

An International Journal on Grey Literature



Spring 2011 – TGJ Volume 7, Number 1

‘TRANSPARENCY IN GREY LITERATURE’

GreyNet

The Grey Journal

An International Journal on Grey Literature

COLOPHON

Journal Editor:

Dr. Dominic J. Farace
Grey Literature Network Service
GreyNet, The Netherlands
journal@grey-net.org

Associate Editors:

Julia Gelfand
University of California, Irvine
UCI, United States

Dr. Debbie L. Rabina
School of Information and Library Science
Pratt Institute, United States

Dr. Joachim Schöpfel
Université Charles de Gaulle Lille 3
France

Gretta E. Siegel
Portland State University
PSU, United States

Kate Wittenberg
Project Director, Client and Partnership
Development at Ithaka, United States

Technical Editor:

Jerry Frantzen, TextRelease

CIP

The Grey Journal (TGJ): An international journal on grey literature / D.J. Farace (Journal Editor); J. Frantzen (Technical Editor) ; GreyNet, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 7, Number 1, Spring 2011. - *The British Library, FLICC-FEDLINK, INIST-CNRS, NTK, and NYAM are corporate authors and associate members of GreyNet International. This serial publication appears three times a year - in spring, Autumn, and autumn. Each issue is thematic and deals with one or more related topics in the field of grey literature. The Grey Journal appears both in print and electronic formats.*
ISSN 1574-1796 (Print)
ISSN 1574-180X (E-Print/PDF)
ISSN 1574-9320 (CD-Rom)

Subscription Rates:

€125 individual, including postage & handling
€240 institutional, including postage & handling

Contact Address:

Reprint Service, Back Issues, Document Delivery,
Advertising, Inserts, Subscriptions:

TextRelease
Javastraat 194-HS
1095 CP Amsterdam
Netherlands
T/F +31 (0) 20 331.2420
info@textrelease.com
<http://www.textrelease.com/gjpublications.html>

About TGJ

The Grey Journal is a flagship journal for the grey literature community. It crosses continents, disciplines, and sectors both public and private. The Grey Journal not only deals with the topic of grey literature but is itself a document type classified as grey literature. It is akin to other grey serial publications, such as conference proceedings, reports, working papers, etc.



The Grey Journal is geared to Colleges and Schools of Library and Information Studies, as well as, information professionals, who produce, publish, process, manage, disseminate, and use grey literature e.g. researchers, editors, librarians, documentalists, archivists, journalists, intermediaries, etc.

About GreyNet

The Grey Literature Network Services was established in order to facilitate dialog, research, and communication between persons and organizations in the field of grey literature. GreyNet further seeks to identify and distribute information on and about grey literature in networked environments. Its main activities include the International Conference Series on Grey Literature, the creation and maintenance of web-based resources, a moderated Listserv, and The Grey Journal. GreyNet is also engaged in the development of distance learning courses for graduate and post-graduate students, as well as workshops and seminars for practitioners.

Full-Text License Agreement

In 2004, TextRelease entered into an electronic licensing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The full text of articles in The Grey Journal (TGJ) can be found in *Library, Information Science & Technology Abstracts* (LISTA) full-text database.

© 2011 TextRelease

Copyright, all rights reserved. No part of this publication may be reproduced, stored in or introduced into a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior permission of the publisher.

Contents

"TRANSPARENCY IN GREY LITERATURE"

Towards a Prague Definition of Grey Literature	5
<i>Joachim Schöpfel (France)</i>	
A Terminology Based Re-Definition of Grey Literature	19
<i>Claudia Marzi, Gabriella Pardelli, and Manuela Sassi (Italy)</i>	
Peering through the Review Process: Towards Transparency in Grey Literature	25
<i>Dominic J. Farace (The Netherlands)</i>	
Scientific Data: Increasing Transparency and Reducing the Grey	33
<i>Bonnie C. Carroll, June Crowe, and J.R. Candlish (United States)</i>	
Global Grey Literature in Health: Identification, Preservation, and Sustainability	41
<i>Janie Kaplan, Lea Myohanen, and Elizabeth Taylor (United States)</i>	
A Bridge over Grey Waters: Collecting and Preserving Public Policy Research	45
<i>Amanda Lawrence (Australia)</i>	

Colophon.....	2
Editor's Note.....	4
Advertisement	
De Gruyter Saur, <i>Grey Literature in Library and Information Studies</i>	24
IIA, Information International Associates, Inc.	32
NYAM, The New York Academy of Medicine.....	40
EBSCO Information Services.....	55
On the News Front	
Thirteenth International Conference on Grey Literature – Call for Papers.....	56
The Grey Circuit, From Social Networking to Wealth Creation – Registration.....	57
About the Authors.....	58
Notes for Contributors.....	59

EDITOR'S NOTE

In the 21st century, social networking is the way the grey literature community remains connected. It encompasses a range of social media and communication tools that enable subject based communities to create, review, process, publish, and make grey literature openly accessible to public domain. Social networking is not new to grey literature; in fact it is inherent to this field of information. What's new however are the technologies available to global grey literature communities in developing, monitoring, and sustaining valued information resources and services. In this context, social networking becomes a mechanism both used and applied by grey literature communities in the processes of knowledge generation and ensuing wealth creation. The Thirteenth International Conference on Grey Literature is beholden to its title: The Grey Circuit, from Social Networking to Wealth Creation. The imagery in the conference logo rekindles the spirit of Franklin in which the uncontrolled discharge of lightening transfers power to controlled networked circuits. Today, in the spirit of science, grey literature communities are called to demonstrate their know-how and merit to wider audiences by responding to the GL13 Call-for-Papers.

Earlier this year, GreyNet established an open LinkedIn Group, in order to promote discussion and dialogue among colleagues in the international grey literature community. More than 120 members have now signed-on and you are likewise encouraged to do so. For those authors and researchers, who have already submitted abstracts to the GL13 Call-for-Papers or who intend to do so, you have the distinct advantage of using the GreyNet LinkedIn Group for purposes of open discussion and relevant feedback well in advance of the Thirteenth International Conference on Grey Literature.

Further with this issue of The Grey Journal, I am pleased to welcome on board a new associate editor, Dr. Debbie L. Rabina of Pratt Institute, School of Information and Library Science. Debbie is widely known for her active involvement of LIS students in education and research in the field of grey literature. She was the recipient of the GreyNet Award in 2009 and authored a chapter in 2010 titled "Course and Learning Objective in the Teaching of Grey Literature".

Dominic Farace, Journal Editor
journal@greynet.org

Grey Literature is a field in library and Information science that deals with the production, distribution, and access to multiple document types produced on all levels of government, academics, business, and organization in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body.

Towards a Prague Definition of Grey Literature*

Joachim Schöpfel (France)

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

* First published in the GL12 Conference Proceedings, February 2011

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* website¹ (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 database². 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 online³. 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).

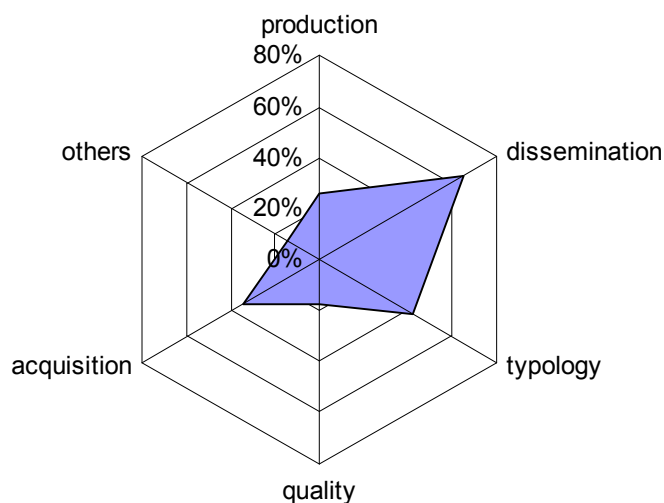


Figure 3: Main topics of GL definitions in corpus

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 uniqueness⁴.

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

Intellectual property: Following De Blaaij (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).

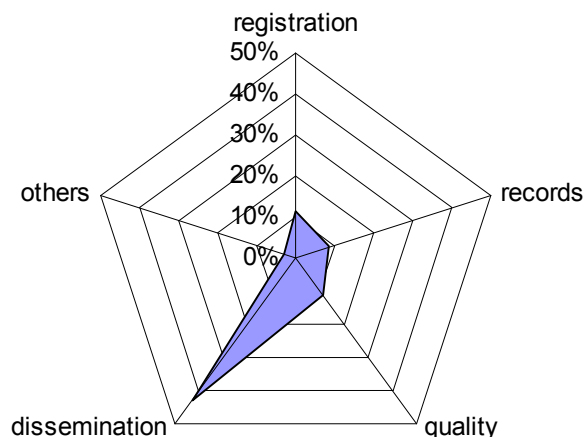


Figure 4: Main functions of GL in corpus

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 conditions⁵. 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).

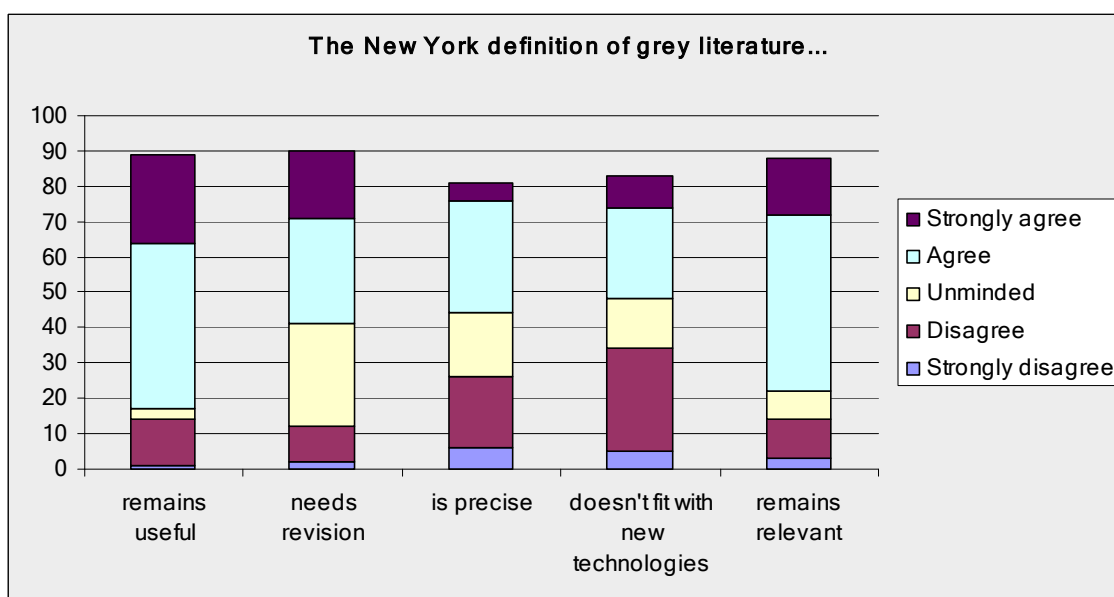


Figure 5: Opinions on the current NY definition

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).

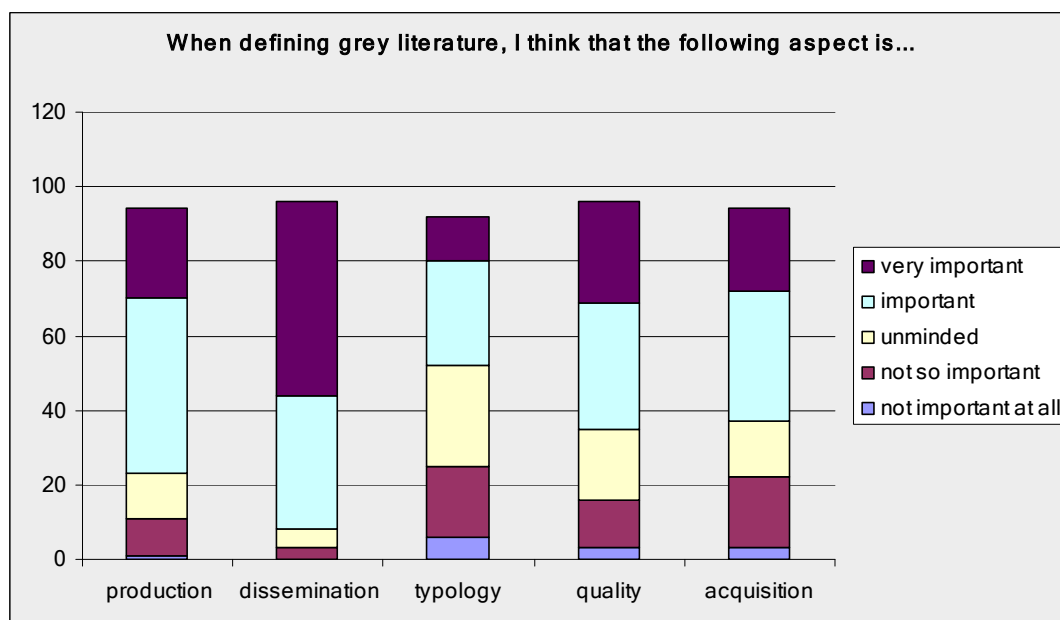


Figure 6: Opinions on essential attributes of a GL definition

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).

	<i>important</i>	<i>not important</i>
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).

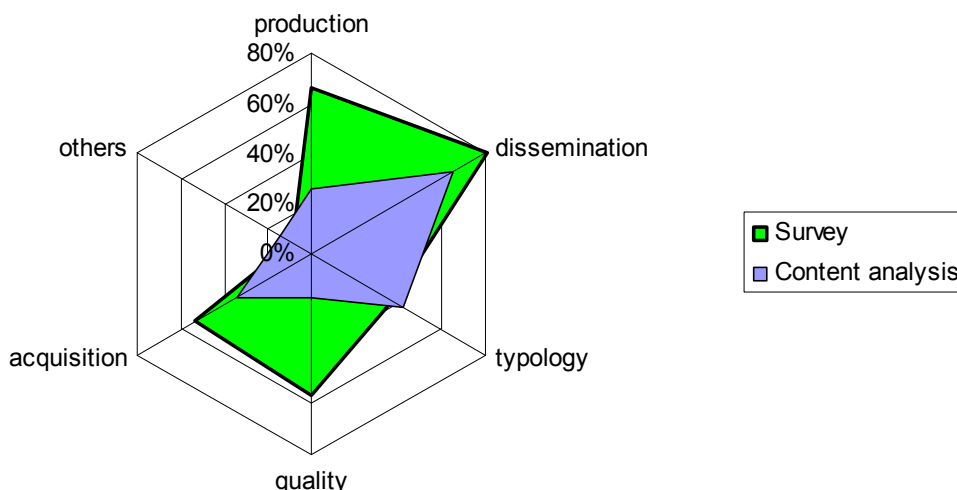


Figure 8: Main topics of GL definitions in corpus (content analysis) and survey
Figure 8 shows clearly that the economic approach to grey literature still prevails, e.g., professionals, scientists and scholars highlight dissemination and production as essential attributes when defining grey literature. 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).

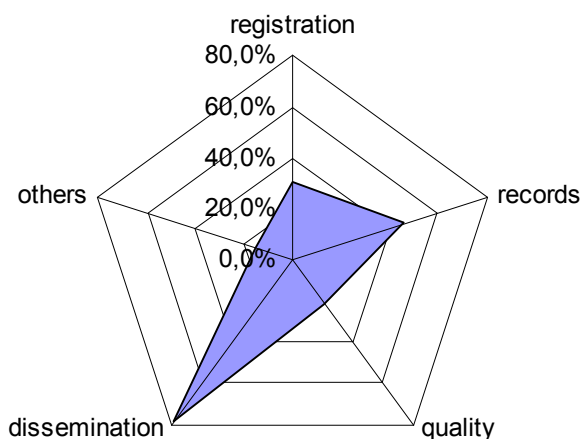


Figure 9: Main functions of GL in survey

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.⁶

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)"⁷. 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 Websites⁸, 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) (Pédaque, 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 GLISC⁹ 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)¹⁰. 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 archives¹¹. 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 intentional 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*¹², 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 H       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. Sch       (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. P         (2003). ‘Document: Form, Sign and Medium, As Reformulated for Electronic Documents’. Tech. rep., CNRS RTP-DOC.
- J. Sch       (2008). ‘Grey literature on bilingualism in Belgium’. *Cahiers de la Documentation / Bladen voor Dokumentatie* **62**(2):22-30.
- J. Sch       & 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.

Notes

¹ <http://opensigle.inist.fr/handle/10068/697753>

² <http://www.refdoc.fr>

³ On the platform <http://fr.surveymonkey.com/>

⁴ 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>

⁵ Concerning relevance of GL for local or regional community, see our recent study on academic publishing in Belgium (Sch      , 2008).

⁶ 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 (...)”

⁷ “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?

⁸ The British Library starts to archive snapshot copies of selected Websites.

⁹ <http://www.glisc.info>

¹⁰ Italics added by JS

¹¹ See the recent study of Luzi (2010).

¹² See Hey et al. (2009).

Annex 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 *4th 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 *8th 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 *5th International Conference on GL: 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. Jeffery (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 *5th International Conference on GL: 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'. *TGJ* 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.

Annex 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 grey literature 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:

A Terminology Based Re-Definition of Grey Literature*

Claudia Marzi, Gabriella Pardelli, and Manuela Sassi (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.

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, Auger (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 from 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 [...]".

* First published in the GL12 Conference Proceedings, February 2011

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.

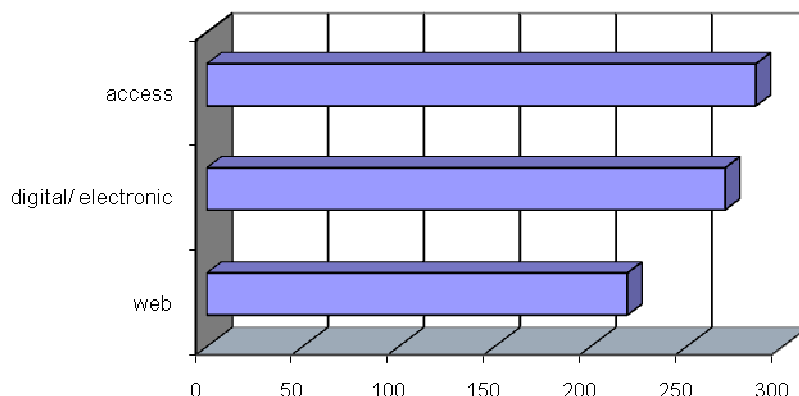
2.2 Data extraction

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

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*¹, *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. mutual information on word pairs), focussing on terminological usages that are closely related to the ontological entities already mentioned above.

The typical collocates of *access* 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 *digital/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.

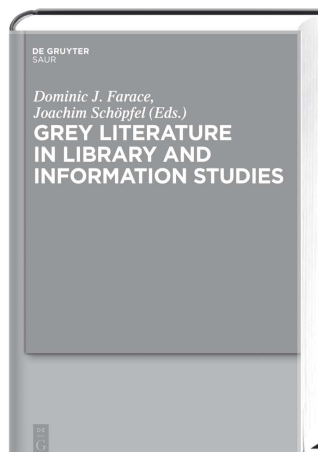
An updated re-definition of GL should take into consideration the key notions of digital medium, web-based distribution channels, information access policy and access and management tools for GL. By bringing these innovative elements into the picture, we are in a position to do justice to recent developments 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

- AUGER, 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 (1), 22-29.
- FARACE D.J.(2006). 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: from 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>

¹ 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.



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

Introduction Grey Literature (*Farace and Schöpfel*)

Part I – Producing, Processing, and Distributing Grey Literature

Section One: Producing and Publishing Grey Literature

Chapter 1 Grey Publishing and the Information Market: A New Look at Value Chains and Business Models (*Roosendaal*)

Chapter 2 How to assure the Quality of Grey Literature: The Case of Evaluation Reports (*Weber*)

Chapter 3 Grey Literature produced and published by Universities: A Case for ETDs (*Južnič*)

Section Two: Collecting and Processing Grey Literature

Chapter 4 Collection building with special Regards to Report Literature (*Newbold and Grimshaw*)

Chapter 5 Institutional Grey Literature in the University Environment (*Siegel*)

Chapter 6 Copyright Concerns Confronting Grey Literature (*Lipinski*)

Section Three: Channels for Access and Distribution of Grey Literature

Chapter 7 Theses and Dissertations (*Stock and Paillassard*)

Chapter 8 Grey Documents in Open Archives (*Luzi*)

Chapter 9 OpenSIGLE – Crossroads for Libraries, Research and Educational Institutions in the Field of Grey Literature (*Farace, Frantzen, Stock, Henrot, and Schöpfel*)

Part II – Uses, Applications, and Trends in Grey Literature

Section Four: Applications and Uses of Grey Literature

Chapter 10 The driving and evolving Role of Grey Literature in High-Energy Physics (*Gentil-Beccot*)

Chapter 11 The Use and Influence of Information Produced as Grey Literature by International, Intergovernmental Marine Organizations: Overview of Current Research (*MacDonald, Wells, Cordes, Hutton, Cossarini, and Soomai*)

Chapter 12 Grey Literature in Karst Research: The Evolution of the Karst Information Portal, KIP (*Chavez*)

Chapter 13 Grey Literature Repositories: Tools for NGOs Involved in Public Health Activities in Developing Countries (*Crowe, Hodge, and Redmon*)

Section Five: Future Trends in Grey Literature

Chapter 14 Blog Posts and Tweets: The Next Frontier for Grey Literature (*Banks*)

Chapter 15 Assessing the Return on Investments in Grey Literature for Institutional Repositories (*Schöpfel and Boukacem*)

Chapter 16 e-Science, Cyberinfrastructure and CRIS (*Jeffery and Asserson*)

Chapter 17 Course and Learning Objective in the Teaching of Grey Literature: The Role of Library and Information Science Education (*Rabina*)

Peering through the Review Process: Towards Transparency in Grey Literature*

Dominic J. Farace (*The 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.

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 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 battlefield appears not to be between grey and commercial publishers but from another affront - that being the black literature, what Carrington [2010] describes as the blogosphere 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

* First published in the GL12 Conference Proceedings, February 2011

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 one's 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 endeavor 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

- The **criteria (functions)** of peer review, including such terms as Validity, Significance, Originality, etc.
- The **roles (stakeholders)** in the peer review process, including such terms as Editor, Reviewer, Publisher, etc.
- The **attributes (characteristics)** of peer review, including such terms as Anonymous, Formal, Evidence based, etc.
- The **process (steps)** in the review of publications, including such terms as Pre-screening, Gate keeping, Rejection, etc.
- 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.



Information International Associates

Creating value from the world's information

- Scientific and Technical Information Management
- Open Source Intelligence Research
- Information Technology Services
- Library and Information Center Management
- Records and Technical Documentation Management
- Education and Training Program Administration



www.iiaweb.com

Information International Associates, Inc (IIa)
P.O. Box 4219
1055 Commerce Park Drive, Suite 110
Oak Ridge, TN 37830
Phone: 865.481.0388 Fax: 865.481.0390
E-mail: inquiry@iiaweb.com

Scientific Data: Increasing Transparency and Reducing the Grey*

Bonnie C. Carroll, June Crowe, and J.R. Candlish (*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.

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).

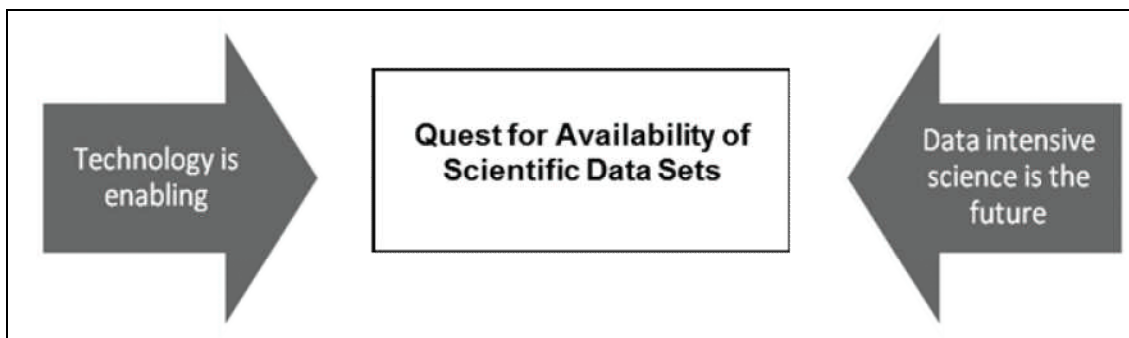


Figure 1. The drivers of scientific data.

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."¹ 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.²

* First published in the GL12 Conference Proceedings, February 2011

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)³

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.

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 data⁴. Data volumes in 2005, for example, were growing at a doubling rate, and datasets were forecasted to reach petabyte size in the near future⁵. 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 2020⁶". 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 properties⁷.

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.

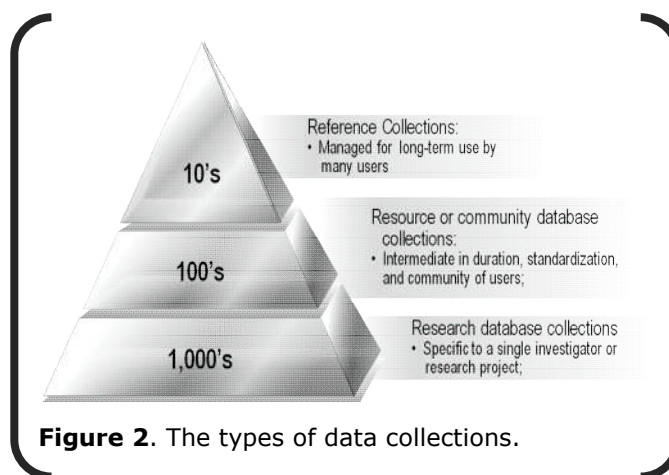


Figure 2. The types of data collections.

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 material⁹.

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 audience¹⁰.

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 areas¹¹.

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 audience¹².

The Information Analysis Center (IAC) is a concept proposed by the Weinberg Report¹³ 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. One 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.gov¹⁵ 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) Clearinghouse¹⁶, and the Electronic Cultural Atlas Initiative (ECAI) Clearinghouse¹⁷. More specific to the scientific community is Mercury¹⁸, 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 States¹⁹ 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 Practices²⁰. 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 α OCT scan enabled quantification of the lumen diameters at the time of the bronchoscopy.

A strong correlation was observed between CT and α 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 α 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.

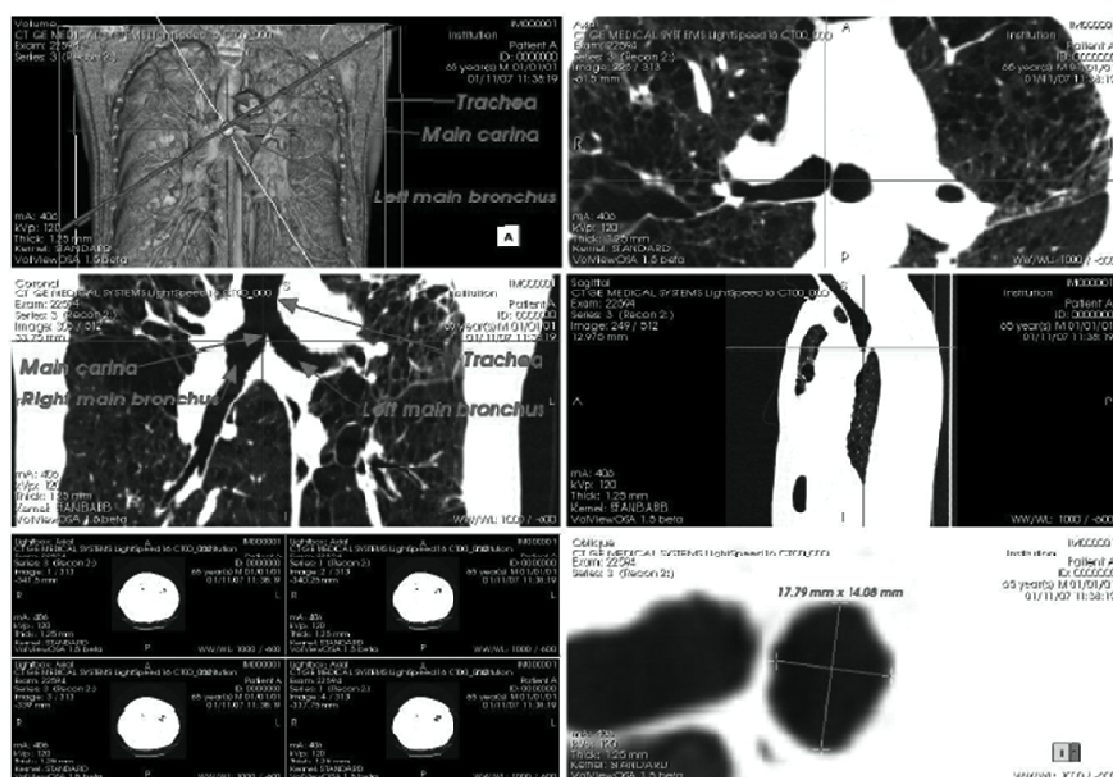


Fig. 2. Patient A. Chest CT depicting the lower airway (View 1). Top row (L-R): 3D view; Axial slice at the level of the main carina. Middle row (L-R): Coronal view; Sagittal view. Bottom row (L-R): Lightbox view; Oblique view measuring airway diameter.

Figure 4. An interactive 2D and 3D dataset derived from ISP software²¹.

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 Dryad²²:

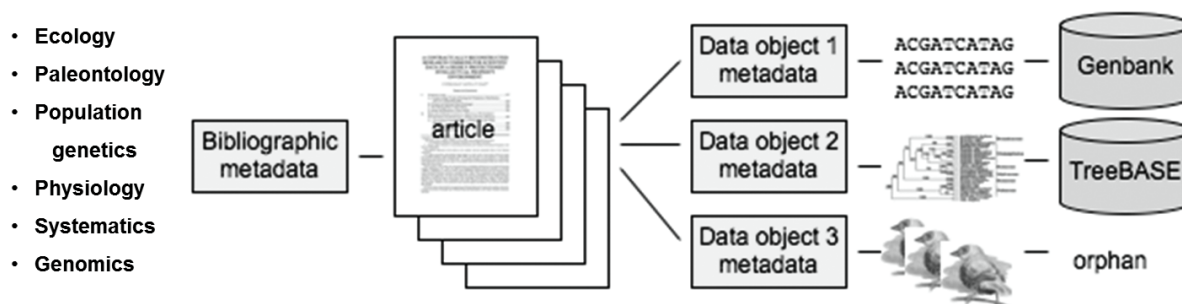


Figure 5: The Dryad repository conceptual flow diagram and general architecture.

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 data²³. Three major data repositories (Atmospheric Radiation Measurement Archive, Carbon Dioxide Information and Analysis Center (CDIAC)²⁴, and the ORNL Distributed Active Archive Center²⁵), comprise the central resources in environmental data. At the core of this enterprise is the Mercury Metadata Clearinghouse²⁶ 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 practitioners²⁷ (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 obtained²⁸. 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.

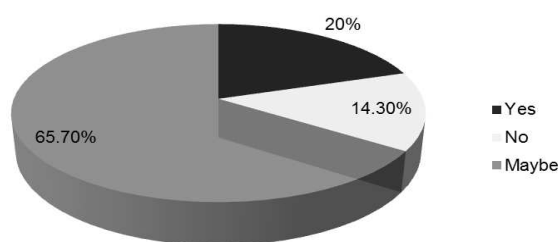


Figure 6: The percentage of survey responses that indicated sufficient knowledge of datasets when planning projects and research programs²⁹.

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 Republic²⁹, laws and Presidential directives in the US have focused on the need for better accessibility to scientific data³⁰. In the UK a new Data Corporation has been proposed which will be 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&orq=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_IWGD_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_IWGD_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>

FIND THE PIECE THAT FITS YOUR PUZZLE



THE GREY LITERATURE REPORT FROM THE NEW YORK ACADEMY OF MEDICINE

Focused on health services research and selected public health topics, the Report delivers content from over 750 non-commercial publishers on a bi-monthly basis.

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

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

For more information visit our website: www.greyliterature.org
or contact us at: greylithelp@nyam.org



**The New York
Academy of Medicine**

At the heart of urban health since 1847

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

Janie Kaplan, Lea Myohanen, and Elizabeth Taylor (*United States*)

Abstract

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's¹ 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.

* First published in the GL12 Conference Proceedings, February 2011

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 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 Witwatersrand², South Africa
- University of KwaZulu-Natal³, South Africa
- University of Dar Es Salaam⁴, Tanzania
- University of Ghana-Legon, Balme Library⁵ 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

References

- ¹ The Grey Literature Report, <http://www.nyam.org/library/online-resources/grey-literature-report/>
- ² University of the Witwatersrand, <http://www.wits.ac.za/>
- ³ University of KwaZulu-Natal, <http://www.ukzn.ac.za>
- ⁴ University of Dar Es Salaam, <http://www.udsm.ac.tz>
- ⁵ University of Ghana-Legon, Balme Library, <http://library.ug.edu.gh/>

A Bridge over Grey Waters: Collecting and Preserving Public Policy Research *

Amanda Lawrence (Australia)

Abstract

This paper looks at the role of Australian Policy Online (APO) as a collection of 'grey' public policy research and considers how this fits within the broader context of open access publishing and policy development. The amount of grey literature now being produced presents challenges for APO as well as other collecting agencies and society in general. This paper looks at who is operating in this space in Australia, the issues for collecting and preserving grey literature, and how bibliographic exchange and interoperability can help us to manage the flood of new material and reduce the duplicated effort that currently exists. While APO has been a foundational player in this field, it has reached a stage where a new solution is required - a national online collaborative collection of full text digital resources created with a vision for the benefits that the semantic web will bring.

Australian Policy Online (apo.org.au) is a web-based, open access collection of Australian public policy related research and information. The core service involves finding, cataloguing and disseminating information on the latest reports, articles, analyses, working papers and statistical documents published 'informally' by research institutes, government and non-governmental organizations, and other agencies. This material, known as 'grey' literature, is not controlled by commercial publishers (Greynet International), meaning it is only collected or catalogued by libraries, publishers or other services in an ad hoc manner. While APO is also an information network and a community space for those working on public policy issues, this paper focuses particularly on APO's role in collecting grey research and considers how this fits within the broader context of open access publishing and policy development. The amount of grey literature now being produced presents challenges for APO as well as other collecting agencies and society in general. This paper looks at who is operating in this space in Australia, the issues for collecting and preserving grey literature, and how bibliographic exchange and interoperability can help us to manage the flood of new material and reduce the duplicated effort that currently exists. APO has been a foundational player in this field, however it has reached a stage where a new solution is required - a national online collaborative collection of full text digital resources created with a vision for the benefits that the semantic web will bring.

APO's strength is that it catalogues and disseminates scattered resources and provides a central archive for further research. It thereby solves a long-standing problem in public policy - ensuring that those who make and implement policy have access to the latest research findings produced by academic institutions and other organisations; and conversely, that researchers and writers are able to disseminate their findings to policy makers in a timely and effective manner.

APO is one of the major ways of bridging the academic/public policy divide, allowing research to flow between academics and practitioners. Around 200 centres and organisations from around the country form the core of APO's content source including: multi-institutional research networks; university research centres; think tanks from all sides of politics; NGOs, charities, associations and professional bodies; government agencies; and government departments and inquiries from Federal, State and Local government.

Public policy issues encompass a great breadth of subject matter and the site covers a wide range of disciplines, as evidenced by its ten major areas (see the screen shot in Figure 1.): creative economy, economics, education, environmental and planning, health, Indigenous issues, international issues, justice, politics and social policy. So although public policy is a subject in itself, APO does not focus on any one particular discipline. It could be described as predominately social sciences but with considerable cross over into areas of technology, health, law and natural sciences.

* First published in 3CMedia Journal of Community, Citizen's and Third Sector Media and Communication, Issue 6 (August) 2010. Available at: http://www.cbonline.org.au/3cmedia/3c_issue6/index.shtml

Edited by the Institute for Social Research, Swinburne University of Technology

Research Commentary Websites Guides Audio Video Events Jobs Submissions A - Z Courses

APO **AustralianPolicyOnline**
ORG.AU "THE BEST ONLINE SOURCE OF AUSTRALIAN POLICY INFORMATION"

Advertise on APO Email Subscription RSS Login TEXT: LARGER | SMALLER

Search APO GO

Home [Creative Economy](#) [Economics](#) [Education](#) [Environment & Planning](#) [Health](#) [Indigenous](#) [International](#) [Justice](#) [Politics](#) [Social Policy](#)

NEW RESEARCH

Enabling Australia: Inquiry into the Migration Treatment of Disability
Joint Standing Committee on Migration
30 June 2010 | Australia's current health requirement for migration to this country is prejudiced against people with a disability and their families, according to this report.

Building a modern committee system
House Standing Committee on Procedure
30 June 2010 | This report recommends a set of complementary measures designed to strengthen the system of committees in the House of Representatives.

IPTV and internet video delivery models: video content services over IP in Australia
Australian Communications and Media Authority
30 June 2010 | Australians are increasingly able to view video content on their own terms, including when, where and how they want it, according to this report.

State of our public hospitals June 2010
Department of Health and Ageing
30 June 2010 | The seventh in an annual series, this report on public hospital performance in 2008-09 includes information on private hospital performance and has chapters on hospital use by Aboriginal and Torres Strait Islander people and on hospital maternity services.

The future of cloud computing
Lee Rainie, Janna Quitney Anderson | Pew Internet and American Life Project
28 June 2010 | Technology experts and stakeholders say they expect internet users will 'live mostly in the cloud' in 2020 and not on the desktop, working through

Compulsory income management: a flawed answer to a complex issue
22 June 2010 | From 1 July 2010 major changes are due to be made to the way income support payments are paid to recipients in the Northern Territory, and potentially across Australia.

COMMENTARY

Campaigning to Christians
30 June 2010 | The elevation of Julia Gillard means the Australian Christian Lobby now needs to engage afresh with a new Prime Minister, writes John Warhurst in Eureka Street

Anxious overture
29 June 2010 | Kevin Rudd's downfall revealed the limits of Labor's ability to introduce reforms opposed by powerful interests, writes Geoffrey Barker in Inside Story

The way ahead is less complicated than it seems
28 June 2010 | Kevin Rudd departed the prime ministership this week

MOST VIEWED:

- 1 Who will benefit from the 1 July 2010 tax cuts?
- 2 Campaigning to Christians
- 3 Our future world: an analysis of global trends, shocks and scenarios
- 4 Review of Discovery laws and practice to improve access to justice
- 5 Senior Innovation & Strategy Officer

THIS WEEK

QUICKLINKS

- Citizenship
- Creative industries
- Disability
- Employment
- Housing
- Mental health
- Schools
- Climate change
- Defence
- Education policy
- Health policy
- Immigration & refugees
- Poverty
- Social problems

SUBSCRIBE TO THE APO WEEKLY BRIEFING

JOBS

Senior Innovation & Strategy Officer
Innovation & Strategy Unit, Victorian Department of Justice
Melbourne

Senior Urban Planner: Consulting Opportunity
Urbis
Brisbane

Urban Planner: Consulting Opportunity
Urbis
Brisbane

EVENTS

WORKSHOP
Developing Monitoring and Evaluation Frameworks
13 May 2010 - 6 Aug 2010
Australian capital cities & Cairns

WORKSHOP
Using purposeful program theory and logic models for evidence-based policy and practice
28 Jun 2010 - 27 Jul 2010
Melbourne, Adelaide, Brisbane, Darwin, Sydney

CONFERENCE

Figure 1. Home page of Australian Policy Online with the site's ten major subject areas forming the navigation.

Audience and funding

According to an online reader survey conducted in 2007, about a third of APO's audience is within the education sector, at least another third come from local, state or federal government sectors, and the final slice of pie is a mix of NGOs, charities, media, professionals and librarians. By definition, public policy is of concern to the public, as well as specialist researchers, which means that a broad range of people engage with APO for a wide variety of reasons in a way that is different to proprietary research databases or specialist subject archives. As a result the service is particularly appreciated by those unable to access alternative sources of scholarly material such as journalists, independent researchers, charities and NGOs.

According to Google Analytics figures, APO's readership is growing by at least 30% annually and the site now attracts over 65,000 visits each month, and over half a million for the year, viewing nearly 1.5 million pages. Over 14,000 people subscribe to the *Weekly Briefing* email newsletter and many access the site via RSS feeds, Twitter or the National Library's Australian Research Online harvester. In 2009 APO was ranked in the top 10 for politics websites in Australia by Hitwise (Hitwise 2009).

The site receives financial support from the Institute for Social Research, Swinburne University and the ARC Centre of Excellence for Creative Industries and Innovation (CCI) and has received a number of Australian Research Council (ARC) grants for infrastructure and equipment funding.

Background and project development

Australian Policy Online was established in 2002 as a project of the Institute for Social Research, Swinburne University of Technology, based on a site in the US called Moving On

(since closed). The original idea was to gather research reports from academic centres and institutes based on free membership of APO and provide a weekly email newsletter alert. Each week 10 – 15 reports were posted along with 1 or 2 original commentary pieces, some events, calls and the rare job.

With the increased use of the internet as a publishing vehicle for all kinds of organisations including government, it became clear that limiting APO's content to only member organisations was too restrictive. In 2006 the collection policy was expanded to include any policy related report or document including those from government, NGOs and associations where appropriate. With this extended scope APO increased its value, becoming a central resource that allowed key players producing policy research to share information with each other. It also massively expanded the amount of information the editors needed to filter and to share with the APO audience; the site now posts up to 50 items each week with around 35 being research publications.

Over time it became clear that APO needed to move onto a new platform to better cope with the rapidly increasing size of the database and to maximise the potential of new software and functionality. Thanks to an ARC grant, the database was migrated from a proprietary content management system to the open source system Drupal in 2009. As part of the upgrade the site had a visual and architectural redesign, a new subject classification structure was created, new metadata fields were added, the type of content collected was extended to include audio, video and web resources, plus new sections for course and book listings. Various other functions and features were added including faceted search, tailored email, most viewed results and a bookmarking system.

A catalogue or a repository?

In upgrading the site, it was intended that APO would have the capacity to upload and store full text PDFs of documents. This was desirable for two reasons. Firstly to avoid the ongoing and inevitable problem of links going dead as a result of documents being moved and URLs changing. Secondly to allow for full text searching of resources on APO. While this is now technically possible and some publications are uploaded two issues have arisen in carrying out this aim. One is copyright and the obligation to secure permission to store a copy of a copyrighted work. The second is that preserving documents involves a much greater degree of time, effort and cost. More importantly, it involves a whole new approach to technical and standards issues than APO has dealt with previously. It means becoming a repository rather than just a catalogue, something that has not been part of the site's conception.

The impact of trying to store full text holdings on the site has brought to the fore the need to understand the larger issue of grey literature collecting and preservation within the broader research and publishing environment in Australia. This is not a bad thing, for as Bryson (2004) points out:

To respond effectively to changes in their environments, public and non-profit organizations (and communities) must understand the external and internal contexts within which they find themselves so that they can develop effective strategies to link these two contexts in such a way that public value is created.

Scholarly publishing and the open access movement

A key question facing us today is, are there new opportunities and new models for scholarly communication that could enhance the dissemination of research findings and, thereby, maximise the economic and social returns to public investment in R&D? (Houghton, Steele & Sheehan 2006)

The answer to Houghton's question is both yes and no. The backbone of academic scholarly publishing and communication was, and still is, the traditional system of peer-reviewed scholarly journals along with peer-reviewed monographs (Harley et al. 2010). These are predominantly assessed by academics but published and distributed by publishers and aggregators. To read these journals, even in an online environment, requires access to subscription only databases paid for by university funds and accessed via the academic library. As Harley et. al. (2010) attest, 'These traditions, which rely heavily on various forms of peer review, may override the perceived "opportunities" afforded by new technologies, including those falling into the Web 2.0 category'.

In contrast to the persistence of traditional models of publishing is the ongoing exploration of the new opportunities and models referred to by Houghton – primarily with the open access movement. In Budapest 2002 it was declared that 'many different initiatives have shown that open access is economically feasible, that it gives readers extraordinary power to find and make use of relevant literature, and that it gives authors and their works vast and measurable new visibility, readership, and impact'(Budapest Open Access Initiative 2002).

This has been no overnight revolution, and there is still no clear answer to what models work and no solution to the financial and institutional barriers to these new models. A recent JISC report has identified three main options for scholars to consider when publishing research as open access: open access journals, open access repositories and open access repositories with overlay services – all of which only account for formally published journal articles and conference papers (Swan 2010).

For many involved in the open access movement in higher education, the publications of interest are peer-reviewed journal articles or their pre-print versions and peer-reviewed conference papers. The management and preservation of material beyond these parameters is often not considered part of the investigation.

While most university based researchers are able to access both subscription and open access research, many in government, NGOs, the community sector, media, professions, and the general public have had limited access to scholarship. As a result, in the public policy arena (which in itself cuts across many social science and some science and medical disciplines) there has developed a strong culture of informal publishing that looks set to continue and in fact, increase massively.

APO's rapid collection of grey literature is particularly important in public policy as new research can be influential if it is timely and accessible (i.e. not held up by the slowness of journal publishing or the need to write for an academic audience) and reports are often produced explicitly to provide evidence for policy decisions.

Also in public policy there are many producers of research outside of academia including government and non-government organisations, associations and think tanks. Grey literature is often the best source of up-to-date research in specific areas and it is generally written in a more accessible style than journal articles, providing a clear, concise introduction to difficult or complex topics. In health, law, environmental studies, economics, creative industries to name a few, grey literature is a major part of the research material published and used. As Tyndal (2008) points out in the health sciences,

While searching the published (black) literature is a given, there is a growing recognition that grey literature should be included to fully reflect the existing evidential base... Literature reviews need to include the most significant research available. And that might mean conference papers, reports, legislation or working papers - all grey.

As a result of simplified tools for desk top and web publishing, the explosion in content management systems, and the power and popularity of internet search engines, it is easier than ever for individuals and organisations to produce and publish their own research online. The amount of quality research material and other relevant and legitimate material being published outside commercial publishing channels is now completely unmanageable without a major, coordinated and collaborative change of approach, plus considerable technical input.

Collecting grey literature

Despite its importance and increased prevalence there is little consensus on how grey literature should be published, collected, disseminated, evaluated or preserved in Australia (or elsewhere). In a recent call for papers for *Cataloguing & Classification Quarterly* on the issue of grey literature the editors state that:

Despite the advantages offered by new information technologies and, in particular by the ability of search engines to retrieve documents on the web, grey literature still lacks bibliographic control. ...unlike journals and books, the absence of commercial stakes has contributed to a situation where the way in which the different types of grey literature are referenced still depends more on choices made by the bodies that produce, collect, or distribute these documents than on any national or international standard (A Library Writers Blog 2010).

In Australia there is still no systematic collection or preservation of grey documents, or any agreement on how this should be done. In terms of government documents, the National Library of Australia collects some government publications but is not obliged to do so in any exhaustive way. The National Archives collects all government documents but not until departments agree to finally hand them over, which may be years after items are published.

This is set to change to some extent with the recent adoption by the Australian Government of the recommendations from the Gov 2.0 taskforce (Australian Government 2010).

Due to a major injection of funding, most Australian universities have instituted repositories in the last 5 years to collect and preserve their own staff's publications (whether open access or not) (ARROW 2008). However the majority of university repositories limit their policies to collecting only scholarly journal publications and conference papers (exceptions include Swinburne University of Technology and University of Technology, Sydney).

As for the not-for-profit or 'third sector', it seems that no one is thoroughly collecting or preserving the output of NGOs, charities, service agencies, associations, corporate researchers, professional sectors or independent scholars. Yet these organisations are commissioning and publishing substantial research material with direct bearing on public policy issues.

Australian Policy Online is one of the largest collections of social science grey literature in Australia and yet it is not comprehensive, does not preserve material and restricts its selection to public policy related material. Of course APO is not the only subject-based gateway to online research, there are in fact over 55 according to a review conducted recently of online collections. Like the objects they collect, there are many terms that are currently in use to describe an online database of resources – clearing house, digital library, resource centre, subject gateway etc. and there may be subtle differences in the aims and functions of each one. Many have been created to resolve a perceived need for "a collection of resources" on a particular topic and thus a new database and website is born. They range from small to large – from the Victorian Women's Disability Clearinghouse to Health Insite for public health guides – but all are subject specific. Another type of collection is the university and organisation repository but here the collection policy is based on author association rather than subject, and, as mentioned earlier, most do not collect grey literature.

The problem with so many clearing houses and subject collections is that currently they are all cataloguing online content which means there is a great deal of duplicated cataloguing of grey literature for subject clearing houses, libraries and specialist collections – mostly at tax payers' expense. There is little sharing of metadata records for materials freely available online and even though many websites are in fact digital libraries, most aren't currently available to browse within the public or academic library systems.

In terms of preservation various problems exist including:

- No agreed frameworks or metadata standards for cataloguing publications and other related items.
- No mandate or agreement to preserve documents and no systematic archival procedure.
- No agreement on the allocation of persistent identifiers and therefore no capacity to link documents and research data or associated content.
- No clearly articulated publishing standards for organisations wishing to publish research informally i.e. using permanent URLs, Dublin core metadata etc.
- No place for individuals or organisations to upload their own documents in an Australian repository.

As the recent report from the Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010 states:

Clarification of the long-term value of emerging genres of digital scholarship, such as academic blogs and grey literature, is a high priority. Research and education institutions, professional societies, publishers, libraries, and scholars all have leading roles to play in creating sustainable preservation strategies for the materials that are valuable to them.

International systems

Internationally there are a number of interesting examples of how open access research and grey literature are being managed. Like Australia there are many clearing houses and specialist subject gateways however there are also a number of large scale collecting operations that provide interesting models for consideration.

Social Science Research Network (www.ssrn.com) is an international repository where researchers are able to upload their documents. The system is run as a business by charging annual subscriptions to subject specific email newsletters in various disciplines. It also includes abstracts from open access journals. This kind of mass collaborative model is set to

take over as the kind of search system researchers want to use. However additional services such as the SSRN's email newsletters could be provided by database partners if a more communal model was adopted.

A more general example is Scribd (www.scribd.com), which describes itself as 'the largest social publishing and reading site in the world.' Scribd allows anyone to post anything and turns all documents into web-based pages, allowing tagging and sharing. Harvard University Press uses Scribd to post its catalogue for readers to look at on Scribd – so it is both a publishing tool for non-publishers and a marketing tool for publishers.

These kinds of systems represent a major change in how individuals and organisations are able to disseminate their own material – they don't even need their own website, nor do they require the editing services of a site such as APO. It therefore appears necessary for APO to seriously question the extent to which it should be cataloguing all its own material and instead work towards creating a larger, open repository where grey literature could be uploaded and catalogued, with layers of services added on including metadata exchange and sharing, filtering and dissemination.

Filtering and evaluation

In the information age, this ability to filter effectively has moved from an essential of survival to one of the primary determinants of success (Dawson 2003).

As the quantity of quality research being made available increases along with the volume of grey literature, so does the need to filter it. It is not enough for APO to find, catalogue and disseminate research, the service must find ways to filter information so that readers are not bombarded with ever increasing amounts of undifferentiated information. Even the traditional systems of peer-review are no longer adequate given the amount of material available and various attempts at post publication peer review are now being made. Houghton, Steele & Sheehan (2006) suggest that from this problem springs a number of opportunities:

there are potential futures for publishers to become value adding service providers overlaying open access content (e.g. peer review services, bibliometrics and webometrics for research evaluation, etc.). In turn, these might enhance research evaluation and lead to better focused R&D expenditures.

Filtering systems range from simple automated systems to massive international collaborations by skilled experts. In terms of the former, filtering systems can be created automatically based on user interaction with a site's content. For example by displaying how many people viewed a particular page or downloaded a report, how many other publications cited it, how many recommended it or found it useful.

Yet while these numbers have a place they are limited within the research domain. Recommended for what? Useful in what way? The rating system on the video site 'Ted talks' (ted.com) doesn't just ask users to say if they liked or disliked a video, or to click a thumbs up or down symbol, it asks users to consider whether the item was jaw-dropping, persuasive, funny, longwinded, unconvincing etc. Perhaps therefore we should be looking at a rating system that might provide options such as: ground breaking, well-written, good synthesis, what works etc. In this way the site might incorporate a web 2.0 type of peer-review for grey literature.

An interesting example of post-publication peer review is Faculty of 1000 which currently has two sites, one for Biology and one for Medicine. The screen shot below shows the rating system that can be applied by over 2000 biologists from around the world, in addition to short paragraphs indicating a publication's research value. The Faculty of 1000 is an interesting model for applying post-publication peer review to grey literature.

For each paper they select, the Faculty member:

1. **Writes a brief Comment** explaining the main findings and why they find the paper interesting. These comments are intelligible to a general biology audience. Faculty members often mention related seminal research articles within the relevant papers.
2. **Assigns a Rating**

 Recommended F1000 Factor 3.1	"Recommended": Recommended reading for a section (i.e. of specialist interest) (F1000 Biology rating of 3)
 Must Read F1000 Factor 6.2	"Must read": Must read for more than one subject/section (i.e. of general interest) (F1000 Biology rating of 6),
 Exceptional F1000 Factor 10.7	"Exceptional": A landmark paper representing the top 1% of publications (F1000 Biology rating of 9).

The ratings assigned to each paper is then used to calculate its F1000 factor (see below), which also takes into account the same paper.
3. **Classifies it** into any of seven paper types
 - ☐ New finding
 - ☐ Technical advance
 - ☐ Interesting hypothesis
 - ☐ Important confirmation
 - ☐ Controversial findings
 - ☐ Refutation
 - ☐ Novel Drug Target

Figure 2. Faculty of 1000 Biology evaluation system <http://f1000biology.com/about/system>

Tyndal (2008) points out that while there is some analysis and acknowledgement of grey literature's importance in the health sciences, there is major uncertainty about what evaluation criteria should be used as there is no clear hierarchy to evaluate grey literature, with different disciplines focussing on different resources. In her view the standard evaluation criteria for websites should be used: authority, accuracy, objectivity, coverage and date – known as AAOCD. This could be used as the basis for discussing collection policies for grey literature – a very difficult issue that is beyond the scope of this paper.

Standards and interoperability

Trends in the web environment are emphasizing interoperability among multiple standards for metadata used and recycled in the book industry, rights management, and library and information sectors (Program for Cooperative Cataloguing Mission Statement 2010).

The library sector has long known the value of bibliographic exchange and has been working on this for the last 30 years (Hider & Harvey 2008). In World Cat, using the Online Computer Library Centre (OCLC) Connexion software, cataloguers can input URLs for web-based materials and Connexion will automatically generate brief bibliographic records for them which human catalogues then edit and add to (OCLC 2010).

Another type of automation is the extraction of keywords for subject classification such as the Open Calais system (www.opencalais.com). Academic reference management and bookmarking software such as Mendeley (www.mendeley.com), Connotea (www.connotea.org), Icyte (www.icyte.com) and Zotero (www.zotero.org) allow readers to post metadata details with the click of a button – but the effectiveness of such systems depends to some extent on all producers of research documents using metadata in the first place. This may require the promotion of Dublin core to informal publishers of research in Australia as well as the installation of the capturing technology in online collection infrastructure.

Mendeley appears to be one of the rising stars in this area. It allows readers to create their 'own personal bibliography online' by indexing and organising PDF uploads. It also looks up PubMed, CrossRef, DOIs and other related document details automatically. The Web Importer allows you to quickly and easily import papers from resources such as Google Scholar, ACM, IEEE and many more at the click of a button.

Protocols such as RSS and the Open Archives Initiative (OAI) allow APO to be harvested by various systems, including the National Library of Australia and the OCLC's World Catalogue via OAIster. This is a good start but there also needs to be collaboration on how information enters the system in the first place. With the combination of technology and metadata standards, there seems to be little justification for research, including grey literature research, to be catalogued multiple times when it is so easy to share metadata if agreed standards are used.

Standards for grey literature publishing and collecting have been agreed upon and implemented to some extent in the sciences (Committee 2007) but there has been less take up and implementation within the social sciences.

Interoperability is the capability that allows different computer systems to share information across a network... By adopting a common set of best practices, controlled vocabularies, and interoperable system architecture, institutions can increase their visibility and provide opportunities for new connections with others to serve the shared needs of constituent communities (CDP Metadata Working Group 2006).

With so many standards and schemas now in existence it is hard to know where to start to identify what might be the best options for grey literature. According to Hider there are two solutions to the problem of standards proliferation - translate or unify; 'The main attempt at unification has been Dublin Core...' (2008 p. 214). Dublin Core stands out as a schema because it aims to be used across information domains rather than be more specialised. However it was designed to describe web pages and is therefore limited in its ability to contextualise information. Despite this no other metadata schema has been adopted as widely. One of Dublin core's great advantages is that it works with the open archives initiative protocol for metadata harvesting (OAI PMH) which has become a popular vehicle for sharing metadata across collections in the library community.

However, according to the Common European Research Information Format project, the current use of Dublin Core or even MARC means that 'the metadata is machine readable but not machine-understandable. The end-result is that the input, update and retrieval processes are human-intensive. With increasing availability of repositories of grey literature this will not scale.' The Europeans are therefore creating their own system - called CERIF. It is not clear at this stage which system if any will become a universal standard so this situation needs continued monitoring.

Warwick Cathro at the National Library of Australia (2009) identifies some of the key standards for data interoperability as:

- MARC (Machine-Readable Cataloguing) - used by library catalogues
- MODS (Metadata Object Description Schema) - developed by Library of Congress to provide a link between MARC and Dublin Core.
- METS (Metadata Encoding and Transmission Standard) Structured for encoding descriptive, administrative, and structural metadata. Designed by the Library of Congress especially for metadata about electronic publications in digital libraries. Includes extra administrative metadata.
- PREMIS (Preservation Metadata - Implementation Strategies) - A data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials.

In the case of software inter-operability, relevant standards include the NISO Circulation Interchange Protocol (NCIP), the Information Retrieval Protocols (Z39.50, SRU), and the Open Archives Initiative Protocol for Metadata Harvesting (OAI).' (Cathro, 2009). Another important standard for maximising the potential of the long awaited semantic web is Resource Description Framework (RDF). This a standard framework for structuring metadata in XML and all metadata schemas expressed in XML should conform to the RDF. The advantage of RDF is that it allows relationships between web documents to be identified by a computer. It is a core element of the semantic web which many are hailing as the next big leap for the internet (O'Reilly 2009).

Finally there is a need for Uniform Resource Identifiers (URI) and the two systems currently operating to provide these: DOI and Handles. These protocols enable a distributed computer system to store identifiers, known as handles, of arbitrary resources and resolve those handles into the information necessary to locate, access, contact, authenticate, or otherwise

make use of the resources – in particular to always be able to locate and link them to other material. DOI's are proprietary while Handles are open source. DOI's have been adopted by publishers while Handles are being used by University repositories.

Linked Data

Interoperability and standards are at the core of the semantic web, which is about making sure that the relationships between information are machine readable and therefore able to be connected and visualised. This raises the issue of how publications and data can be linked. In biology researchers have been working on how to connect publications and data for some time and Philip Bourne (2005) presents a compelling case for what might be possible:

...what if the data in an online paper became more alive? Some databases let you download data into spreadsheets or other client-side applications that render and analyze data. Papers could be treated this way, too. The technology is there to create these ubiquitous clients that are independent of operating systems and hardware and that are downloadable on demand. New levels of comprehension might be achievable.

A great deal of public policy and social science research is quantitative and the opportunity to analyse the data as well its interpretation would be a great boon for evidence-based policy. A major opportunity presented by the semantic web is the potential to connect research publications with the data held in collecting agencies such as the Australian National Data Service and the Australian Social Science Data Archive (ASSDA). But to do this there needs to be a commitment to preserving grey literature and implementing standards using handles or DOI on all grey literature publications and have relatable identifiers on datasets.

Conclusion

As a catalogue of public policy grey literature, Australian Policy Online needs to understand the context in which its services are used and the way in which the type of content it holds is perceived, created, managed, preserved and disseminated. This analysis has revealed that in fact the service that APO provides is a unique and valuable one but that there is a greater issue that APO could be assisting to resolve – that of collaboratively preserving, cataloguing and sharing metadata about Australian grey literature. At the same time there are many efficiencies that APO could gain from implementing automated data extraction now where possible. Cataloguing and preservation need to be separated from services such as specialist subject access, evaluation, filtering systems, value added services, and community of practice stewardship. APO aims to be a driving force in helping to solve this national research infrastructure problem while continuing to provide specialist services for its constituent community of public policy researchers and practitioners. We welcome any comments or thoughts from others interested in creating such a resource.

References

- A Library Writers Blog 2010. *CFP: Special issue: Cataloging Grey Literature (Cataloging & Classification Quarterly)*. Available at: <http://librarywriting.blogspot.com/2010/01/cfp-special-issue-cataloginggrey.html> (accessed 17 February 2010).
- Australian Research Repositories Online to the World (ARROW). Available at: <http://arrow.edu.au/> (accessed 10 April 2010).
- Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010, *Sustainable economics for a digital planet*, OCLC. Available at: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf (accessed 4 June 2010).
- Bourne, P. 2005. Will a Biological Database Be Different from a Biological Journal? *PLoS Computational Biology*, Vol 1, Issue 3. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1193993/pdf/pcbi.0010034.pdf> (accessed 24 May).
- Bryson, J. M. 2004. *Strategic planning for public and nonprofit organizations: A guide to strengthening and sustaining organizational achievement*, 3rd edition., John Wiley & Sons, Hoboken.
- Committee, GLIS 2007. *Guidelines for the production of scientific and technical reports: how to write and distribute grey literature*. Available at: <http://www.glisc.info/> (accessed 15 March 2010).
- Budapest Open Access Initiative, 2002. Open Society Initiative. Available at: <http://www.soros.org/openaccess/read.shtml> (accessed 15 May 2010).

- Cathro, W. 2009. Collaboration Strategies for Digital Collections: The Australian Experience, International Conference on Libraries, Seoul, Korea, 25-26 May 2009. Available at: <http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1433/1738> (accessed 22 April 2010).
- CDP Metadata Working Group 2006. *Dublin core metadata best practices*, Collaborative Digitization Program, University of Denver. Available at: <http://www.bcr.org/dps/cdp/best/dublin-core-bp.pdf> (accessed 13 June 2010).
- Dawson, R. 2003. *Living networks*, Prentice Hall, Upper Saddle River, New Jersey.
- Faculty of 1000. Available at: <http://f1000.com> (accessed 10 April 2010).
- Australian Government, 2010, *Government Response to the Report of the Government 2.0 Taskforce*, Canberra. Available at: <http://www.finance.gov.au/publications/govresponse20report/index.html> (accessed 30 June 2010).
- Greynet International, nd. Available at: <http://www.greynet.org/greynethome.html> (accessed 15 May 2010).
- Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S. and King, C. J. 2010. *Assessing the future landscape of scholarly communication: an exploration of faculty values and needs in seven disciplines*, Center for Studies in Higher Education, University of California, Berkeley. Available at: http://escholarship.org/uc/cshe_fsc (accessed 10 June 2010).
- Hearst, M. A. 2006. *Design recommendations for hierarchical faceted search interfaces*. School of Information, UC Berkeley. Available at: <http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf> (accessed 15 March 2009).
- Hider, P. and Harvey, R. 2008. *Organising knowledge in a global society*, Centre for Information Studies, Charles Sturt University, Wagga Wagga.
- Hitwise 2009. *Hitwise Top 10 Online Performance Award 2009*, 2009. Available at: <http://www.hitwise.com/awards/popup.html?market=au&sDomain=www.apo.org.au&iDate=200904&iCatnum=295&Cal=&semi=2> (viewed 15 June 2010).
- Houghton, J. Steele, C and Sheehan, P. 2006. *Research communication costs in Australia, emerging opportunities and benefits*, Centre for Strategic Economic Studies, Victoria University, Melbourne. Available at: <http://www.cfes.com/documents/wp24.pdf> (accessed 27 April 2008).
- OAIster, nd. Available at: <http://www.oclc.org/oaister/> (accessed 15 June 2010).
- OCLC 2010. Connexion, Ohio. Available at: <http://www.oclc.org/connexion/> (accessed 16 May 2010).
- OECD 2010. *The economic and social role of Internet intermediaries*. Available at: <http://www.oecd.org/dataoecd/49/4/44949023.pdf> (accessed 16 April 2010).
- O'Reilly, T. & Battelle, J. 2009. *Web Squared: Web 2.0 Five Years On*. Available at: http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper.pdf (accessed 1 September 2009).
- Program for Cooperative Cataloging Mission Statement, 2010. Library of Congress. Available at: <http://www.loc.gov/catdir/pcc/stratdir-2010.pdf> (accessed 13 June 2010).
- Research and the scholarly communications process – principles. 2006. Research Information Network, London. Available at: <http://www.rin.ac.uk/system/files/attachments/Research-scholarlycommunications-principles.pdf> (accessed 29 March 2010).
- Shirky, C. 2009. *Here comes everybody: the power of organizing without organizations*, Allen Lane, London.
- Swan, A. 2010. *Modelling scholarly communication options: costs and benefits for universities*, JISC, London. Available at: <http://ierepository.jisc.ac.uk/442/> (accessed 13 May 2010).
- Tyndal, J. 2008. *How low can you go? Towards a hierarchy of grey literature.*, Dreaming08: Australian Library and Information Association Biennial Conference, 2-5 September 2008, Alice Springs. Available at: <http://hdl.handle.net/2328/3326> (accessed 18 May 2010).

Library, Information Science & Technology AbstractsTM with Full Text

Available via EBSCOhost[®]

The definitive professional information
resource designed for librarians and
information specialists...

*Library, Information Science & Technology AbstractsTM
with Full Text* is an indispensable tool for librarians
looking to stay current in this rapidly evolving field.

Comprehensive content includes:

- Full text for more than 270 journals and nearly 20 monographs
- Indexing for more than 550 core journals, 50 priority journals and nearly 125 selective journals
- Includes books, research reports, proceedings and author profiles
- Access to 6,800 terms from reference thesauri
- Coverage extends back as far as the mid-1960s

Subject coverage includes:

- Bibliometrics
- Cataloging
- Classification
- Information Management
- Librarianship
- Online Information Retrieval
- And much more...

Contact EBSCO Publishing to learn more about
Library, Information Science & Technology AbstractsTM with Full Text,
or to request a free trial.

Phone: 800.653.2726

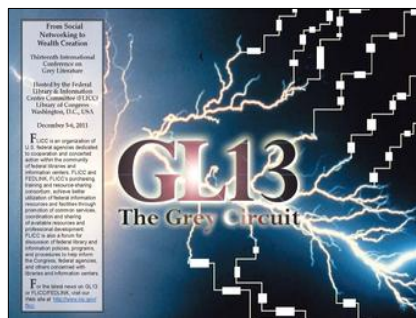
Email: request@ebscohost.com

www.ebscohost.com



Thirteenth International Conference on Grey Literature

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



The Grey Circuit From Social Networking to Wealth Creation

Call for Papers

Guidelines for the Submission of Abstracts

Participants who seek to present a paper at GL13 are invited to submit an English abstract between 300-400 words. The abstract should deal with the problem/goal, the research method/procedure, an indication of costs related to the project, as well as the anticipated results of the research. The abstract should likewise include the title of the proposed paper, topics most suited to the paper, name(s) of the author(s), and full address information. Abstracts are the only tangible source that allows the Program Committee to guarantee content and balance in the conference program. Every effort should be made to reflect the content of your work in the abstract submitted. Abstracts not in compliance with the guidelines will be returned to the author for revision.

Title of Paper, Author and Address Information

Title of Paper:	Topics suited to Paper:
Author Name(s):	Telephone:
Organization(s):	Fax:
Address/P.O. Box:	Email:
Postal Code - City - Country:	URL:

Indicate the Topic(s) most suited to your Paper

Social Networking – Subject based Communities – **Quality Control** – Enhanced Publication
Open Access – Open Source – **Special Collections** – Discovery Services
New Frontiers – *e-Science* – *Legal Concerns*
Data – **Datasets** – **REPOSITORIES** – **Metadata**
Competitive Intelligence – **USE AND USER STATS** – *Lean Thinking* – *Wealth Creation*

Due Date and Format for Submission

Abstracts in MS Word must be emailed to conference@textrelease.com by **April 10, 2011**. The author will receive verification upon receipt. Shortly after the Program Committee meeting on April 28th 2011, authors will be notified of their place on the conference program. This notice will be accompanied by further guidelines for the submission of the full-text papers as well as PowerPoint slides. Authors will then have 30 days to complete Conference Registration and confirm their place on the GL13 Conference Program.

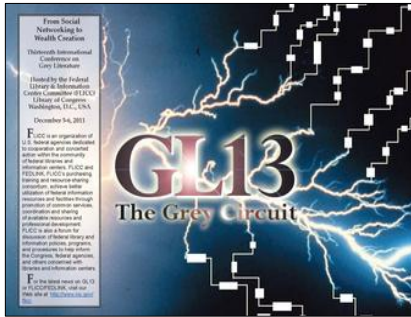
GL13 Program and Conference Bureau

TextRelease

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

Thirteenth International Conference on Grey Literature

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



The Grey Circuit From Social Networking to Wealth Creation

Registration Form

The Conference Registration Fee includes attendance during the Sessions, a copy of the GL13 Program and Abstract Book, the full-text of the Conference Papers and PowerPoint Slides, as well as the conference pouch and participant badge. Also included are lunches, refreshments during the morning and afternoon breaks, the Inaugural Reception, and post-conference tour.

Payment after October 1, 2011 add a US\$ 75 surcharge

☐ **US\$ 595** Participant Fee ☐ **US\$ 495** Author Fee ☐ **US\$ 350** Day Pass ☐ **US\$ 175** Student Fee



I work for a U.S. Government Agency and am entitled to 20% reduction on the fee: ☐

I am a *GreyNet* Member and am entitled to 20% reduction on the conference fee: ☐

Registrant's Name: _____

Organization: _____

Address/P.O. Box: _____

Postal Code/City/Country: _____

Tel/Fax/Email: _____

Please Check One of the Boxes below for the Method of Payment:

☐ Direct bank transfer to TextRelease, Account 3135.85.342 - Rabobank, Amsterdam, Netherlands
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to GL13 and Registrant's Name

☐ MasterCard ☐ Visa Card ☐ American Express

Card No. _____ Expiration Date: _____

If the name on the credit card is not that of the registrant, print the name that appears on the 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.

GL13 Program and Conference Bureau

TextRelease

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

Candlish, J.R.

33

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). Email: jcandlish@iiaweb.com

Carroll, Bonnie C.

33

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

Crowe, June

33

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

Farace, Dominic J.

25

Dominic Farace is Director of TextRelease, an Amsterdam based information bureau specializing in grey literature and networked information. He founded GreyNet, Grey Literature Network Service, in 1993 and has since been responsible for the international GL-Conference Series. In this capacity, he serves as Program and Conference Director as well as managing editor of the conference proceedings. Since 2004, he is a Guest Lecturer on Grey Literature in the Masters Program at the University of Amsterdam; Instructor of Grey Literature via UNO Distance Education, and Editor of TGJ, The Grey Journal. Email: dominic.farace@textrelease.com

Kaplan, Janie

41

Janice J. Kaplan is Director of the Library & Special Collections at the New York Academy of Medicine. She received her Master of Librarianship from Emory University and Bachelor of Arts in English Language and Literature from Boston University. She has been involved in the field of medical grey literature and has been an Associate Member of GreyNet. Email: JKaplan@nyam.org

Lawrence, Amanda

45

Amanda Lawrence is managing editor of Australian Policy Online (apo.org.au) and Creative Economy (creative.org.au) at the Institute for Social Research, Swinburne University of Technology. She also manages the website of the ARC Centre of Excellence for Creative

Industries and Innovation (CCI). Amanda created the information architecture and oversaw the upgrades for APO and CCI websites in 2009. She recently completed a Graduate Diploma of Library and Information Management at RMIT. Amanda has a Bachelor of Arts (Hons) from the University of Melbourne and was Manager of the Literature Program at the University's Asialink Centre from 1996 – 2005. She has also worked as a bookseller and researcher. alawrence@swin.edu.au

Marzi, Claudia

19

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 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

Myohanen, Lea

41

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

19

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

Sassi, Manuela

19

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

5

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 GERiCO 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

Notes for Contributors

Non-Exclusive Rights Agreement

- I/We (the Author/s) hereby provide TextRelease (the Publisher) non-exclusive rights in print, digital, and electronic formats of the manuscript. In so doing,
- I/We allow TextRelease to act on my/our behalf to publish and distribute said work in whole or part provided all republications bear notice of its initial publication.
- I/We hereby state that this manuscript, including any tables, diagrams, or photographs does not infringe existing copyright agreements; and, thus indemnifies TextRelease against any such breach.
- I/We confer these rights without monetary compensation and with the understanding that TextRelease acts on behalf of the author/s.

Submission Requirements

Manuscripts should not exceed 15 double-spaced typed pages. The size of the page can be either A-4 or 8½x11 inches. Allow 4cm or 1½ inch from the top of each page. Provide the title, author(s) and affiliation(s) followed by your abstract, suggested keywords, and a brief biographical note.

A printout or PDF of the full text of your manuscript should be forwarded to the office of TextRelease. A corresponding MS Word file should either accompany the printed copy or be sent as an attachment by email. Both text and graphics are required in black and white.

REFERENCE GUIDELINES

General

- i. All manuscripts should contain references
- ii. Standardization should be maintained among the references provided
- iii. The more complete and accurate a reference, the more guarantee of an article's content and subsequent review.

Specific

- iv. Endnotes are preferred and should be numbered
- v. Hyperlinks need the accompanying name of resource and date; a simple URL is not acceptable
- vi. If the citation is to a corporate author, the acronym takes precedence
- vii. If the document type is known, it should be stated at the close of a citation.
- viii. If a citation is revised and refers to an edited and/or abridged work, the original source should also be mentioned.

Examples

Youngen, G.W. (1998), Citation patterns to traditional and electronic preprints in the published literature. - In: College & Research Libraries, 59 (5) Sep 1998, pp. 448-456. - ISSN 0010-0870

Crowe, J., G. Hodge, and D. Redmond (2010), Grey Literature Repositories: Tools for NGOs involved in public health activities in developing countries. – In: Grey Literature in Library and Information Studies, Chapter 13, pp. 199-214. – ISBN 978-3-598-11793-0

DCMI, Dublin Core Metadata Initiative Home Page http://purl.oclc.org/metadata/dublin_core/

Review Process

The Journal Editor first reviews each manuscript submitted. If the content is suited for publication and the submission requirements and guidelines complete, then the manuscript is sent to one or more Associate Editors for further review and comment. If the manuscript was previously published and there is no copyright infringement, then the Journal Editor could direct the manuscript straight away to the Technical Editor.

Journal Publication and Article Deposit

Once the journal article has completed the review process, it is scheduled for publication in The Grey Journal. If the Author indicated on the signed Rights Agreement that a preprint of the article be made available in GreyNet's Archive, then browsing and document delivery are immediately provided. Otherwise, this functionality is only available after the article's formal publication in the journal.

An International Journal on Grey Literature

'TRANSPARENCY IN GREY LITERATURE'

SPRING 2011 – TGJ VOLUME 7, NUMBER 1

Towards a Prague Definition of Grey Literature	5
<i>Joachim Schöpfel (France)</i>	
A Terminology Based Re-Definition of Grey Literature	19
<i>Claudia Marzi, Gabriella Pardelli, and Manuela Sassi (Italy)</i>	
Peering through the Review Process: Towards Transparency in Grey Literature	25
<i>Dominic J. Farace (The Netherlands)</i>	
Scientific Data: Increasing Transparency and Reducing the Grey	33
<i>Bonnie C. Carroll, June Crowe, and J.R. Candlish (United States)</i>	
Global Grey Literature in Health: Identification, Preservation, and Sustainability	41
<i>Janie Kaplan, Lea Myohanen, and Elizabeth Taylor (United States)</i>	
A Bridge over Grey Waters: Collecting and Preserving Public Policy Research ..	45
<i>Amanda Lawrence (Australia)</i>	

Colophon.....	2
Editor's Note.....	4
Advertisement	
De Gruyter Saur, <i>Grey Literature in Library and Information Studies</i>	24
IIA, Information International Associates, Inc.	32
NYAM, The New York Academy of Medicine.....	40
EBSCO Information Services.....	55
On the News Front	
Thirteenth International Conference on Grey Literature – Call for Papers.....	56
The Grey Circuit, From Social Networking to Wealth Creation – Registration.....	57
About the Authors.....	58
Notes for Contributors.....	59