA) in 2005, and had been working for management and dissemination of JAEA research results information. Since 2009, he is in charge of selection, classification and indexing for INIS (International Nuclear Information System) Database. He is also a member of editorial committee of the Journal of Information Science and Technology Association (INFOSTA) since 2008. He holds a degree in information science (M.A.) from the Graduate School of Library, Information and Media Studies, University of Tsukuba. Email: gonda.mayuki@jaea.go.jp

Author Information *(continued)*

Henrot, Nathalie **59**
Hélène Prost works since 1995 as a librarian at INIST-CNRS and is specialised in the evaluation of collections and document supply. Actually she is preparing an interface in order to compare different databases. She obtained a Master in History in 1991 and a Master in Scientific Information in 1993. She published on statistical and bibliometric analysis of information.
Email: helene.prost@inist.fr

Ikeda, Kiyoshi **113**
Kiyoshi Ikeda works as librarian at the Central Library of Japan Atomic Energy Agency (JAEA). He joined JAERI (former JAEA) in 2004, and had been working for acquisition of library materials (domestic/foreign books, journals and technical reports). Since 2007, he is in charge of editing work at "JAEA Reports" and "JAEA R&D Review" which are grey literatures. He is also a member of editorial board of the Japan Special Library Association (JSLA) since 2008. He holds a degree in library and information science (M.A.) from the Graduate School of University of Library and Information Science (in Japan).
Email: ikeda.kiyoshi@jaea.go.jp

Jeffery, Keith G. **95**
Keith Jeffery is currently Director, IT and International Strategy of STFC (Science and Technology Facilities Council), based at Rutherford Appleton Laboratory in UK. Previously he was Head of Business and Information Technology Department with a staff of 140 supporting over 360000 users, developing software for business and science and doing leading edge R&D. STFC hosts the UK and Ireland Office of W3C and develops and supports the largest OA (Open Access) institutional repository in UK. Keith is a Fellow of both the Geological Society of London and the British Computer Society. He is a Chartered Engineer. He is an Honorary Fellow of the Irish Computer Society. He is president of euroCRIS (www.eurocris.org) and of ERCIM (www.ercim.org) and holds three honorary professorships. He has extensive publications and has served on numerous programme committees and research grant review panels. He has particular interests in 'the research process' and the relationship of hypotheses, experiments, primary data and publications based on research in information systems, knowledge-based systems and metadata.
Email: k.g.jeffery@rl.ac.uk

Luzi, Daniela **72**
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 **27**
Claudia Marzi graduated in Foreign Languages and Literature at Pisa University in 1998, with the dissertation "The power of words: language creativity in Edgar Allan Poe's narrative". Since 2001 she has been working at the CNR Institute for Computational Linguistics in Pisa. She contributed to many national (funded by Italian Ministry of Research) and international (funded by European Commission) projects. Principal organizer of the European Science Foundation Exploratory Workshop "Words in action: Interdisciplinary approaches to understanding word processing and storage" (Pisa, Oct. 2009); coordinator of a ESF Research Networking Programme proposal (call 2009) on "The European Network on Word Structure. Cross-disciplinary approaches to understanding word structure in the languages of Europe". Member of ILC scientific board at Institute for Computational Linguistics. Email: claudia.marzi@ilc.cnr.it

Mynarz, Jindřich **65**
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

Myohanen, Lea **49**
Lea Myohanen (BS, MS, MSILS) has worked as a reference librarian at the New York Academy of Medicine library since 1999. She is the library liaison to the Divisions of Health & Science Policy and Public Health and to the Center for Urban Epidemiological Studies. Email: lmyohanen@nyam.org

Pardelli, Gabriella **27**
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

Pejšová, Petra **39**
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

Ramasesh, C.P. **101, 118**
C.P. Ramasesh obtained Master's Degree in Philosophy from the University of Mysore in 1976 with specialization in Vedanta Philosophy. Also obtained Master's Degree (1978) and Doctorate Degree (1989) in Library and Information Science from the University of Mysore. Recipient of Dr. S.R. Ranganathan Memorial Gold Medal for securing first rank in the Master's Degree. Served as professional librarian for three decades and responsible for organization of several

Author Information *(continued)*

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 Deputy Librarian at the University Graduate Library, University of Mysore and supervising research programmes in the field of 1) Grey Literature and 2) Online Journals. Email : cpramasesh@gmail.com

Ricci, Marta **72**
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 Grey Literature.
Email: biblio.irpps@irpps.cnr.it

Ruggieri. Roberta **72**
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 the National Research Council (CNR) in research activities related to the field of Grey literature and Institutional repositories.
Email: biblio.irpps@irpps.cnr.it

Sassi, Manuela **27**
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 **11**
Joachim Schöpfel is head of the department of information and communication 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

Searing, Susan E. **52**
Susan E. Searing is an associate professor of library administration at the University of Illinois, Urbana-Champaign, where she is in charge of services and collections for the field of library & information science and is affiliated with the Graduate School of Library and Information Science. Sue previously worked for the University of Wisconsin-Madison as Deputy Director and Associate Director for Public Services; for the multi-campus University of Wisconsin System as Women's Studies Librarian-at-Large; and for Yale University as a reference librarian. Her current research interests include library services to distance education students, embedded librarianship, and reference sources in women's studies. Email: searing@illinois.edu

Škuta, Ctibor **65**
Ctibor Škuta is working in the Department of Polythemathic Structured Subject Heading System (PSH) at the National Technical Library, Czech Republic. His tasks mainly involve the automation of processes related to the administration of PSH and cooperation on other projects with the Development of Electronic Services Department. At the same time, Škuta is studying Applied Informatics in Chemistry at The Institute of Chemical Technology in Prague. His professional interests are programming (Python, Java, XML technologies), data mining, and semantic web.
Email: ctibor.skuta@techlib.cz

Smith II, Plato L. **125**
Plato L. Smith II is currently Head of the Digital Library Center Department at Florida State University with rank of Associate University Librarian, my career in research libraries includes over 7 years total Association of Research Libraries experience with 5 years of increasing managerial responsibilities at Florida State University. As the Digital Library Center Department Head at Florida State University since June 2005, I manage major aspects of library's digital collections development, institutional repository, digital assets management system, oversee scanning and digitization projects, and serve as FSU's statewide digital initiatives subcommittee (DISC) representative, Council of State University Libraries DISC Chair, and technical expert and Co-PI on the grant funded MetaArchive project (LOCKSS), a multiinstitutional digital preservation partnership funded by the Library of Congress NDIIPP program. Working closely with special collections, technical services, faculty, librarians, and systems staff, some projects have included digitization of electronic theses and dissertations, special collections materials, and various intellectual output of the University community to provide value to the Libraries and the University. Email: psmithii@fsu.edu

Stock, Christiane **59**
Christiane Stock is the Head of the Monographs and Grey Literature service at INIST, in charge of the repositories LARA (reports), mémSIC (master's theses in information sciences) and OpenSIGLE. Member of the Technical Committee for the SIGLE database from 1993 to 2005, she also set up the national agency for ISRN (International Standard Report Number). She is member of the AFNOR expert group who prepared the recommended metadata scheme for French electronic theses (TEF).
Email: christiane.stock@inist.fr

Vaska, Marcus **39**
Marcus Vaska is a librarian for the Physician Learning Program (PLP), a new collaborative initiative between the Universities of Alberta and Calgary, funded via an Alberta Medical Association (AMA) trilateral agreement. Marcus is responsible for assisting physicians in their research, and addressing their perceived and unperceived learning needs. Prior to this 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

List of Participating Organizations

Arizona State University, ASU	United States
BMS College of Engineering	India
British Embassy Prague	Czech Republic
British Library, BL	United Kingdom
Central Technical University, CVUT	Czech Republic
Centre National de Recherché Scientifique, CNRS	France
Charles University	Czech Republic
Dalhousie University, DAL	Canada
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
Food and Agriculture Organization of the United Nations	Italy
Grey Literature Network Service, GreyNet	Netherlands
Information International Associates, IIa	United States
Institut de l'Information Scientifique et Technique, INIST	France
Institute for Chemistry of Organometallic Compounds, ICCOM	Italy
Institute of Information Science and Technologies; CNR	Italy
Institute of Information Studies and Librarianship	Czech Republic
Institute of Research on Population and Social Policies	Italy
Int. Council for Scientific & Technical Information, ICSTI	France
Istituto di Linguistica Computazionale, ILC	Italy
Japan Atomic Energy Agency, JAEA	Japan
Korea Institute of Science & Technology Information, KISTI	Korea
Law Library of Congress, LLC	United States
Library of Congress, LC	United States
Library of the Academy of Sciences	Czech Republic
Masaryk University	Czech Republic
Ministry of Culture	Czech Republic
Multidata	Czech Republic
National Medical Library, NLK	Czech Republic
National Research Council, CNR	Italy
National Technical Library, NTK	Czech Republic
New York Academy of Medicine, NYAM	United States
Oklahoma State University	United States
PricewaterhouseCoopers, PwC	Netherlands
Science and Technology Facilities Council, STFC	United Kingdom
Slovak Centre of Scientific and Technical Information	Slovak Republic
Tempere University of Technology Library	Finland
Texas A&M University	United States
TextRelease	Netherlands
Total Support Systems Corporation, TOSS	Japan
Université Charles de Gaulle Lille 1	France
Université Charles de Gaulle Lille 3	France
University of Bergen, UiB	Norway
University of Calgary	Canada
University of Economics Prague	Czech Republic
University of Illinois, UIUC	United States
University of Ljubljana, UNI-LJ	Slovenia
University of Mysore	India

GreyNet Members
less 20%

Transparency in Grey Literature

Publication Order Form

TWELFTH INTERNATIONAL CONFERENCE ON GREY LITERATURE *Prague, Czech Republic 6-7 December 2010*

Forthcoming February 2011	*No. of Copies*	*x*	*Amount in Euros*	Subtotal
GL12 CONFERENCE PROCEEDINGS – Printed Edition ISBN 978-90-77484-16-6 ISSN 1386-2316 *Postage and Handling excluded*[*)]		x	99.00 =	€
GL12 CONFERENCE PROCEEDINGS - CD-Rom Edition ISBN 978-90-77484-16-6 ISSN 1386-2316 *Postage and Handling included*		x	99.00 =	€
EXCLUSIVE OFFER: GL12 Conference Proceedings - CD-Rom Edition With accompanying PowerPoint Presentations *Postage and Handling included*		x	119.00 =	€
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 "GL12 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

GL12 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

GL13
The Grey Circuit

From Social Networking to Wealth Creation

Thirteenth International Conference on Grey Literature

Hosted by the Federal Library & Information Center Committee (FLICC)
Library of Congress
Washington, D.C., USA

December 5-6, 2011

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

For the latest news on GL13 or FLICC/FEDLINK, visit our Web site at http://www.loc.gov/flicc.