https://doi.org/10.26069/greynet-2025-000.542-tgj
ISSN 1574-1796

An International Journal on Grey Literature

Spring 2015, TGJ Volume 11, Number 1

'PUBLIC AWARENESS TO GREY LITERATURE'

www.textrelease.com Grey Literature Network Service www.greynet.org

https://doi.org/10.26069/greynet-2025-000.542-tgj

The Grey Journal

An International Journal on Grey Literature

COLOPHON

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 11, Number 1 2015. - *CVTISR, DANS, EBSCO, FEDLINK-Library of Congress, Inist-CNRS, ISTI-CNR, KISTI, NIS-IAEA, NTK, and NYAM are Corporate Authors and Associate Members of GreyNet International. This serial publication appears three times a year - in spring, summer, 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:
€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/publications.html

About TGJ

The Grey Journal is a flagship journal for the international 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.

Contents
'Public Awareness To Grey Literature'

https://doi.org/10.26069/greynet-2025-000.542-tgj

A Slide Rule Approach to Quality Issues in Grey Literature*

When asked to write a short article on quality issues dealing with grey literature, my first reaction was to turn to an article I had written in 2011 dealing with the review process of grey literature compared with that of commercially published literature[1]. However, while peer review certainly is of importance to grey literature there are also other issues that determine its quality both in content and form. Hence, upon second thought I set about to identify key issues that influence the quality of grey literature and bring these together on a continuum ranging from pre-publication to publication onto post-publication. While neither the list of variables compiled for this purpose is in any way exhaustive, neither is the discussion given them in this article. Instead, a framework is constructed that offers a cognitive approach to a rather complex and still explorative area of grey literature.

PRE-PUBLICATION PHASE	PUBLICATION PHASE	POST-PUBLICATION PHASE
policy … standards … review … education	*policy … standards … review … education*	*policy … standards … review … education*
• Source data • Document type • …	• Metadata • References • …	• Open Access • Preservation • …

Diagram: Slide rule approach to quality issues in grey literature

While the diagram above provides a rather systematic approach to quality assessment, it is not static. In other words, issues or variables identified by the bullet points in one of the three phases of publication are not strictly bound to a given time sequence. For example, the issue of preservation listed in the post-publication phase may already be anticipated during pre-publication, but only carried out in post-publication. Another aspect in this slide rule approach is that with the introduction of the middle range, key issues of policy, standards, review, and education are shown to run throughout all three phases of publication. In fact, it is this middle range that demonstrates the dynamics of the cognitive instrument allowing the user a fuller understanding, assessment, and weight that one assigns to grey literature.

If we consider the review of grey literature, one of the key issues in the middle range of the diagram, then we have to address how this is carried out in all three phases of publication. The peer review model applied in various degrees by commercial publishers – primarily with journal articles – is not the only model suited for grey literature. Other models have been introduced which extend beyond the pre-publication phase and which deserve the attention of grey literature publishers and corporate authors. Two of which are the Open Peer Review Model[2] and the Peer-to-Peer Review Model[3]. Further considerations in this regard are that grey literature document types far exceed journal articles and thus require alternative models for review. Also, community sourcing lends itself to the review of grey literature in which its range and focus contributes more to the process of review, than to any one particular publication.

* Mesurer la qualité de la littérature grise / Dominic Farace ; traduction Joachim Schöpfel. – In: i2D - Information, données et Documents : Pratiques & Recherche, p.60-61, N° 1, Mars 2015. - (Text drafted in English, then translated and first published in French)

If we were to address one of the issues cited in the lower range of the diagram, for example the quality of metadata, we would have to consider the policy, standards, and review process that fuel and sustain metadata. And, this in some way, would then have to be correlated with how well informed or educated the user is in grey literature.

As mentioned earlier, determining or assessing the quality of grey literature is complex and still an explorative area in information science. On the bottom line, it relies on community sourcing. And for this purpose, the GreyGuide, Guide to Good Practices and Resources in Grey Literature[4] was recently constructed – intent upon serving information professionals, practitioners, and web users on an open access platform.

Dominic Farace
journal@greynet.org

[1] Farace, D..J. (2011), Peering through the Review Process: Towards Transparency in Grey Literature, http://www.opengrey.eu/item/display/10068/700007

[2] 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

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

[4] GreyGuide, Guide to Good Practices and Resources in Grey Literature, http://greyguide.isti.cnr.it/

https://doi.org/10.26069/greynet-2025-000.542-tgj

How Information in Grey Literature Informs Policy and Decision-Making: A Perspective on the Need to Understand the Processes

Bertrum H. MacDonald, James D. Ross, Suzuette S. Soomai, and Peter G. Wells (Canada)

Abstract

Effective advocacy for grey literature must be based on understanding the environments in which it is used. As advances in communications technologies continue to occur at a breath-taking pace, all forms of information are being affected. Evolving publication practices are presenting new communication opportunities, in addition to disruptions of established patterns, as long-standing genres are being reshaped by powerful technological and societal changes. Disruptions can cause discomfort and anxiety, but opportunities to promote the value of particular information genres also arise. Grey literature, for example, continues to be produced in large quantities, which suggests that its importance in communication may be increasing rather than diminishing. Advocates of grey literature may believe this genre is undervalued or misunderstood, but lobbying for grey literature in the absence of understanding the contexts in which it is or can be used will likely fail unless information activity in those settings is understood. One prominent context encompasses public policy and decision-making where grey literature is often present but typically unnoticed. Policy and decision-making are complex processes and increasing attention is being placed on developing an understanding of the science-policy interface and evidence-based policy making in particular. Conferences (e.g., Science Advice to Governments, Auckland, New Zealand, August 2014), evidence information services (e.g., one launched in the United Kingdom in 2014), research programs and institutes (e.g., Environmental Information: Use and Influence, Dalhousie University), and a considerable body of literature emphasizes the importance of understanding the relationship between research and policy. Drawing on findings from research conducted within the Environmental Information: Use and Influence program, which involves governmental, intergovernmental, and non-governmental organizations, we outline roles for grey literature in policy and decision-making contexts. We describe types of grey literature used in these contexts, identify preferences for specific features of useable information by managers and policy makers, and outline pathways of research evidence, some of which is produced as grey literature.

To be effective, advocacy for grey literature must be based on understanding the environments in which this literature is generated and used. A recent case about hydraulic fracturing, a controversial subject in North America and Europe (see for example, "Hydraulic Fracking," 2014), illustrates this point. After months of public consultations, publication of numerous open access papers, and extensive study, an independent review panel on hydraulic fracturing (fracking) appointed by the Province of Nova Scotia published its final report on the subject in August 2014 (Wheeler et al., 2014). Five days after the release of the report, the provincial government announced that it would introduce legislation "to prohibit high volume hydraulic fracturing for onshore shale gas" (Government of Nova Scotia, 2014). In a front page news story two weeks after the government passed the legislation to enforce the ban, David Wheeler, chair of the review panel, stated categorically that the panel's report was "very clear in not suggesting a moratorium" (McLeod, 2014). From Wheeler's perspective the government's opposition to fracking was "for the most part, not based on scientific evidence." He claimed "science doesn't cut it in the public-policy realm….Instead, a combination of social and political factors…come together to determine how a community assesses risk," which "includes everything from knee-jerk fear to an understandable skepticism of the government's track record of environmental management." Wheeler believes that "these views can change with research and discussion" (McLeod, 2014).

The relevance of this incident to the topic at hand, i.e., emerging and changing markets for grey literature and public awareness of this genre, relates not to the subject of hydraulic fracturing *per se* but rather to the purpose of the review report, its audiences (both expected and unintended), and how its use can inform discussion about how to lobby for or highlight the significance of information published as grey literature. The final report by that independent panel, and the ten discussion papers released earlier in 2014, are all examples of grey literature. Such documents are produced in the thousands by governments and their myriad agencies globally each year. The final report was designed primarily to inform the provincial government, although some might argue that politics rather than the evidence held sway in the government's decision announced very quickly after the report was released. The report served purposes other than simply to inform the government. Over several

months, scientists, engineers, environmental groups, First Nations (aboriginal) groups, industry, and the public, principally the interested public, offered their views on the discussion papers, and media coverage increased awareness of the panel's work. Many people encountered and engaged with the grey literature produced by the panel. Moreover, as the documentation was made available primarily via an online website, residents and non-residents alike could, and did, stay abreast of the work of the panel and contributed to the panel's understanding of the science, engineering, economic potential, and public interest in the subject.

This Nova Scotian example demonstrates that information in grey literature and current information technologies facilitated an assessment and consultation process that is characteristic of modern policy making. Different types of grey literature were used and they served different purposes. However, it is likely that most who were involved with the process were unaware that a category of information that we call grey literature fulfilled an important role in the consultation process, or that, in the absence of such material, the ability to become informed about the issues and participate in the debates would have been very difficult to achieve. Because of the generally transparent nature of the consultative process, made possible by social media, web-based technologies, and media coverage, the Minister of Energy in the provincial government was likely aware of public reactions and responses to the panel's consultation activities prior to receiving the final report. He would not have been aware of the panel's conclusions and recommendations before receiving the report, but he would have known about vocal opposition to fracking. While the government acknowledged receipt of the report, and in fact requested it, it is believed that advocacy for grey literature or the merits of the scientific advice it contained would not have given this report any more authority at the political level in this decision-making context. By identifying and understanding the variety of steps in this consultation and decision-making process, the place of information presented as grey literature and its use by numerous individuals can be appreciated. If advocacy for this literature is needed, then its promotion is likely to be more effective by elucidating the life cycle of such reports.

Advocates of grey literature may believe that this genre is undervalued or misunderstood, and evidence suggests that is the case (McKimmie & Szumak, 2002; Ravindranath, 2010; Thaler, 2010). Advocacy for grey literature can be affected by confusion about what actually constitutes this genre and whether it is reliable (credible). Grey literature experts may assume that the genre is clearly demarcated, but many information users do not recognize the distinction or consider it important.

Grey Literature in Policy and Decision-Making Contexts – Some Challenges

Grey literature is commonly used in policy and decision-making contexts, but it may not be noticed or recognized as such. In these contexts, attributes of salience, credibility, and relevance can far outweigh peculiarities of definition (McNie, 2007). In a recent review, an author questioned why the reports produced by the Intergovernmental Panel on Climate Change were labelled as grey literature. After all, the reviewer pointed out, these reports "are probably some of the most thoroughly peer-reviewed publications in the world" (personal communication, October 2014). This reviewer simply did not view peer-reviewed technical reports as grey literature even though that individual acknowledged that the reports are "intergovernmental documents."

Within policy contexts, requested or contracted reports may be preferred and relied upon because their credibility can be confirmed, the relevance of the information they contain matches the expectations of readers, and the scale (local or regional versus national or global) fits the requirements for the decisions that need to be made. Moreover, the networks in which the policy specialists, managers, and/or scientists operate provide the assurance that the information is credible. These practitioners may not realize that this information is grey literature even though they may work with it on a daily basis. The term "grey literature" may unintentionally convey negative connotations in policy contexts, a point that advocates for the genre must take into account.

Public policy and decision-making can have different manifestations according to particular political regimes or governance models, a point that merits emphasis (Liverani, Hawkins, & Parkhurst, 2013). For example, recognizing the similarities and differences in the governance structures of countries, e.g., as regards public consultation, can be important for understanding the role(s) that grey literature fulfills and ultimately how the relevance and importance of this literature could be promoted.

Numerous studies (e.g., Lawrence, Houghton, Thomas & Weldon, 2014; Luzi, 2000; Schöpfel & Farace, 2010; Thelwall, Klitou, Verbeek, Stuart, & Vincent, 2010; Webster & Collins, 2005) have pointed to the growing production of information, often as grey literature, by government departments, international inter-governmental bodies, and non-governmental organizations, resulting in greater reliance on this type of publication for disseminating information and assisting with policy decisions. The quantity of information is now very large but the proportion that is published or otherwise distributed as grey literature is unknown. The number of scholarly documents available on the web, not including grey literature, was recently estimated at 114 million (Khabsa & Giles, 2014; see also Bornmann & Mutz, 2014 and Van Noorden, 2014). Another recent study estimated that a new journal article is published every 20 seconds (Munro, 2013). *Nature* reports that 47 x 10^3 primary journals currently exist adding more that 10^6 new articles per year to the primary literature. With the addition of grey literature, the number of available publications increases markedly. As of five years ago, 95% of the web, composed of over 220 billion pages, was still to be indexed by search engines (Scheeran, 2012). The growth of the web over the past five years now makes it very difficult to estimate the extent of web-based information, hence it may be impossible to determine the quantity of existing grey literature (BrightPlanet, 2014).

Unfortunately, the massive volume of new web-based information has affected abilities of decision makers to discover relevant, timely information. This situation merits a question: what proportion of the grey literature component of the information universe is worth paying attention to and, therefore, merits efforts to increase its visibility to search engines by enhanced information architecture and embedded metadata? Examination of information activities in decision-making contexts, including how people deal with information overload (Bawden & Robinson, 2009; Parsons, 2010), could provide grey literature advocates with insights to address this question.

Science-Policy Interface(s)

Since 2002, the Environmental Information: Use and Influence (EIUI) research program at Dalhousie University has strived to understand the use and influence of marine environmental and fisheries information and the processes at work at the science-policy interface (EIUI, 2014). Results from our recent studies support the view that an appreciation of the context(s) in which grey literature is used will inform efforts to recognize and promote its value.

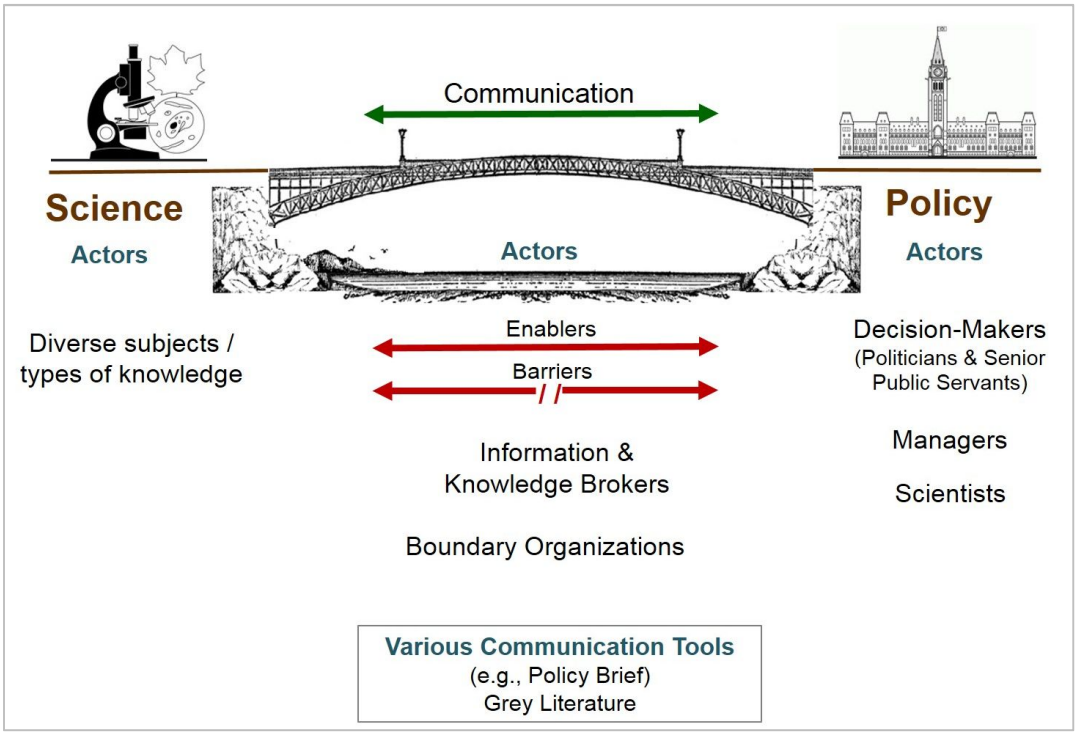


Figure 1. Science-Policy Interface(s)

Figure 1 provides a schematic illustration of our understanding of the science-policy interface(s). The "information universe" is multidimensional and information flow is often non-linear, leading to the complexity of activity at the science-policy interface. However, in its most simple form, the interface is

characterized by information production (science) and information use (policy) with often bi-directional communication between these two broad categories. In reality, the science-policy interface is more inclusive than the label implies, as it encompasses many social processes. In fact, numerous interfaces exist due to many decision-making contexts in which issues, information types, actors, jurisdictions, and other factors change (e.g., Sarkki et al., 2014). Actors or individuals at the interface compose both the "science" and "policy" arenas and they influence the production of and preferences for information types, especially grey literature. Bridging these two are a suite of additional actors who serve in various communication roles, such as editors, publishers, and distributors, now functioning in mostly digital capacities. Numerous enablers and barriers to communication can influence information flow and in some instances a particular factor, e.g., terminology used in research, or sensitivity to the topic, may enable or impede communication.

A diversity of subjects and types of knowledge contribute to the multiplicity of information products generated by the "science" component in Figure 1. Besides the natural and physical sciences, the social sciences and humanities, as well as traditional knowledge, may contribute significantly to the pool of information that can, and often should be, considered in decision making. Within government, policies themselves can be administrative, legislative, or regulatory. Various types of individuals participate in the policy and decision-making bodies, e.g., politicians and senior public servants, among whom are managers and scientists who conduct research within the mandates of government departments and agencies. Furthermore, actors at the science-policy interface can include information and knowledge brokers whose synthesis and communication skills enable them to translate the information generated by either the science or policy communities into a form suitable for either one. Some organizations, such as non-governmental organizations, are particularly effective as boundary organizations at this naturally complex interface. All of the individuals who work and pursue careers in the various components of this interface operate within a diversity of world views, educational capacities, and constraints that affect the production, distribution, reception, and use of information. Communication can be affected, and is often complicated by many converging factors. As a consequence, various tools have evolved to promote communication among the array of actors fulfilling the roles found at this interface.

Measuring Use of Information Available as Grey Literature

While grey literature is produced in large quantities by many organizations and is frequently used in decision-making contexts, the use of this large body of information in policy development is deemed extremely important but poorly understood (Ascher, Steelman, & Healy, 2010; Briggs & Knight, 2011; Holmes & Lock, 2010; Likens, 2010; McNie, 2007; Stojanovic, Ball, Ballinger, Lymbery, & Dodds, 2009). Surprisingly few organizations have undertaken an analysis of the use of their publications, and information pathways in decision-making contexts are still being elucidated (Economic Commission for Europe, 2003; MacDonald, Cordes, & Wells, 2004; Soomai, MacDonald, & Wells, 2011a, 2011b; Wells, 2003; Wells, Duce, & Huber, 2002).

As stated above, information use in policy and decision-making contexts is a complex phenomenon. It operates at various levels of scale: geographic, institutional, political, and temporal. Most models of scientific communication ignore the use of research information in public policy making, where information assimilation is different than in pure research contexts (Doern, 2001; Duff, 1997; Dunn, 2005; Søndergaard, Andersen, & Hjørland, 2003; Van der Veer Martens & Goodrum, 2005). Decision makers or their advisors may find it challenging to choose which information to use when there is an absence of consensus among scientific experts and the presence of competing views advocated by stakeholders. In policy settings, information in grey literature may be given greater importance than that in peer-reviewed journals because the language is more accessible and a more rapid and flexible delivery of relevant information can facilitate knowledge diffusion where decisions are based on competing factors, e.g., the pressures of political processes (Bremer & Glavovic, 2013; Pielke, 2007; Shanley & López, 2009). However, as open-access, peer-reviewed journals proliferate, papers published in these venues are becoming more easily accessible to decision-makers than has been the case previously (e.g., PLOS – Public Library of Science), which may affect the future mix of grey and primary literature used in decision-making contexts.

While information in grey literature may have a direct influence on the development of a policy, this type of "use" is only one of several. "Use" of information can, in fact, extend in meaning from general awareness, to increased understanding, to a change in attitude about a subject, or to actual implementation of the information in practice or policies (Figure 2) (Nutley, Walter & Davies, 2007;

Young, 2014). This spectrum of possible uses lays out a challenge to those wishing to promote the value and use of grey literature. An example drawn from one of our research projects illustrates this point.

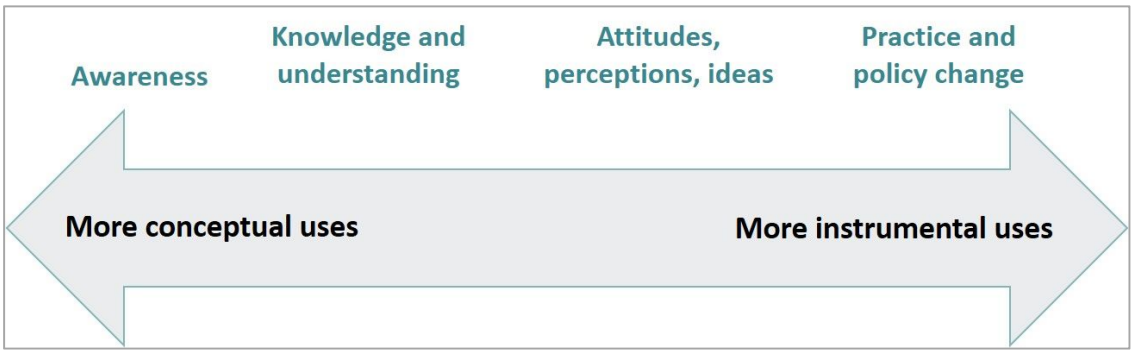


Figure 2. Continuum of Information Use (Nutley, Walter, & Davies, 2007)

The Gulf of Maine Council on the Marine Environment is an American-Canadian intergovernmental organization consisting of representatives of three American states, two Canadian provinces, and the federal governments of both countries, as well as representatives from selected non-governmental organizations. This organization has produced several hundred publications over its history related to its mandate to maintain and enhance environmental quality in the Gulf of Maine and Bay of Fundy (MacDonald, Cordes, &Wells, 2004; Ross, Hubbard, Cordes, MacDonald, & Wells, 2014). In one of our studies, we interviewed members of the main Working Group of the Council about their understanding of the production and use of the Council's publications (Cossarini, MacDonald, & Wells, 2014). A member of that Working Group said the following about one of the publications:
We have a five or six page handout on [the] American Eel…and its status in the Gulf of Maine. So, I gave that to our pelagics advisor, fisheries advisor, and he found it very informative, a good synopsis and was quite impressed with it…now he never told me … that changed how [he] recommended to the Minister what our position is on it…but he found it informative in terms of getting his knowledge, in terms of playing that role as an advisor to the Minister of the issue.

This manager's description about the "use" of the handout, a fact sheet on the American Eel, underscores the different aspects of information "use." Both the manager and the fisheries advisor were clearly aware of the fact sheet. That awareness led to greater understanding about this fish species; however, whether the increased understanding led to direct implementation of the information in policies revised or developed by the particular department of government is not revealed. Nonetheless, the manager's observations illustrate three types of information use. Although only one example, the manager's comments also show that advocates of the use of grey literature cannot assume that simply drawing such literature to the attention of potential users is sufficient. Each stage of use (awareness, understanding, change in attitude, and implementation) may require different promotional techniques, or a focus on different features of the information so that how it might be used becomes obvious to a targeted recipient.

Types or Formats of Grey Literature

Experts on grey literature are well aware of the wide variety of types of information formats and products that constitutes this genre: internal policy documents, briefing documents, technical reports, contract or consultants' reports, internal reviews, fact sheets, infographics and posters, conference proceedings, newspapers, monographs, websites, social media posts, and a host of others. Acknowledgment that there is an array of types may lead to overlooking the role(s) that each fulfills and the need for different strategies for raising awareness and use of each type. Furthermore, given ongoing constraints on resource allocation and the vast quantity of grey literature, efforts to improve its use will benefit from understanding the different functions.

Briefing notes and memos, for example, are frequently used in public policy and decision-making contexts (see Figure 3). Policy briefs are frequently used to convey information to politicians and decision makers in many, if not most, government organizations. This type of document, while very common, is often not readily accessible, due to confidentiality and the short life span of the documents. Few studies have investigated the effectiveness of this form of communication (see Beynon, Gaarder, Chapoy, & Masset, 2012; Masset, Gaarder, Beynon, & Chapoy, 2013), although two of us (Soomai and Wells) in previous government positions saw briefing notes being used for their

intended purpose, e.g., briefing a Minister. Interestingly, in contrast one recent study found little evidence that policy briefs actually led to a change in prior beliefs (Rajabi, 2012). In our view, the quantity and wide-spread use of this form of grey literature points to the importance of briefing notes and memos. As stated, our experience in government (Soomai and Wells) showed that briefing notes submitted to managers and decision-makers at various levels were regularly used as an effective way to distill information on a problem and offer options for solutions.

Type	Role
Briefing Notes/Memos	Advice / The "Facts"
Technical Reports	Document Detailed Research Results
Social Media	Brief, Rapid, and Frequent Communication

Figure 3. Selected Types of Grey Literature

Technical reports have been produced by governments and other organizations in very large numbers and in various formats and styles for decades. This type of grey literature has received extensive attention by grey literature experts, librarians, and authors alike. Often, the resources required to produce technical reports can be substantial. State of the environment reports, for example, may entail the commitment of many contributors, editors, illustrators, designers, and publishers, at considerable expense (Wells 2003). Enhancing awareness and use of these reports is merited on several grounds, one of which is accountability for fiscal resource allocation, including the costs associated with not preparing such environmental reports. Other types of technical reports may have a smaller audience and consequently a lower probability of use. As the roles for technical reports vary, the expectations about readership need to be taken into account (very few politicians have the time or the expertise to read lengthy technical reports, for example) and should have a direct impact on decisions about promoting this form of grey literature.

Rapid communication and rapid-fire feedback are not new to the research fields. But now widespread online publishing and social media are changing how people learn about new developments and how people incorporate this information into decision-making processes. The degree to which researchers and decision-makers engage in and monitor social media is also evolving. There is little doubt that this form of communication is taking on increasing importance and that its role merits careful consideration. However, the sheer volume of information flowing through social media presents considerable challenges for deciphering what portion is significant enough to warrant development of strategies for promoting its awareness and use.

Communication and Information Pathways

While understanding the types and roles of grey literature can inform advocacy for this genre, further appreciation of decision-making processes is needed to increase the likelihood that the advocacy will be relevant and effective. Communication patterns are changing due to the prominence of the internet and social media (Cossins, 2014), and other ever advancing information technologies, and this trend will likely continue. Non-governmental organizations have become more active and successful at disseminating scientific information to policy communities at various levels of government (Grainger, 2013). However, "inherent differences between the fundamental structures and traditions of science and policy" (Francis, Whittaker, Shandas, Mills, & Graybill, 2005, p. 35) contribute to sub-optimal "flow of knowledge between researchers, policy makers, and resource managers" (Roux, Rogers, Briggs, Ashton, & Sergeant, 2006, p. 5). Therefore, tracing out information pathways that identify the points where grey literature plays a role is important. For example, information pathway models have

informed our recent marine environmental and fisheries research (as seen in Avdic, 2013; Cano Chacón, 2013; Cossarini, MacDonald, & Wells, 2014; Hutton, 2009; Soomai, Wells, & MacDonald, 2011).

In the first pathway model (Figure 4), the production and movement of information among five significant groups in fisheries activities in the Caribbean was studied, namely, industry, scientists, fisheries managers, advisory bodies, and policy makers (Soomai, Wells, & MacDonald, 2011). Identifying which information was assembled, produced, and used and in what format by each group provided an opportunity to determine where information flow and use were working well and where barriers to information exchange and use existed. This information flow model equips an advocate for grey literature with an understanding of where promotion of the value of this literature needs strengthening and also what information formats are preferred by each group. For example, while some members of this specific fishing industry can read and understand technical reports, most cannot. Thus, this model shows less use of technical reports by the fishing industry. It is quite likely that efforts to promote use of this literature by this group would be unsuccessful.

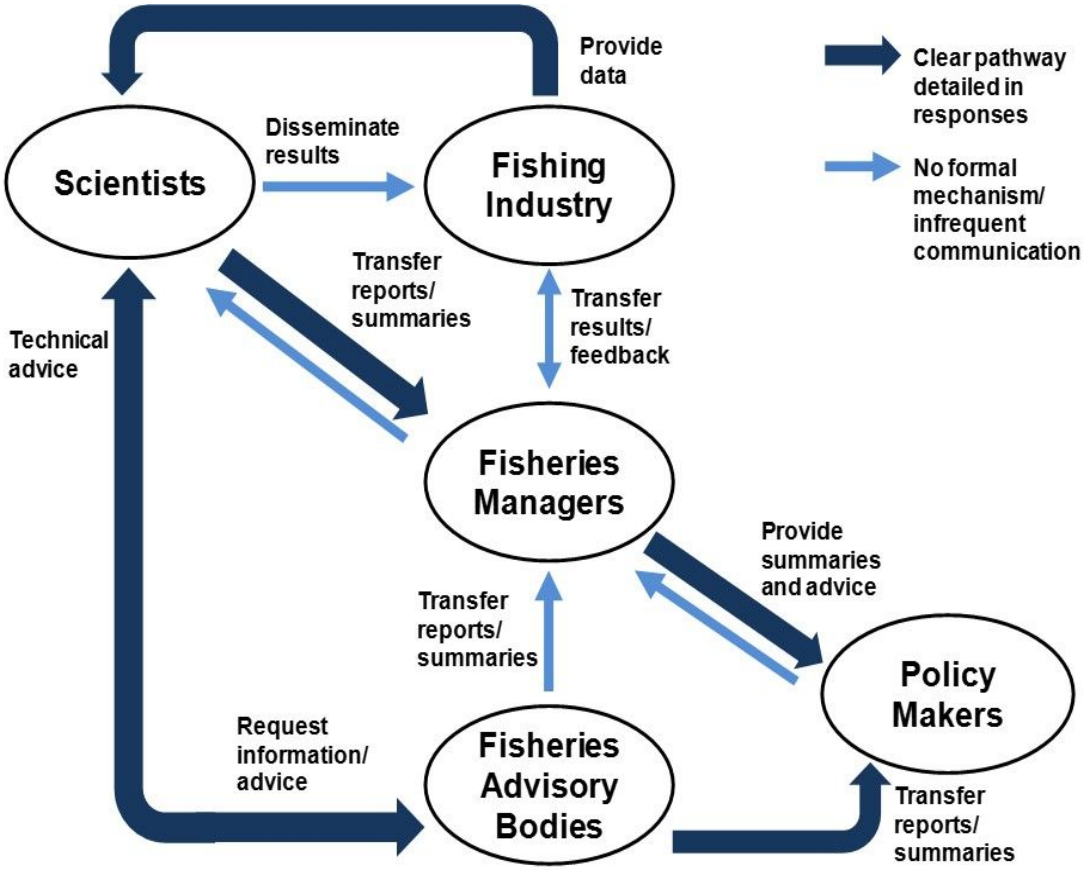


Figure 4. Information Flow: Fisheries Technical Reports
(Soomai, Wells, & MacDonald, 2011)

In the second pathway model (Figure 5), the production and movement of information in the certification of sustainable fisheries by the international Marine Stewardship Council (MSC) process are set out (Cano Chacón, 2013). This process results in labelling consumer products when a fishery is determined to be sustainable. Information as grey literature is found at a number of points in the decision-making process. Learning that the major reports are typically only produced in English and made accessible primarily by posting to the Council's website enhances an understanding of how to tailor advocacy tactics. For example, simply drawing attention to the existence of the reports is not likely to increase use by small fishers in countries where the MSC certification process is being applied.

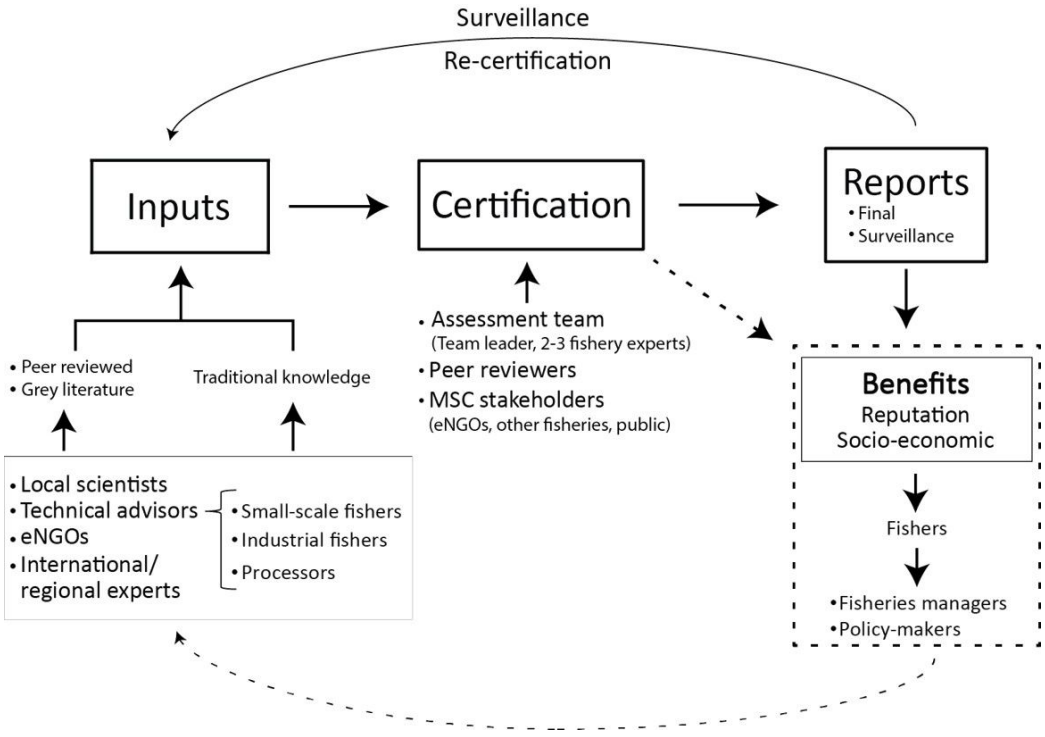


Figure 5. Information Flow: Marine Stewardship Council (Cano Chacón, 2013)

Conclusions

As illustrated by the Nova Scotian example given earlier, information published as grey literature can fill an important role in policy-making contexts. Promotion of the value of such literature, in our view, is merited, particularly when it offers evidence that will help to resolve societal problems. In an age of information abundance, or even information excess, many factors affect which if any information resources will be used to inform decision-making. As well, use itself is a complex phenomenon that is closely related to context. Published evidence clearly competes with other factors that are brought to bear in decision making.

Lobbying for the value of grey literature should not begin with the assumption that the value is obvious to potential users. Rather, advocacy should begin from a position informed by an understanding of the range of settings in which it is used and with the various audiences or readers taken into account. The role that grey literature fulfils must be viewed within the larger background of information production and communication where the relative importance of information in grey literature versus information in primary literature can and will vary. While understanding the types and roles of grey literature can inform advocacy for this genre, appreciation of decision-making processes is needed to increase the likelihood that the advocacy will be meaningful and effective. Further research on information use is needed. With increased understanding resulting from such study, advocacy for grey literature will be better informed.

References

Ascher, W., Steelman, T., & Healy, R. (2010). *Knowledge and environmental policy*: *Re-imagining the boundaries of science and politics.* Cambridge: The MIT Press.

Avdic, V. (2013). Measuring use and influence: An assessment of FAO's flagship report *The State of world fisheries and acquaculture*. (Unpublished Master's thesis). Dalhousie University, Canada.

Bawden, D., & Robinson, L. (2009). The dark side of information overload, anxiety, and other paradoxes and pathologies. *Journal of Information Science, 35,* 180-191.

Beynon, P., Gaarder, M., Chapoy, C., & Masset, E. (2012). Passing on the hot potato: Lessons from a policy brief experiment. *IDS Bulletin, 43*(5), 65-75.

Bornmann, L., & Mutz, R. (2014). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Retrieved from http://arxiv.org/abs/1402.4578v2.

Bremer, S., & Glavovic, B. (2013). Mobilizing knowledge for coastal governance: Re-framing the science-policy interface for integrated coastal management. *Coastal Management, 41*, 39-56.

Briggs, S.V., & Knight, A.T. (2011). Science-policy interface: Scientific input limited. *Science, 333*(6043), 696-697.

BrightPlanet Corporation. (November 5, 2014). How big is the internet? *Deep Web University.* Retrieved from http://www.brightplanet.com/2014/11/big-internet/

Cano Chacón, M. (2013). The role of information of the Marine Stewardship Council certification process in developing countries: A case study of two MSC fisheries certified in Mexico. (Unpublished Master's thesis). Dalhousie University, Canada.

Cossarini, D. M., MacDonald, B. H., & Wells, P. G. (2014). Communicating marine environmental information to decision makers: Enablers and barriers to use of publications (grey literature) of the Gulf of Maine Council on the Marine Environment. *Ocean & Coastal Management, 96,* 163-172.

Cossins, D. (2014, October 1). Setting the record straight. *The Scientist.* Retrieved from http://www.the-scientist.com/?articles.view/articleNo/41056/title/Setting-the-Record-Straight/

Doern, G. B. (2001). Science and scientists in regulatory governance: A mezzo-level framework for analysis. *Science and Public Policy, 28,* 195-205.

Duff, A. S. (1997). Some post-war models of the information chain. *Journal of Librarianship and Information Science. 29*(4), 179-187.

Dunn, K. (2005). Impact of the inclusion of grey literature on the scholarly communication patterns of an interdisciplinary specialty. Paper presented at the Sixth International Greynet Conference, New York Academy of Medicine, December 2004. Amsterdam: TextRelease.

Economic Commission for Europe. (2003). *Environmental policy in transition: Ten years of UNECE environmental performance reviews.* New York: United Nations.

EIUI. (2014). Environmental Information: Use and Influence. Retrieved from www.eiui.ca.

Francis, T. B., Whittaker, K. A., Shandas, V., Mills, A. V., & Graybill, J. K. (2005). Incorporating science into the environmental policy process: A case study from Washington State. *Ecology and Society, 10,* 35. Retrieved from http://www.ecologyandsociety.org/vol10/iss1/art35

Government of Nova Scotia. (2014, September 3). Government to prohibit hydraulic fracturing. Retrieved from http://novascotia.ca/news/release/?id=20140903005

Grainger, R. (2013). How does information influence policy? The role of fishery organizations in policy-making for fisheries. Public Lecture, Dalhousie University, Halifax, 20 September.

Holmes, J., & Lock, J. (2010). Generating the evidence for marine fisheries policy and management. *Marine Policy, 34*, 29-35.

Hutton, G. R. G. (2009). Developing an inclusive measure of influence for marine environmental grey literature. (Unpublished Master's thesis). Dalhousie University, Canada.

Hydraulic fracturing. (2014, November 28). In *Wikipedia*. Retrieved 3 December 2014 from http://en.wikipedia.org/wiki/Hydraulic_fracturing

Khabsa, M., & Giles, C. L. (2014).The number of scholarly documents on the public web. *PLoS ONE, 9*(5), e93949.

Lawrence, A., Houghton, J., Thomas, J., & Weldon, P. (2014, November 17). *Where is the evidence? Realising the value of grey literature for public policy & practice.* Swinburn Institute for Social Research, Melbourne, Australia. Retrieved from http://apo.org.au/research/where-evidence-realising-value-grey-literature-publi-policy-and-practice

Likens, G. E. (2010). The role of science in decision making: Does evidence-based science drive environmental policy? *Frontiers of Ecology and the Environment 8,* 1-9.

Liverani, M., Hawkins, B., & Parkhurst, J. O. (2013). Political and institutional influences on the use of evidence in public health policy. A systematic review. *PLoS One, 8*(10), e77404.

Luzi, D. (2000). Trends and evolution in the development of grey literature: A review. *The International Journal on Grey Literature, 1*, 105-116.

MacDonald, B. H., Cordes, R. E., & Wells, P. G. (2004). Grey literature in the life of GESAMP, an international marine scientific advisory body. *Publishing Research Quarterly, 20*(1), 25-41.

Masset, E., Gaarder, M., Beynon, P., & Chapoy, C. (2013). What is the impact of a policy brief? Results of an experiment in research dissemination. *Journal of Development Effectiveness, 5*(1), 50-63.

McKimmie, T., & Szumak, J. (2002). Beyond grey literature: How grey questions can drive research. *Journal of Agriculture and Food Information, 4,* 71-79.

McLeod, P. (2014, November 26). Nova Scotia fracking report author slams ban. *The Chronicle Herald*. Retrieved from http://www.thechronicleherald.ca/novascotia/1253658-nova-scotia-fracking-report-author-slams-ban

McNie, E. (2007). Reconciling the supply of scientific information with user demands: An analysis of the problem and review of the literature. *Environmental Science & Policy, 10,* 17-38.

Munro, R. (2013, October 4). How much science is there? *Science.* Retrieved from http://www.sciencemag.org.ezproxy.library.dal.ca/site/special/scicomm/infographic.jpg

Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using evidence: How research can inform public services.* Bristol: The Policy Press.

Parsons, P. (2010). Information overload. *New Scientist, 205*(2743), 38-39.

Pielke, R. A. (2007). *The honest broker: Making sense of science in policy and politics.* New York: Cambridge University Press.

Rajabi, F. (2012). Evidence-informed health policy making: The role of the policy brief. *International Journal of Preventive Medicine, 3*(9), 596-598.

Ravindranath, N. H. (2010). IPCC: Accomplishments, controversies, and challenges. *Current Science, 99,* 26-35.

Ross, J. D., Hubbard, L. D., Cordes, R. E., MacDonald, B. H., & Wells, P. G. (2014). *Celebrating 25 years of knowledge on the Gulf. A bibliography of publications of the Gulf of Maine Council on the Marine Environment.* Halifax: Dalhousie University.

Roux, D. J., Rogers, K. H., Briggs, H. C., Ashton, P. J., & Sergeant, A. (2006). Bridging the science-management divide: Moving from unidirectional knowledge transfer to knowledge interfacing and sharing. *Ecology and Society, 11,* 4. Retrieved from http://www.ecologyandsociety.org/vol1/iss1/art/

Sarkki, S., Niemelä, J., Tinch, R., van den Hove, S., Watt, A., & Young, J. (2014). Balancing credibility, relevance, and legitimacy: A critical assessment of trade-offs in science-policy interfaces. *Science and Public Policy, 41,* 194-206.

Scheeren, W. O. (2012). *The hidden web: A sourcebook*. Santa Barbara, California: Libraries Unlimited.

Schöpfel, J., & Farace, D. J. (2010). Grey literature. In *Encyclopedia of Library and Information Science* (3rd ed.) (pp. 2029-2039). London: Taylor & Francis.

Shanley, P., & López, C. (2009). Out of the loop: Why research rarely reaches policy makers and the public and what can be done. *Biotropica, 41,* 535-544.

Søndergaard, T. F., Andersen, J., & Hjørland, B. (2003). Documents and the communication of scientific and scholarly information: Revising and updating the UNISIST model. *Journal of Documentation, 59*(3), 278-320.

Soomai, S. S., MacDonald, B. H., & Wells, P. G. (2011a). The 2009 State of Nova Scotia's Coast Report: *An initial study of its use and influence*. Halifax: Environmental Information: Use and Influence.

Soomai, S. S., MacDonald, B. H, & Wells, P. G. (2011b). The State of the Gulf of Maine Report: *An initial study on awareness, use, and influence of the theme papers*. Halifax: Environmental Information: Use and Influence.

Soomai, S. S., Wells, P. G., & MacDonald, B. H. (2011). Multi-stakeholder perspectives on the use and influence of "grey" scientific information in fisheries management. *Marine Policy, 35,* 50-62.

Stojanovic, T. A., Ball, I., Ballinger, R. C., Lymbery, G., & Dodds, W. (2009). The role of research networks for science-policy collaboration in coastal areas. *Marine Policy, 33,* 901-911.

Thaler, A. (2010, November 15). Shades of gray: Gray literature, peer-review, and the struggle for fisheries management. Retrieved from http://www.southernfriedscience.com/?p=8641

Thelwall, M., Klitkou, A., Verbeek, A., Stuart, D., & Vincent, C. (2010). Policy-relevant webometrics for individual scientific fields. *Journal of the American Society for Information Science and Technology, 61*, 1464-1475.

Van der Veer Martens, B., & Goodrum, A. A. (2005). The diffusion of theories: A functional approach. *Journal of the American Society for Information Science and Technology, 57*, 330-341.

Van Noorden, R. (2014, May 7). Global scientific output doubles every nine years. *Nature.* Retrieved from http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html

Webster, J. G., & Collins, J. (2005). *Fisheries information in developing countries. Support to the implementation of the 1995 FAO code of conduct for responsible fisheries*. Rome: Food and Agriculture Organization of the United Nations.

Wells, P. G. (2003). State of the marine environment reports – a need to evaluate their role in marine environmental protection and conservation. *Marine Pollution Bulletin, 46,* 1219-1223.

Wells, P. G., Duce, R. A., & Huber, M. E. (2002). Caring for the sea – accomplishments, activities, and future of the United Nations GESAMP (the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). *Ocean and Coastal Management, 45*, 77-89.

Wheeler, D., et al. (2014). *Report of the Nova Scotia Independent Panel on Hydraulic Fracturing*. Retrieved from http://energy.novascotia.ca/sites/default/files/Report%20of%20the%20Nova%20Scotia%20Independent%20Panel%20on%20Hydraulic%20Fracturing.pdf

Young, S. P. (Ed.). (2014). *Evidence-based-policy-making in Canada: A multidisciplinary look at how evidence and knowledge shape Canadian public policy*. Don Mills: Oxford University Press.

Mapping Italian Grey Communities: What Is There Beyond The Academy?*

Silvia Giannini, Stefania Biagioni, Sara Goggi, and Gabriella Pardelli (Italy)

1. Introduction

The following title was published on an influential Italian newspaper, *La Stampa*, on November 7, 2013: "Tra i tesori della 'Letteratura Grigia' un'Eneide in napoletano del '600". The article is about the presentation of the "Fondo De Mauro" on the Italian Network of Popular Culture [1]: this fund originates from a private collection built up in several decades and donated to the Network by Tullio De Mauro[i] and his wife Silvana Ferreri in 2011; it is made up of thousands of books, brochures, pamphlets relating to Italian dialects and minority languages.

In the field of healthcare, while browsing on the web, we found the title of a PhD thesis of last year: "Letteratura Grigia nelle meta-analisi delle prove ripartite con scelta casuale degli interventi di sanità". In the legal environment, the title of a Seminar at the University of Siena emerges from the web [2]: "La "centralità" della legge e la letteratura grigia. Profili di politica del diritto in Italia tra Otto e Novecento"[ii]. From these findings, the idea of a survey on the wide variety of grey material available on Italian web portals arose. A first analysis showed that this material is available in different forms and dissemination is carried out through various means such as thematic bibliographies, newspapers articles, various types of documents published in PDF format or simple descriptions on web sites.

The following are a few examples excerpted from some home pages belonging to our corpus:

- ✓ Collection of *grey literature*. The Historical Archive of Women candidates for becoming the repository where the memories about these themes will be stored…[iii]
- ✓ *Grey literature* ... Master copies · Reprints · Unreleased copies · *Grey Literature*. Archives for the history of Education…[iv]
- ✓ *Grey literature*. The high quality brand of parks…[v]
- ✓ … International *grey literature*; national and international legislative data on the topic of drug addiction and related themes; documentary archive ...[vi]
- ✓ Besides literature in German, there is literature in other languages and *grey literature* as well – in particular catalogues of museums and exhibitions…[vii]
- ✓ It is about a few thousand of books, brochures and documents of *grey literature* concerning two topics, Italian dialects and minority languages…[viii]

Given this scenario, this research aims at verifying whether – and eventually how much – the grey literature available on the web is actually structured, accessible or even managed by systems dealing with its organization and aiming at its retrieval and storing. The utmost goal is to build up a map of non-academic communities and their mechanisms for managing, presenting and disseminating this type of material. It is a sort of journey among the streams of the Web, which channel meeting minutes, manifests, fliers, pictures, newspapers articles, journalistic services and audio/video material on various topics. These "grey" products – by conveying basic information about social and popular culture – store, represent and spread knowledge.

2. Materials and Methods

In order to get a specific range of data, we selected a sample data from the Italian Web portals using the following query criteria:

1. Search for >"letteratura grigia" OR "letteratura non convenzionale" OR "documentazione grigia" OR "materiale grigio".
2. Creation of the corpus
3. Analysis of the communities and the grey material retrieved
4. Statistical processing of data

We gathered the

- ✓ Type of Communities and stakeholders
- ✓ Type of production
- ✓ Terminology

* First published in the GL16 Conference Proceedings, March 2015.

and analysed the documents by

- ✓ a) Type of products
- ✓ b) Format
- ✓ c) Availability/Access

We got about 42,000 indexed documents considered as "interesting" with respect to the terms of the query. We have to consider that the list ordered by relevance of URL provided by Google never exceeds 1000 results. In order to have the largest number of results, we used the option proposed by Google about the data visualisation…

…In order to show you the most relevant results, we have omitted some entries very similar to the 390 already displayed. If you like, you can repeat the search with the omitted results included.

The first step was to "skim" the information retrieved by removing the duplicates and the irrelevant information. We obtained results consisting of 800 URLs for building the sample data (Corpus). Data was collected between May and July 2014. Then we extracted the information and built the Corpus by grouping the gathered data in 12 informative classes (Categories) assigned to each URL:

Stakeholder, *Type of Stakeholder*, *Name of Stakeholder*, *Infrastructure/service*, *Field*, *Topic*, *Type of material*, *Format*, *Access*, *Domain*, *Description*, *Terminology*. Finally, an analysis of the results followed.

3. Results

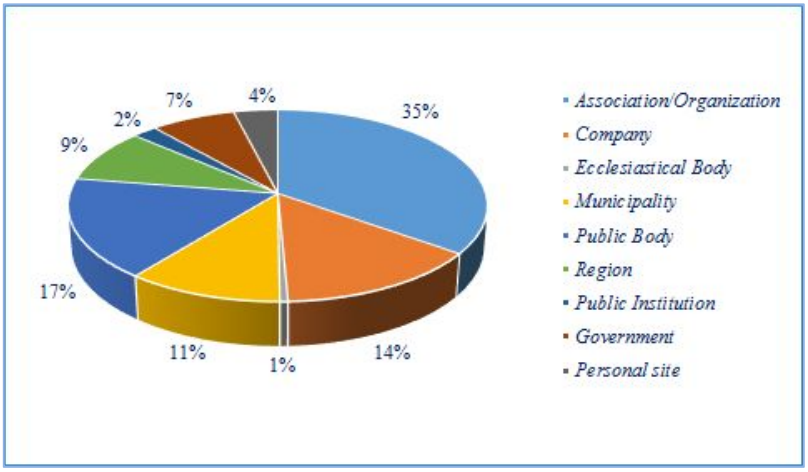
3.1 Stakeholders

The institutional nature of the analyzed stakeholders is either public or private (Graph 1); in addition to a certain amount of personal sites, the highest percentage is represented by websites of associations/organizations as well as public bodies/institutions and companies which indeed have both public and private nature.

In the public sector, the institutional websites are those prevailing: they are those of Ministries, schools, public bodies, local or territorial authorities, healthcare agencies (in Italian, "Azienda Sanitaria" - ASL) whose websites contain detailed information about the structure and the services offered.

There are also thematic sites carried out by several collaborating administrations with a view to a precise purpose such as the presentation of a project or event, the provision of a service or the promotion of a specific field: an example is the portal called "Innovatori PA: la rete per l'innovazione nella Pubblica Amministrazione Italiana" where the title *Letteratura grigia e dati: accessibili, intellegibili e riutilizzabili* (about the advertisement of a 2014 workshop) can be found.

This website does not contain collections of grey literature but aims at disseminating the digital culture pertaining to the grey material, as for instance the sharing of databases for the complete digitalization of Italian public administration[ix]. Amongst the associations/organizations there are professional, healthcare and research associations, Foundations and Consortia; to a small extent, also trade unions and socio-cultural and political groups are represented.



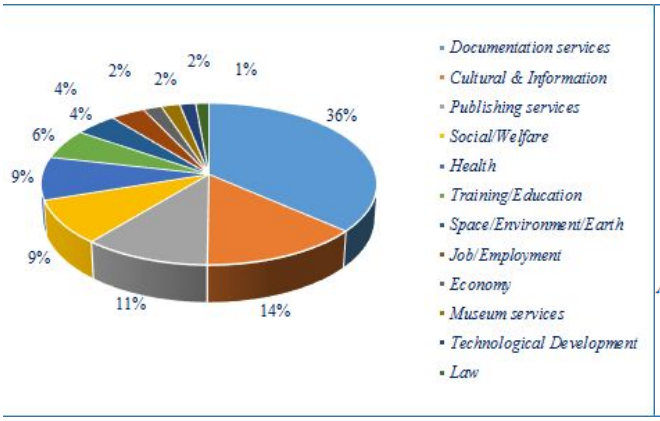
Graph 1 - Stakeholders

In the private sector, there are commercial companies, software houses, publishing houses, banks and management consulting companies, together with personal homepages, fora and blogs dealing with different topics which might have a social relevance such as alcoholism, obesity and violence against women. These stakeholders often restrict themselves to simply present grey literature as citations only but sometimes they do provide interesting editorials and grey literature of various nature.

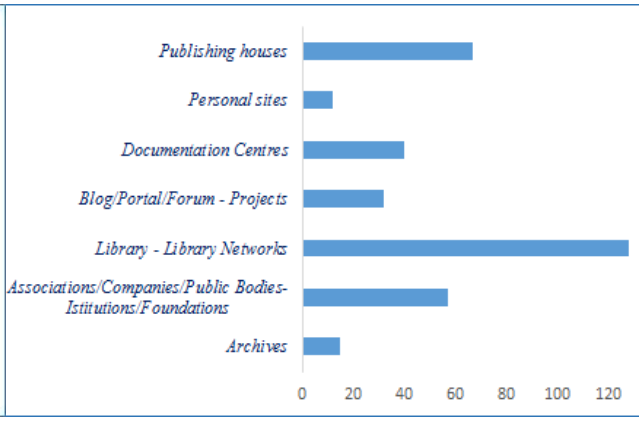
3.2 Thematic areas

Graph 2 shows the fields of application of the stakeholders: documentation services (36%) and cultural/information services (14%) are the areas where grey literature is more easily retrievable compared with fields as Training/Education or Job/Employment which are far more specific. Moreover, the management of grey literature is traditionally assigned to documentation and information services and this area actually embraces a wide variety of sub-fields: from technological development to e-commerce and services for online degree thesis, as well as services for online dictionaries and encyclopedias.

In most of the cases the infrastructure through which the producer manages the grey material is easy to identify: Graph 3 unequivocally shows that in most of the cases libraries and library networks keep the documentation though also archives and documentation centers are sometimes involved in handling this material. It frequently happens that also publishing facilities of associations, companies or institutions maintain this type of material.



Graph 2 – Fields of application



Graph 3 - Infrastructures

In some cases the identity of the supporting infrastructure was not specified and then we assumed that the stakeholder itself was taking care of the grey material, sometimes by creating ad-hoc fora, blogs or portals.

3.3 Communities

Our attention moved towards communities particularly up-to-date and socially relevant. Significant examples have been identified on websites presenting a heterogeneous set of material, which could prove to be especially interesting to researchers, scientists, professionals and simple fans of the various subjects if ever made available and usable.

Several communities handling and disclosing “grey” documentation in various ways come to light from this analysis. The ensemble of information collected allowed to identify and delimit four illustrative communities presenting the following matters: history of women’s culture and of their movement in some specific Italian regions [3]; projects dealing with the sustainability of urban environment with respect to childhood and adolescence [4]; parks and other natural protected environments [5]; archaeological documentation such as draft reports, diaries from the site, letters and miscellaneous documentation [6] [7]; nursing and health-related disciplines which produce guidelines, diagnostic and therapeutic courses, informative material for patients and their families [8]. Also the theatrical culture is nourished by “grey products” as video archives, collections of music LPs and CDs, brochures, scripts, autograph manuscripts (i.e.. letters, correspondences, fliers, musical scores) [9] [10].

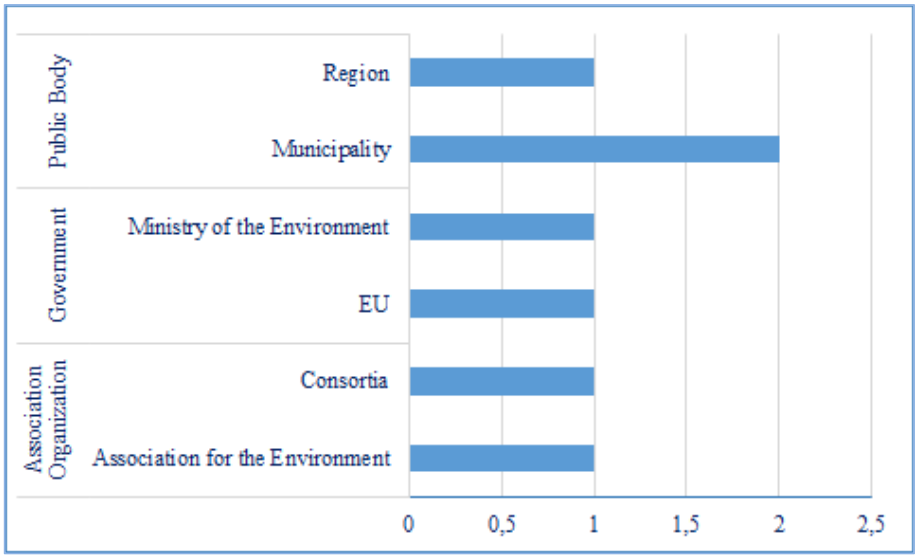
Finally we chose the following communities as case studies of this work: Environment, Women, Healthcare and History.

3.3.1 Environment

Environmental conservation is one of the main issues of today: and many governmental institutions and associations are in charge of preserving the national environmental heritage. Their nature is mostly public, in fact these are several municipalities that take care of managing documentation.

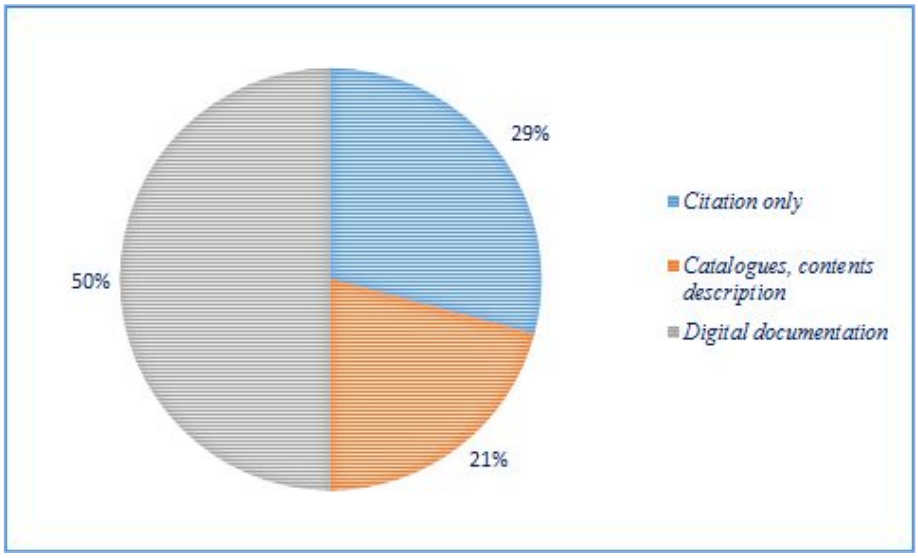
This community deals with the promotion of citizen's environmental awareness and solicits the adoption of adequate politics directly involving the citizenship in the care of the territory.

Several contents related to the environment can be retrieved from the web by means of grey material such as maps, data on city and marine pollution as well as thermal data collection.



Graph 4 – Environment community – Nature of Stakeholders

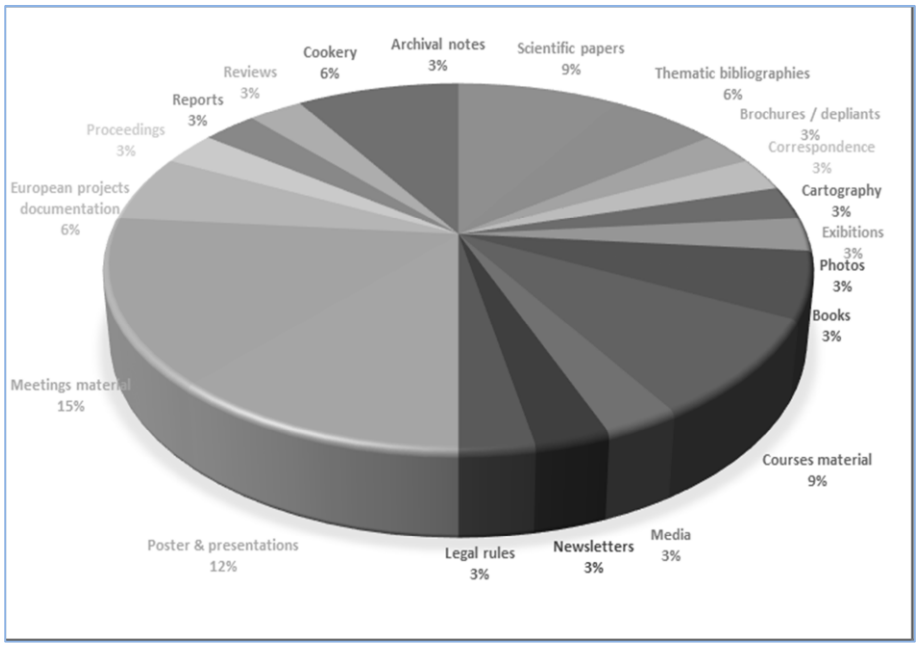
Just to make a few examples, there are websites illustrating interesting projects of eco-sustainable cities for enhancing the quality of life, others – mostly institutional - endorsing some specific geographic areas where peculiar activities are carried on (for instance, mussel farming in the Puglia region in southern Italy) [11].



Graph 5 – Environment community – Document expression

Grey documentation of the environment community is varied: material related to meetings and events (fliers, brochures, etc.) and posters/presentations has the higher percentage but collections of photos and videos as well as thematic bibliographies are remarkable too.

Only in the 21% of the cases the community manages its contents by means of a structured database with search and information retrieval features, therefore an enhancement of content management and then fruition would be highly beneficial.

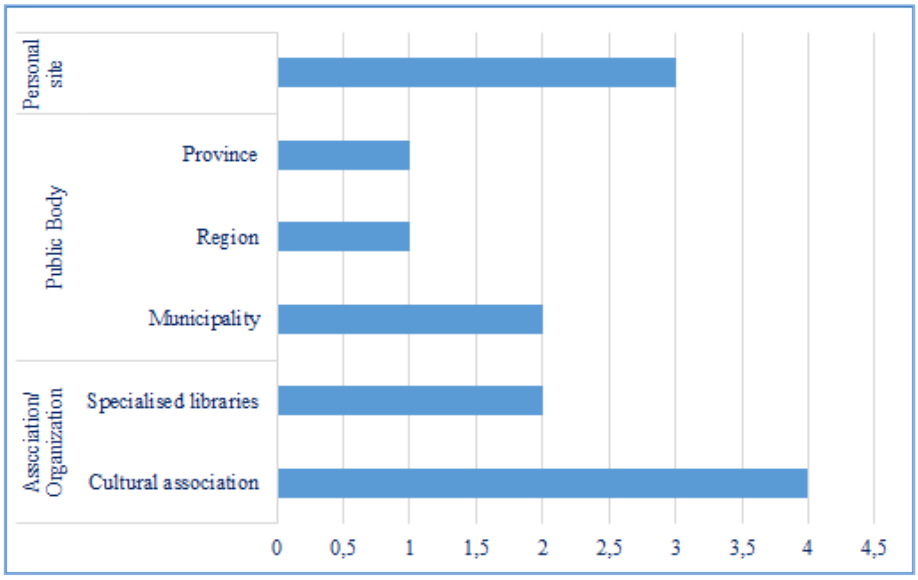


Graph 6 – Environment community – Type of contents

The material is freely accessible and the common format is html (70%) but in several occasions html is used in conjunction with the PDF format so that the latter is used in the 50% of the cases.

3.3.2 Women

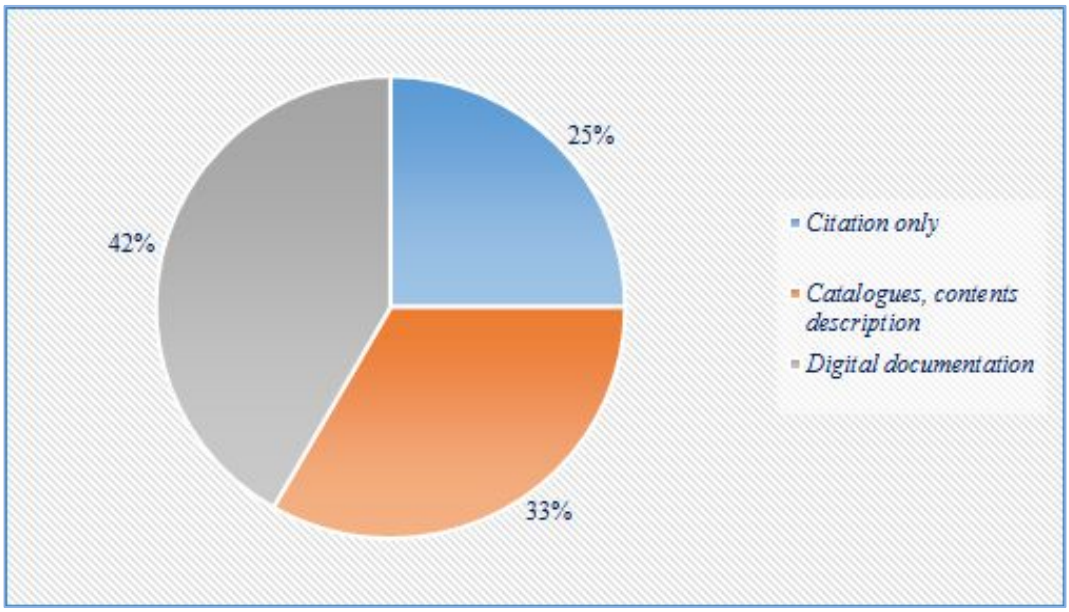
There is a wide documentation about problems related to women on Italian web sites of associations and institutions dedicated to these topics.



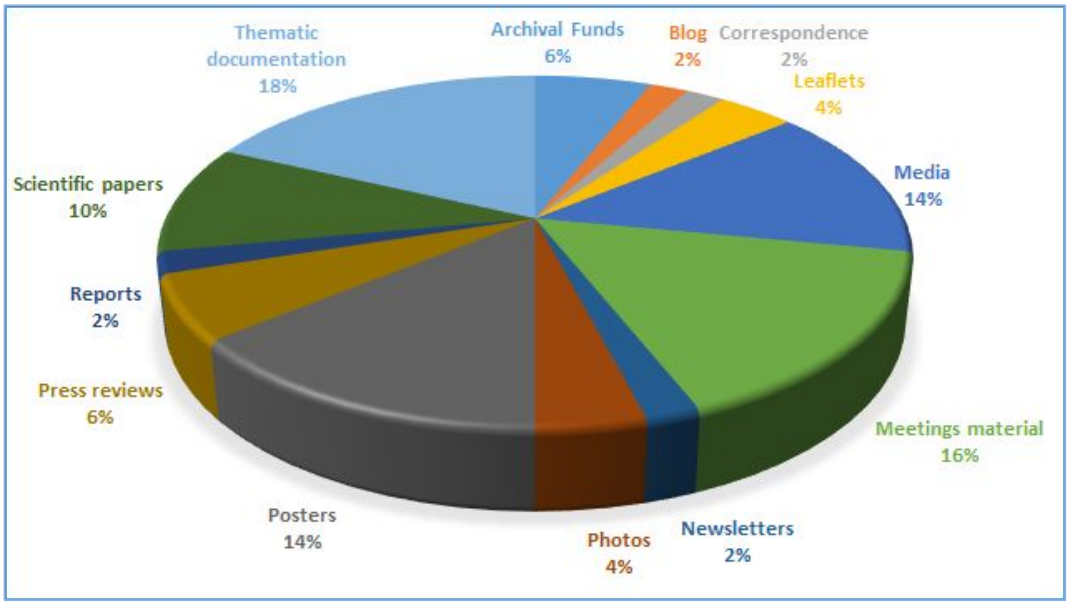
Graph 7 – Women community – Nature of Stakeholders

History of women is disseminated thanks to grey material stored in archives but also by means of websites which publish literature of various ages pertaining with the historical, cultural, political and social memory of Italian women and of their feminist movement for emancipation.

A good example is provided by the "Biblioteca Unione Femminile Nazionale" [12], a library established by the Milanese suffragettes more than a century ago and still offering nowadays a wide catalogue on feminine literature. This community usually pays more attention than the other three to the organization of the grey material in databases/catalogues (33% of documents); in the 42% of the cases, documentation is digitalized and is about the historical, cultural, political and social memory of Italian women: it is worthy mentioning the material on the feminine movements for emancipation of the XIXth century – first half of XXth century.



Graph 8 – Women community – Document expression

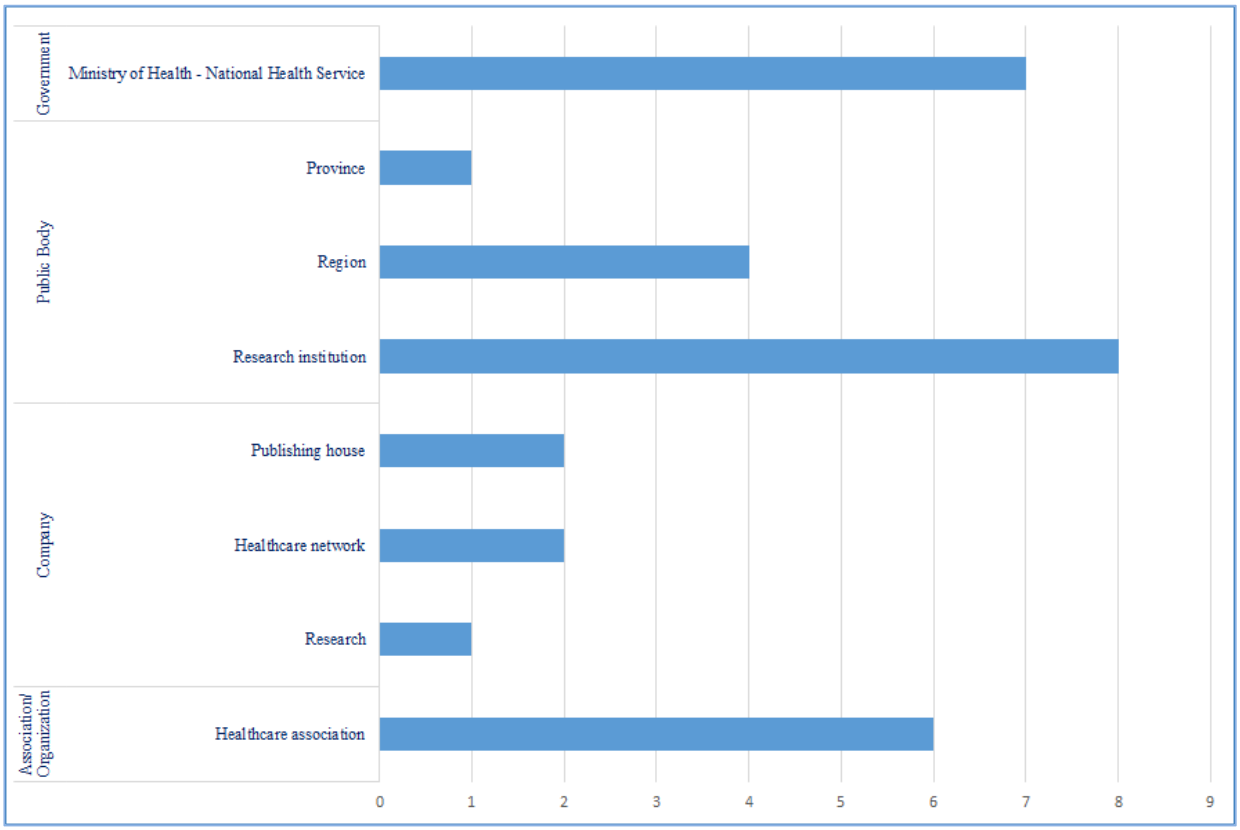


Graph 9 – Women community – Type of contents

As for the type of contents, thematic documentation focusing on women's rights, thoughts, wishes and creativity stands out: on these websites – alongside with worrying statistical data on feminicide and violence against women - there are maps locating the national and foreign Centers dealing indeed with violence against women. Then it comes various material on every kind of meetings, events, manifestations organized on this theme.
A good percentage is represented by the section called "Media" containing collections of movies, audio-visuals and photographs.
Also in this case the material is freely available and the most common format is html (85%) while the use of PDF is around 45%.

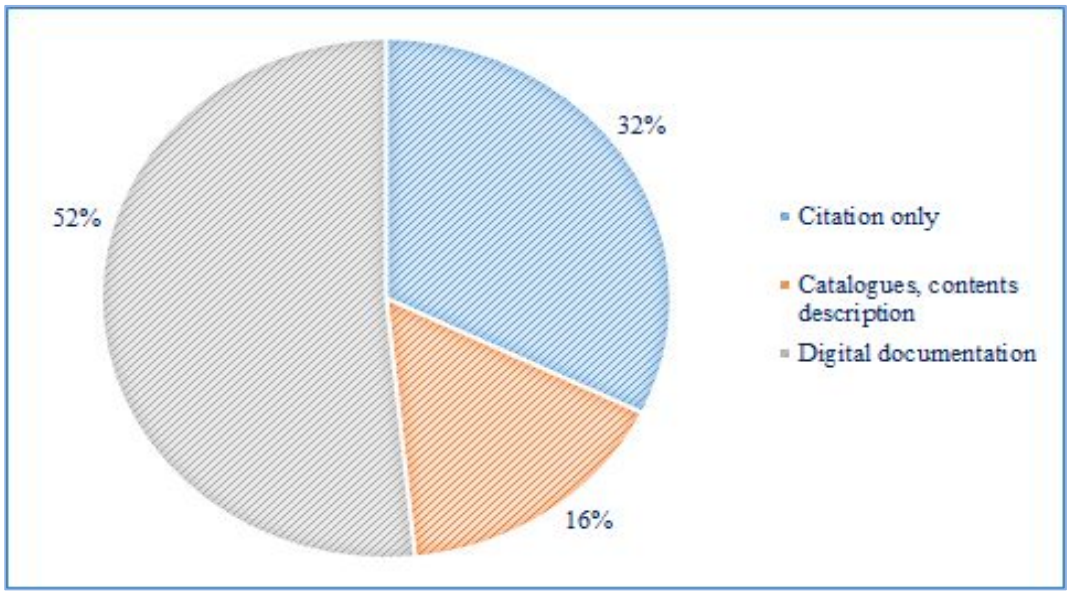
3.3.3 Healthcare

The Italian healthcare community is very rich in terms of initiatives and material available: the Ministry of Healthcare [13] is the central body of the National Health Service and is dealing with the safeguard of public health. This means that the main stakeholders of this community are mainly governmental and research organizations - both at central and outlying level - but there are also several public and private associations or institutions which devote ever more web pages to the fight against alcoholism, drug dependency and to the prevention of severe diseases (e.g. obesity). A significant example is represented by the "Istituto Ortopedico Rizzoli"[x] which is carrying on a project of an open archive of the various material produced within the Emilia-Romagna healthcare agencies, paying special attention to non-conventional literature while recalling the importance of knowledge and dissemination both in terms of diagnosis and therapy [14].



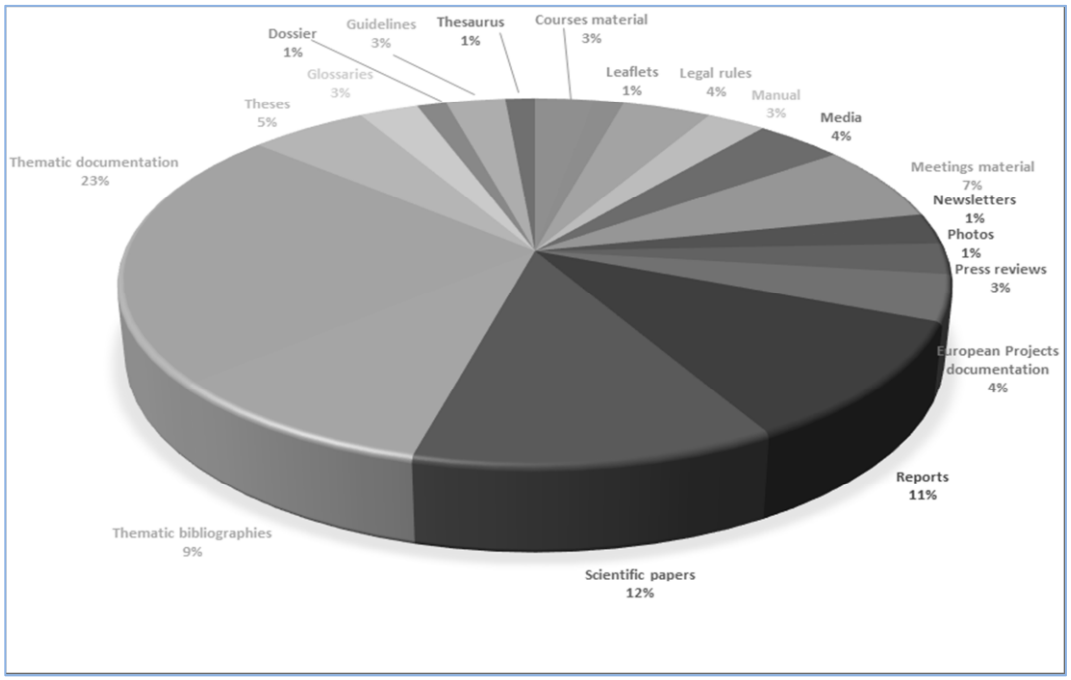
Graph 10 – Healthcare community – Nature of Stakeholders

Other issues, amongst the most studied, concern mental health affections and youth problems, health and safety at work; furthermore, it can be noted the slow appearance of websites where grey material about complementary and alternative medicine (*CAM*)[xi] starts being collected (thought not in a very incisive way yet). Although within this community the percentage of stakeholders using ad-hoc instruments to provide access to the documentation is quite low (about 10%), interesting initiatives are detectable, such as the Archives of grey literature of the "Oncological Center" of Aviano [15] which collect educational material related to training events, protocols and guidelines for welfare activities, references to degree theses, surveys and research works, working manuals and any other document of public interest produced by the Center. Access to consultation of the database is free as well as the downloading of documents, both in html (50%) and in PDF format (50%).



Graph 11 – Healthcare community – Document expression

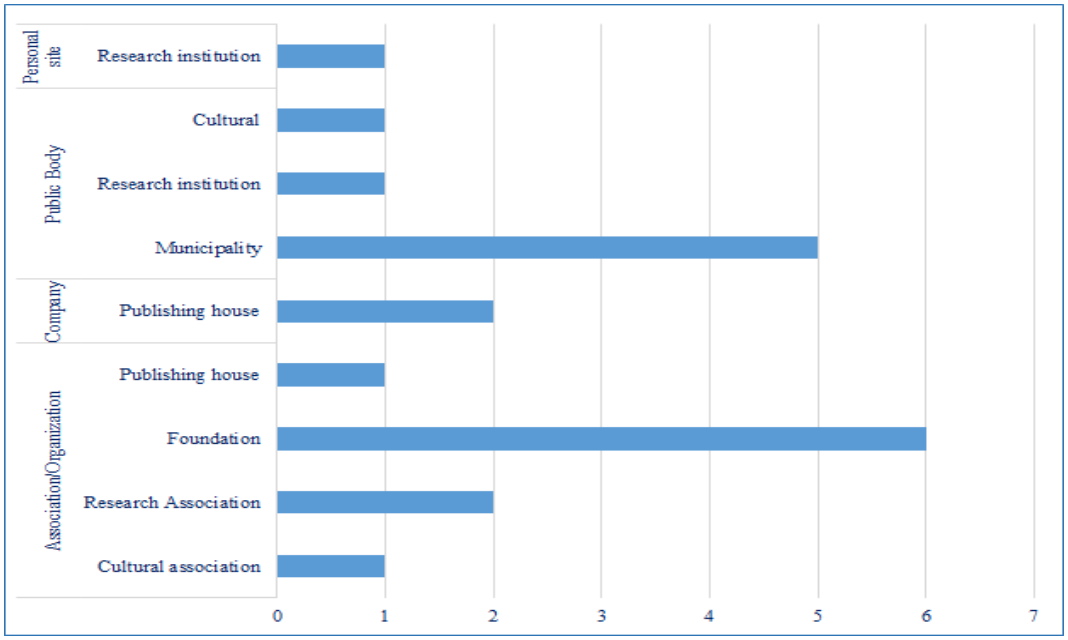
Graph 2 describes the partitioning of documentation within the healthcare community, highlighting how the thematic documentation definitely prevails but also pointing out how much the scientific production (papers, reports, etc.) is relevant in this field; on the other hand, by comparison with the environment and women communities, a smaller percentage is represented by conference, workshop and tutorial material.



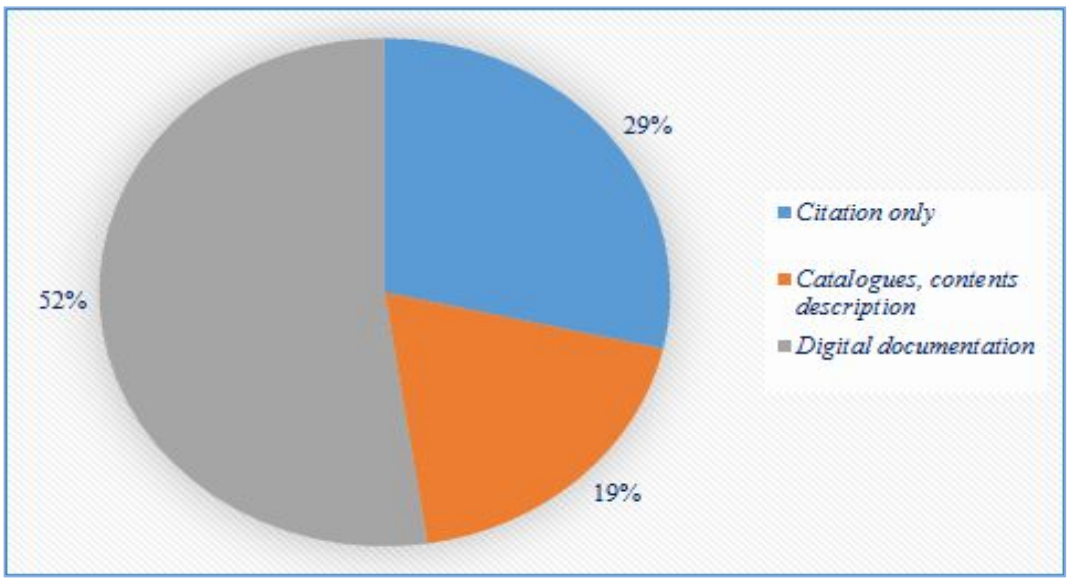
Graph 12 – Health community – Types of documents

3.3.4 History

This community is obviously very rich in terms of documentation: there are hints of local history provided by institutional sites (municipalities and provinces) and usually concerns territorial geography and customs; Foundations are nevertheless the major producers of grey literature of this community and a good percentage is also represented by research associations.

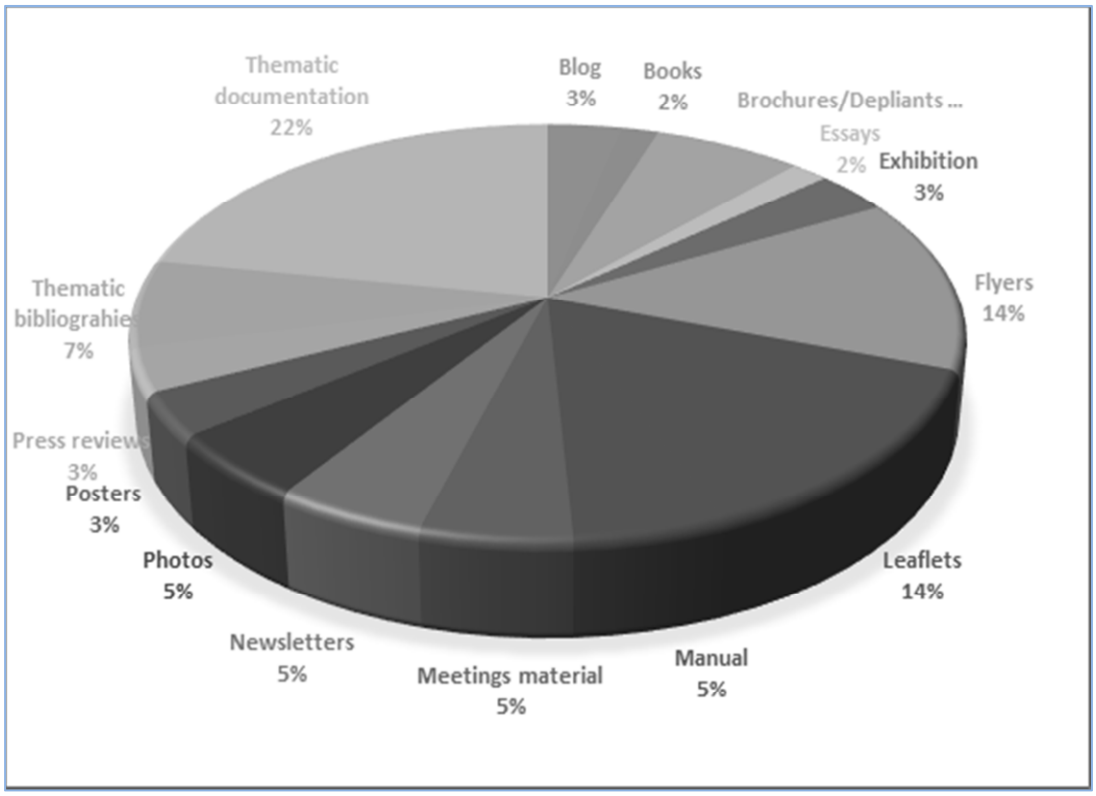


Graph 13 – History community – Nature of Stakeholders



Graph 14 – History community – Document expression

Of course information on national history disseminated by institutes such as the "Istituto di Storia del Risorgimento Italiano" [16] can be easily retrieved: in these cases, grey material reflects refresher courses or more specific activities for both lecturers and students and is mainly made of miscellaneous documents – historiographical, archival, literary, films and direct evidences.
There are also websites devoted to the history of political movements or of single politicians which collected a huge quantity of material such as correspondences, photographs, cuts out, leaflets, etc., which are made available on the Web for avoiding the dispersal of significant pieces of the Italian history.



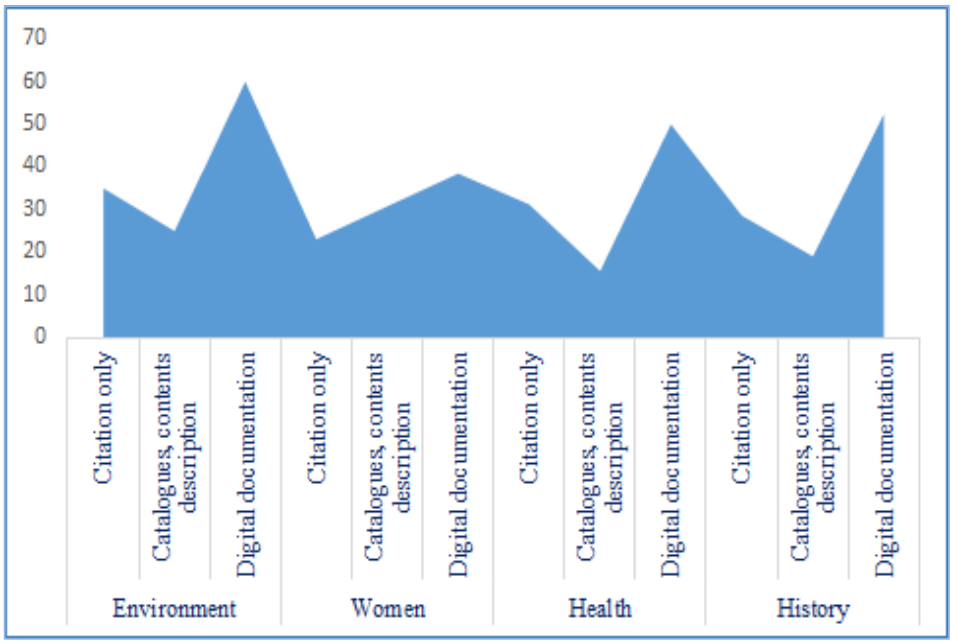
Graph 15 – History community – Type of contents

The material produced by this community largely belongs to the category of thematic documentation, followed by a good amount of documents like leaflets and flyers; there is also the typology 'essays', completely absent in the others. The contents available are offered completely for free, but the percentage of documents downloadable in PDF is lower than in other communities - around 30% - compared to html format (70%).

Within this community, we would like to illustrate an example of "work in progress" such as the "Galileo Galilei Foundation" [17] of our own town, Pisa. The Foundation takes care of disseminating scientific culture by promoting and enriching the heritage of the "Museum of Computing Instruments" which represents a kind of journey through the history of calculation along at least 500 years, from the calipers to the most modern computers. This newly established Foundation has a wealth of very important grey literature consisting of: manuscripts and unpublished correspondence of the major names in the history of science; material related to exhibitions and conferences; educational initiatives and guided tours for schools, operating manuals, etc. Unfortunately this amount of documentation is not accessible because has not been catalogued yet.

Graph 16 shows the progress of the different forms of expression of the documentation proposed by the four Communities. For all of them the highest percentage is the one that aims at the usability of content, compared to the citation only or the bibliographic description and/or cataloguing.

The Environment community offers the greatest amount of content, while the smaller amount belong to the Women community.



Graph 16 – Document expression

Rather equivalent seems to be the number of producers who only cite the presence of grey documents, while the highest percentage in the creation of bibliographic descriptions and/or cataloguing belong to the community of Women.

4. Multi-grey-expressions

From a terminological point of view, the expression *Letteratura Grigia* is the mostly adopted (86%) on the websites of this survey; the other lexical forms used in the query are nevertheless present: the term *materiale grigio* is used in the 12% of the cases only, *documentazione grigia* even less (2% of the cases) while *letteratura non convenzionale* can be found only on two websites.

Quite often the term "letteratura grigia" (and its variants) is followed by the explanations of its meaning as if the producers feel the need to explain what this "niche" term really stands for with words easily understandable to non-experts as well. A few examples excerpted from Italian websites:

- ✓ *…letteratura non convenzionale; quella vasta porzione di preziosi documenti che esula dal circuito commerciale...;*
- ✓ *…detta letteratura grigia scientifica non verificata, mentre quella controllata verrà successivamente pubblicata nelle riviste …;*
- ✓ *…"letteratura grigia" che, per le sua stessa natura, risulta difficilmente reperibile ed accessibile; questa parte del fondo, costituita prevalentemente dal cosiddetto "materiale grigio" (volantini, opuscoli, bollettini, ma anche appunti…;*
- ✓ *…letteratura non convenzionale (materiale grigio)…;*
- ✓ *…nella biblioteca si conserva anche il cosiddetto "materiale grigio", cioè tutte le informazioni su eventi relativi a spettacoli teatrali, cinematografici, convegni…;*
- ✓ *…con l'espressione letteratura grigia si indicano i documenti che non sono stati pubblicati da un editore, ma che possono costituire materiale di riferimento…;*
- ✓ *…materiale grigio (documenti non editi, materiale di corsi di formazione, ricerche della scuola…)...;*
- ✓ *…frutto dell'attività culturale e didattica prodotti nel corso degli anni da studenti e professori (materiale grigio) e materiali in forma multimediale e audiovisiva…;*
- ✓ *…il termine "letteratura grigia" indica la letteratura ed il materiale (opuscoli di organizzazioni, associazioni, ditte, unioni, relazioni su congressi e convegni)…;*
- ✓ *…il rischio sarebbe la scomparsa della cosiddetta 'letteratura grigia', ricerca pubblicata in forma non commerciale, spesso in copia unica, che...;*
- ✓ *…letteratura grigia … quella galassia della produzione scientifica che precede l'offerta dei risultati di una ricerca al peer review…;*
- ✓ *…si va così creando una particolare forma di letteratura grigia digitale, non nuova in termini assoluti, essendo l'erede degli archivi cartacei di «pre-print»..*

5. Conclusions & Remarks

First of all, we have to admit that the processing of data has been time-consuming and little fulfilling for the following reasons: i) there is a big amount of disorganized and badly-structured material; ii) there are many web pages with access denied; iii) the term "letteratura grigia" often appears simply as a description of the material but the content is not always available (e.g. catalogues, lists of references, glossaries, encyclopaedia entries, etc.).

The use of Google as a source for building our grey literature corpus has been useful for evaluating the state-of-the-art and for starting sketching a map of communities and stakeholders outside the academic scenario. However, using Google also presents some disadvantages: i) there is an "explosion" of information and is often difficult to discern who is producing it (a single person, a society, a municipality, a foundation). ii) the pages might contain a large amount of non-pertinent references to grey material (e.g. grey stones, grey furniture and so on); iii) many websites cannot be accessed, change their URLs or are not updated and the information is either cancelled or duplicated; iv) it is impossible to have more than 1000 results.

What is evident is the great variety of non-academic communities, each of them having a different approach to grey material; grey documentation is definitely ubiquitous and belongs to the most diverse fields but at the same time is more fragmented and variously represented than conventional.

The Web surely helps in disclosing what is available as well as in pointing out the difficulties which the communities still face in organizing the grey material: repositories dedicated to grey literature are inadequate and the shared feeling is the lack of willingness or possibility to build platforms for storing great quantity of grey material and equipped with engines for querying and retrieving information as well as semantic search engines for channeling information to the user: documentary fragmentation usually means loss of knowledge for the citizens.

However, we should admit that beyond the Academy, the Italian scenario of grey literature is rather rich and despite the lack of a methodical processing of the documentation, the awareness of the importance of grey literature and its dissemination comes to light.

References

[1] <http://www.reteitalianaculturapopolare.org/it/archivio/fondi-in-rete/fondo-tullio-de-mauro.html>
[2] <http://www3.unisi.it/eventi/prin2007/index.htm>
[3] <http://www.archiviodonne.bz.it/it/progetti/raccolta-di-letteratura-grigia>
[4] <http://www.cittasostenibili.minori.it/bibliografia/grigia.htm>
[5] <http://ambiente.regione.emilia-romagna.it/parchi-natura2000/consultazione/pubblicazioni/letteratura-grigia>
[6] <http://www.openpompei.it/2014/01/31/open-archaeological-data-for-dummies/>
[7] <http://generazionediarcheologi.myblog.it/2013/01/30/gli-open-data-in-archeologia-chi-li-conosce-chi-li-usa/>
[8] <http://www.news-medical.net/news/2007/04/19/17/Italian.aspx>
[9] <http://www.archive.org/movies/movies.php>
[10] <http://www.luccaterre.it/it/dettaglio/7423/Teatro-del-Giglio-Miranda.html>
[11] <http://www.sceglereimitili.eu/raccolta-di-fotografie-video-letteratura-grigia-e-scientifica-pertinenti-alla-filiera-mitilicola>
[12] http://www.unionefemminile.it/biblioteca/
[13] http://www.salute.gov.it/
[14] http://www.news-medical.net/news/2007/04/19/17/Italian.aspx
[15] http://www.cro.sanita.fvg.it/
[16] http://www.risorgimento.it/php/index.php
[17] http://www.fondazionegalileogalilei.it/index_en.htm

[i] Famous linguist and former Italian Ministry of Education.

[ii] The Seminar presents the results of a 2007 PRIN research project funded by the Italian Minister of Research: «Perpetue appendici e codicilli alle leggi italiane». Le circolari ministeriali, il potere regolamentare e la politica del diritto in Italia tra Otto e Novecento.

[iii] www.archiviodonne.bz.it/it/progetti/raccolta-di-letteratura-grigia?

[iv] www.historied.net/portal/index.php?option=com_content...id...?

[v] http://ambiente.regione.emilia-romagna.it/parchi-natura2000/consultazione/pubblicazioni/letteratura-grigia

[vi] http://www.ceisroma.it/upgrade/biblioteca-agora/

[vii] http://artlibraries.net/allg_infos_it/KUNST_GDK.php

[viii] http://www.reteitalianaculturapopolare.org/archivio/fondi-in-rete/fondo-tullio-de-mauro.html

[ix] Please visit the portal of the "Agency for Digital Italy" of the Presidency of the Council of Ministers (http://www.agid.gov.it/dati-pubblici-condivisione/basi-dati-pa)

[x] http://www.ior.it/

[xi] http://www.treccani.it/enciclopedia/dalla-medicina-alternativa-alla-medicina-integrata_%28XXI-Secolo%29/

https://doi.org/10.26069/greynet-2025-000.542-tgj

Data: Is it Grey, Maligned or Malignant?*

Julia M. Gelfand and Daniel C. Tsang (United States)

Abstract:
Cancers, growths, past events, social issues, conditions, and trends are each proverbially described as on a spectrum from maligned to malignant and scientists, physicians, journalists, commentators, politicians and other specialists offer opinions and commentary on what frames the answer to this question of the title. This paper explores not just the color and tone of data, but attempts to resolve what characterizes whether data is maligned or malignant. Hues of greyness distinguish the perils of failing to share, publish nor make accessible research data and the contemporary consequences to scholarship and open access are critical objectives in today's information arcade. Access to data is determined by those who can afford it, discover and know about it, and can thus manipulate it. Grey literature can take the offensive approach to further the role of data, and promote it to advance the common good, contribute to social responsibility and human actions. Data, while increasingly ubiquitous and abundant is the driver of evidence-based foundations, and the link to academic credibility, communication, discourse, dialogue, and the platform for greater open access. Grey data, possessing some of the attributes of grey literature, difficult to identify, acquire and access, when endangered or threatened, not archived or preserved, requiring methods to organize, sort and stratify, forces nontraditional publishing to pursue data publication to enhance perpetual access and new interpretations for its utility in future learning and research applications. We know that there is a somewhat elevated likelihood that open data policies lead to more widespread knowledge and information sharing, greater self-confidence among information providers and scholars alike, but we know less of whether these patterns have any short or long term benefits or disadvantages for individuals or society and of the factors that moderate and mediate these effects. In the meantime, the new reality is that data is central to the work of science, social sciences and basic human conditions of health and wellbeing and data policies mostly proceed from a grey containment to this new reality. The argument that as libraries become active publishers by digitizing content, creating new content, supporting researchers by addressing new domains and formats, that other interpretations of grey data and data more generally are increasingly plausible and that further research on the factors moderating and mediating the effects of data management is needed. This paper explores the continuum for data from maligned to malignant and anticipates data approaching the benign stage emphasizing new hues of grey and open access.

Definitions of Grey Literature & Grey Data

There is much to reconcile – the analogies versus disconnects between Grey Literature and grey data and the relationships between maligned and malignant and how those terms can apply or describe the expanding sphere of grey data today. The grey literature movement with origins now nearly four decades old documents how parallels between original grey sourcing impacts new components of scholarship such as data in the contemporary research setting. From the 1970s when technical and government report literature was the domain of grey literature championed by Charles Auger in his early compilations[1] or a more expansive definition, coined in 1997 that became known as the "Luxembourg definition," which included, "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."[2] An additional clause was added in 2004 noting that grey literature "is not controlled by commercial publishers, or "where publishing is to the primary activity of the producing body."[3]

Fast forward to 2010 when this conference met in Prague and there was a call for some revisions to the definition that reflected more online, electronic or born digital publishing. It was proposed then that four attributes should now qualify the basic definition:

- Documents character of grey literature – more multidisciplinary in content and form
- Includes legal nature of works of the mind – protecting intellectual property
- Reinforces quality level for peer review, validation
- Links to intermediation bridging collection status to readers or users

* First published in the GL16 Conference Proceedings, March 2015.

And the new "Prague Definition" for grey literature was defined: "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, ie, where publishing is not the primary activity of the producing body."

This definition has been challenged as still being too narrow considering new forms and practices of scholarship and research underway and the methods of publishing now widely available. The emphasis on different forms of publishing and the added dissemination streams of content reaching people in new ways; the constant influx of new technologies; incorporation of new media; containing raw research including data, interviews, previews; new forms of single and combined authorship; and the added concerns for preservation, stewardship and perpetual access regardless of future revisions; are all concerns for a current definition of what constitutes being grey or called "grey literature."

Data for the purposes of this paper is defined as research data and we have embraced the standard OMB definition, "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings."[4] Data can be as effusive as grey literature, reflect disciplinary intersections as well as stand on its own and be as varied as spreadsheets, laboratory or field notebooks, codebooks, protein or genetic sequencing, spectra, test plotting, samples, specimens, digital objects, database elements, models, algorithms, formulas, simulations, methodologies, workflows, operating procedures, protocols, and the like. What research data does not include are trade secrets, commercial information, information protected under law, and personnel or medical information that would clearly constitute a clearly unwarranted invasion of privacy. Balancing privacy and data usage will remain one of the biggest challenges in working with data and will contribute to the ongoing "greyness" associated with data.

Data Science as a discipline is also becoming a programmatic area that crosses mathematics, statistics and economics and spans across all disciplines. The definition at our own institution includes how it "may encompass the full spectrum of theories and methods that use data to understand and make predictions about the world around us." This includes "fundamental research on statistical methods, prediction algorithms, data management techniques and policy issue; as well as a broad range of domain-specific data-driven research problems in the sciences, engineering, humanities, education, medicine and business/management."[5] It is a broad range of "theories, algorithms, methodologies and tools that allow us to use data to better understand and make predictions about the world around us."[6]

Some examples of common elements that bridge grey literature with grey data include:
correspondence, project files, grant and ethics applications, technical reports, research reports, signed consent forms.[7]

Maligned or Malignant?

Why, you may inquire do we pose the spectrum of maligned or malignant to define our thesis? Maligned is used in the sense of speaking ill of someone or something, being injurious, slanderous or defaming by causing one to conclude very differently about given circumstances. An example may be when shown to be evil in disposition, nature or intent in order to conclude different outcomes. Malignant, known for its pathological or cellular definition of tending to produce a bad outcome such as death from cancer and can be highly invasive, dangerous, cause harmful influence or effect suggests something out of a researcher's control but will have to be explained due to some other force or context. Thus our imagery shows the extremes of confusion to an example of the recent anthrax episodes that generated fear and uncertainty about the potential exposure and outcomes.

We were influenced by the psychologist, Professor K. Alison Clarke-Stewart who framed some important questions about social issues such as infant day care as "maligned or malignant?"[8] She assessed what kinds of available data perhaps best suggest there could be separation anxiety among infants from their mothers who returned to work during their early infancy and what risks

these infants would have for future emotional insecurity and social maladjustment when placed in daycare. She also asks what the empirical evidence is concerning the effects of infant day care – "it is truly bad for babies or it is undeservedly maligned,"[9] and goes on to examine what evidence there is to support either supposition.

The conclusions of social science and other forms of scientific research reveal that there is much to learn about the effects of something being studied or the outcomes. We can conclude that perhaps there are elevated or declining likelihoods of something happening, we can establish patterns, review longevity data, and understand that different forces contribute to maturation, change and outcomes. We know about situational evidence or trends in work environments, lifestyles that can influence outcomes without observing or studying those contributing factors, such as climatic or environmental occurrences, catastrophic or unexpected events, political interpretations or activities, market surges or declines, and thus the issue today is not whether data is sufficient for the research being undertaken but what else can we learn from the entire research experience and landscape that influences what we observe and learn. That begins with the research question, how the research is conducted and whether sufficient data is collected to offer any conclusions, definitive or not.

Perceptions of data by different communities

Data used to be the stepchild of the research publication process – considered as not worthy of including with a publication, too complex to describe adequately, or too bulky to attach and beyond the needs of most readers. Funding agencies until recently never gave much thought about how to manage research data, even data generated with taxpayer funds, so data were left to rot literally (as computer tapes or disks deteriorate over time), without any thought of preservation and retention of the data underlying scientific research. Like grey literature, research data can be termed as largely fitting the concept of grey data – unfit to see the light of day. They tended to be left on floppy disks or other removable storage devices, often unintelligible to anyone but the original data compiler and over time were unable to be deciphered. Unless the data are collected as part of a well-funded large-scale study, individual datasets are often forgotten after the research is finished and a paper produced or publication issued. Although the scientific ethos mandated reproducibility, research data was more often than not left in filing cabinets or on computer drives inaccessible to others.

That has begun to change. With the advent of concepts like Data Paper[10] and the development of mandates from US federal agencies including the National Institutes of Health (NIH) and the National Science Foundation (NSF) to share and deposit research data for public access, one suspects that data has suddenly been elevated and given a more prestigious status. The concept of data paper is to create a searchable metadata document where instead of focusing on a journal article where hypotheses and conclusions are central, the data becomes the primary emphasis. Increasing examples of this are known. This has been done by the environmental sciences community that studies biodiversity, and these scholars achieve recognition like in other forms of scholarly and creative outputs, by increasing the visibility, usability and credibility of the data resources now cited and published.

When Megan Smith, Chief Technology Officer of the United States responded in an interview when asked how her office is working to make large sets of government data public, "Scientists and universities and the general public can do extraordinary things with it. It could be weather or climate data; it might be data from the Department of the Interior or NASA or water data. Whole industries are being built from things that taxpayers have helped the government know."[11]

Still, efforts to develop new ways of managing research data while encouraging researchers to share their data have been stymied at the individual level. Researchers may be unwilling to share their data, or see little utility in making it available to a public likely to be uninformed about how to interpret the data. Efforts to teach individual researchers about research data management are time consuming and in the main not supported by funding agencies. Funding for research data management is more likely to become more available to institutions than to individual researchers or research projects. Also, the mandates encourage repurposing data may become a more central practice than constantly funding variations of the same research question. Data has

demonstrated that it is more than "nice to have," or being relegated to second class citizenry, for when data is misunderstood it changes the tenor or disposition in fundamental and profound ways.

Risks & Conditions of Data: Unevenness, authoritativeness, misinterpretation

Data is often easily misinterpreted, especially if research results do not contain enough information to place the research in context. How the data was collected, the sample size versus the universe, whether the data was adequately described, and who funded the research are issues any outsider might want to have answered.

Metadata is often inadequate and inconsistent, filenames are subject to private whim, and documentation of steps in the research process are often lacking. Researchers thus can make bold and extravagant statements about research findings – unfortunately not backed up by their data upon closer perusal. A study may not be in fact generalizable but that may not deter the mass media to tack on to the latest research and proclaim a scientific "breakthrough." Traditionally research subjects in some fields were limited to who were readily available, say, college undergraduates of a certain ethnicity and demographic. Whether such findings are in fact generalizable to a larger population, another setting, or another time period is dubious.

Repurposing the research frontier

With the last few decades demonstrating that an increasing amount of research is done collaboratively with different specialists sharing how to examine and study a problem, bringing their specialized skills and objectiveness to the frontlines, and having the desire to share research findings with other likeminded citizenry, several conclusions are evident. They include the need to share research data widely. This can be from the vantage point of international or global sharing with the focus on making sure colleagues and inhabitants in developing countries who are already challenged by not having as dependable or consistently good access to information, get it.[12] It also serves the ongoing development of ICTs, statistical analysis and the intensity of interdisciplinarity across many subject areas and the entire academic landscape.

The directive of working or conducting research in a more open and networked environment, now increasingly commonplace suggests that the cyberinfrastructure that we come to rely upon is more complex than we perhaps anticipated as it was launched. With the unknowns came opportunities and the role of the Internet over the last two decades has promoted connectivity, exchange, collaboration, and reduced barriers and thus redefined new indicators for competitive intelligence and for intellectual competitiveness, a sense of evidence-based models and the ongoing need for best-practices. Supercomputing now available in a handheld device offers new metrics and ease of computation that can be achieved more economically and efficiently. Being committed to the tenets of sustainability, whether it be economic, social or environmental[13] reinforces how the research process ensures integrity, accessibility and stewardship of research data in the digital age.

Data is often delegated to the cloud. Out of sight but not out of mind, data has found a parking spot in the cloud where costs are less and options more plentiful in terms of what to park.

Role of Libraries: Finding & supporting data

Library catalogs, finding aids and most recently the discovery systems that consolidate access to all content via this new breed of discovery tools support data in new and novel ways. By engaging with social networks, media, and other innovative technologies such as QR Codes, discovery systems from a user, contextual and content owner or publisher perspective will impact a greater reliance on data.

The Internet Manifesto 2014[14] widely adopted by the library community and its publishing partners ensures equitable access to the Internet and its services in support of freedom of access to information and freedom of expression and data related content is no exception to this. As the current chair of IFLA's Committee on Freedom of Access to Information and Freedom of Expression, (FAIFE) writes about this update,

"The world has changed significantly since 2002 both physically and digitally, and we now have a greater experience and understanding of the role of the Internet and digital resources in our

services, and in developing connected societies where individuals have the skills that they need to exploit the opportunities that technologies can bring. We also have a greater understanding of the threats that can be posed through the Internet including the impact on human rights of inappropriate monitoring and surveillance, and from criminal activity….reflects this experience and reinforces the vital role of library and information services in ensuring equitable access to the Internet and its services in support of freedom of access to information and freedom of expression."[15]

Libraries, even experiencing these trends remain insufficiently staffed and ill-equipped to handle research data management, but they welcome the opportunity to manage data much like they have taken on different elements of library holdings. Managing research data is not the same as oversight for digitizing printed materials or archiving and preserving digital objects. Data needs to be curated, backed up but also restructured as formats and platforms change. In other words data curation is not just dumping data in a repository, providing access, telling users where it is and considering that sufficient. The level of data curation will vary depending on the resources, skill-sets and time a library or institution can devote to it.

Typically a library or data repository receives a dataset at the end of a research life cycle, often decades after – thus posing often intractable challenges to the data archivist.
Efforts now are being devoted in some institutions to involve researchers at the beginning or during the research life cycle in best practices in research data management. But such efforts are labor intensive, varies by discipline, and often unfunded.

Data Professionalization
With the emphasis of this paper positioned on the role of libraries and librarians handling research data, we should not discount how the commercial world is also responding to the need to create more reliable data management services. The advertising industry is partnering with the computer industry to explore how it "turns Big Data into new ideas." Ogilvy & Mather, among the largest agencies with global revenue, has a new idea about stepping up a longstanding commitment to data-driven decision-making by forming a unit named OgilvyAmp where "Amp" is short for "amplify" and is being directed by a leader with the new job title of global chief data officer. The previous emphasis on information evangelists and such gurus has led to this latest series of tasks that include data strategy and planning, analytics services and data management continuing the way innovative companies such as Ogilvy Mather has amplified the creative process when advertising is developed by paying attention to data with sayings like "Advertising people who ignore research are as dangerous as generals who ignore decodes of enemy signals." The goal of companies such as Ogilvy is to "remove data as a distraction and position it as a tool in the creative palette." Descriptions note, "As data becomes more complex, we want to move data up in the process I support of the creative…You need someone to translate stories out of the data instead of just giving the creative a dashboard." "Creative departments should embrace data as part of their raw material, instead of seeing it as taking their power away." "Data doesn't replace creativity…but great creativity has lasting power that data may not have predicted."[16]

In other innovative business and management applications, data has value if it contributes to efficient and cost-effective savings. Data gains tractions if it has positive outcomes for:
- Reduces building & investment costs
- Allows for better and improved cash flow during projects
- Encourages new opportunity costs – using new apps to interpret results of search or data analysis
- Promotes a more green and sustainable environmental landscape[17]

Data mastery is highly associated with a digital vision. As companies and organizations create their digital vision, they are experiencing a digital transformation, although a compelling vision of a digital future still remains unclear. Technology can remove obstacles and extend capabilities but to enhance users or customers' experiences, streamline operations and transform operational methods and business models is a never ending process.[18]

Back to the library

Today we see positions posted in nearly every research library around the world trying to recruit for professional librarians who can offer appropriate and relevant services to be delivered by a team of research data curation librarians. Sometimes called eScience or eResearch Librarians, Data Librarians or Digital Initiatives Librarians, the work that these professionals now conduct reflects the scholarly communication environment we find very commonplace, where we want to engage users as authors, and develop collections and services that directly support teaching, learning and research missions of the institutions that are trying to re-imagine the contemporary research library.[19] These positions are commanding a lot of attention from library users seeking directions in where to publish, how to publish, how to provide accompanying research data and how to share and repurpose data. Not easy to fill, there remains insufficient competition for all the slots that are being created. Related is the province of statisticians and indicators of professional forecasting does not suggest that there will be an adequate supply of them either for the needs in the marketplace.

In addition, new data intensive environments rely upon the work of semantics, digital humanities, web technologies, human computer interaction, data-mining, information retrieval, communication, strong web programming skills to forecast for tomorrow's new discoveries. As Steve Lohr writes, we still must be mindful that this popular new trade contains far too much handcrafted work or "data wrangling," "data munging" or "data janitorial work." It has been estimated that between "50-80% of data scientists' time" is devoted to "collecting and preparing unruly digital data, before it can be explored for useful nuggets."[20]

Repositories

Many libraries increasingly have institutional repositories and digital commons to contain the scholarship from their affiliates. These repositories are in addition to the thematic or subject repositories that have evolved where one stop deposit and consultation is increasingly the norm to stay informed about the latest contributions to a field. Examples of this include ArXiv, with nearly a million ePrints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics (http://arxiv.org/), the Social Sciences Research Network (http://www.ssrn.com/en/) that supports a score of subject areas. There are also major incentives to not disenfranchise non-recognized or non-distinguished members of the academy but to encourage undergraduates and the student population to also deposit their creative outputs. Examples[21] of journals containing student work, independent projects and group work are increasingly gaining recognition as seen by the detailed work at Illinois Wesleyan University, a small independent liberal arts institution. Such efforts are examined in the increasingly common efforts to promote open data, open education and just reduce the barriers for access. A recent emphasis that addresses these issues is the intersections between scholarly communication and information literacy that librarians have been embracing for the past few years where new library services have been developed to offer supplemental instruction about the value of each of these significant roles.[22]

Just as we continue to explore grey literature, grey data, open data, open science, open education we celebrate ten years as plans are underway for the Tenth International Conference on Open Repositories to be hosted cooperatively by the Coalition of Networked Information (CNI) and several institutions in June 2015. That proposed agenda is a call for many topics to be explored including how repositories can support the needs of research data via data registries, storage, curation lifecycle management, management and digital preservation tools. In addition, the CNI is also tracking the establishment of digital scholarship centers or labs in universities and colleges and hosted a program in April 2014 that offered participants a way to examine the intersections of Scholarly Communication, Visualization and Digital Scholarship.[23] The data element was very visible in this work environment.

How data is structured

At a recent forum on our campus highlighting the new Data Science Initiative[24] we learned how distinctions between statistics and computing have merged and blurred. Related to this may be the parallels of machine learning to the operations of brain functioning. IBM has Watson and it has been shown that learning has similarities between the human and a programmable robot. The same deep learning problems can be applied across the sciences to physics, chemistry,

biology and related disciplines. Depending on computer science, programming, machine learning, statistical analysis, varied technologies, one can integrate applications with understandings so that they can be ready for analysis.

Impacts of Open Access

The Open Access movement has contributed significantly to perceptions, expectations and directions of open data. Publishing opportunities and scholarly communication directives illustrate many examples of how government and the private sector have independently and collaboratively produced new information products, resources and methods of handling data. The pricing of free content does not always come without other associated costs and as soon after the launch of the Educational Resources Information Center (ERIC) by the US Department of Education in 1966 we have seen a very broad range of products and methods of information production.

Highlights include the following launches as noted on Peter Suber's Open Access Timeline:

- Agricola database in 1970 from the National Agriculture Library
- USENET created in 1979
- Text Encoding Initiative (TEI) in 1987
- Psycoloquy among first free online journal that became peer-reviewed in 1989
- First web page was debuted by Tim Berners-Lee in 1990
- Preprint Archives, Mathematical Physics and arXiv launched in 1991
- GenBank launched by the National Center for Biotechnology Information (NCBI) in 1992
- CERN launched its preprint server in 1993
- Human Genome Project, Social Science Research Network (SSRN) and NASA Technical Report Server went live in 1994
- Scientific Electronic Library Online (SciELO) went online in 1997
- Wikipedia was born in 2001
- Budapest Open Access Initiative (BOAI) becomes reality in 2002
- Several scholarly communication initiatives launched in 2002
- Public Library of Science (PLoS) obtains funding to start tow open access journals in 2002
- Google began to digitize and index millions of public domain and copyrighted books from five major libraries in 2004
- From 2004 on, aggressive activity on a global scale took place exploring how to produce open access content and share worldwide – from governments, academic institutions, non-profits, philanthropic foundations, and hosts of individuals with journal declarations being among the most common examples of new information resources[25]

Examples of Data

One flavor of data does not apply across everything. Some will be greyer than others but making generalizations remain problematic. One should be careful and cautious when defining data and its properties. Some examples and current experiences follow:

Survey Data

Common elements include but are not restricted to consumer behavior, trends and, demographics. In the social sciences, survey data is typically structured as a rectangular alphanumeric file, arranged by respondent responses to multiple-choice questions. Areas where such data are relied upon include demographic, public opinion, political participation and consumer behavior research. Given the current importance of social media in social movements etc., researchers are likely to become less interested in structured survey data and will want ways to analyze data that are structured differently, e.g. as tweets. In other words data mining will become a necessary core skill.

Big Data

Getting the basics down for understanding Big Data is easily done by reviewing Cathy O'Neil and Rachel Schutt's book, *Doing Data Science*.[26]

The report, *Big Data: Seizing Opportunities, Preserving Values* issued by the White House this year and currently under review[27] suggests the importance placed on big data due to the

increasingly available sensors, cameras and geospatial technologies that can track global movements. The purpose of this report was to determine how big data will transform the way we live and work and alter the relationships between government, citizens, businesses and consumers. The perception is that big data technologies will be transformative in every sphere of life. Big data according to some experts is "fundamentally reshaping how Americans and people around the world live, work and communicate. It is enabling important discoveries and innovations in public safety, health care, medicine, education, energy use, agriculture and a host of other areas. But big data technologies also raise challenging questions about how best to protect privacy and other values in a world where data collection will be increasingly ubiquitous, multidimensional and permanent."[28] As exciting as knowledge discovery is, and how intensely the Internet has changed how we live our lives, big data as it is defined does not come without challenges and concerns.

The definition of big data is "The capability to manage a huge volume of disparate data, at the right speed and within the right time frame, to allow real-time analysis and reaction." Big data is typically broken down by three characteristics, including volume (how much data), velocity (how fast that data is processed), and variety (the various types of data).[29] Data virtualization and data warehousing is critical for businesses as well as for research practices.

What the US Federal Government has learned so far is that there is concern with data practices.

- Big data tools can alter the balance of power between government and citizen
- Big data tools can reveal intimate personal details
- Big data tools could lead to discriminating outcomes[30]

The policy recommendations made so far include:

- Advance the Consumer Privacy Bill of Rights
- Pass National Data Breach Legislation
- Extend Privacy Protections to non-US Persons
- Ensure Data Collected on Students in School is used for Educational Purposes
- Expand Technical Expertise
- Amend the Electronic Communications Privacy Act[31]

It is easy to agree with Emma Uprichard that we are all consumed by the "challenges of big data." We are led to believe that big data "brings new hope to big social problems and social policy that big data will help us deal with crime and terrorism, intervene with social problems and social policy and may be cheaper to use than organizing large-scale official surveys."[32]

From this we can conclude that technologies are driving the potential for current practices that may accurately or inaccurately represent what we value most. In the spheres of science, technology, and medicine, some ongoing concerns include, data formats, the ambiguity of human language, the need to repeat applications in data projects.

Examples of tools that make data less grey include:

- Spreadsheets – Excel has been around a long while and offers methods to create charts and calculations to allow for models
- ClearStory Data[33] – software that recognizes many data sources, pulls them together and presents the results visually
- Statista[34] – commercial sourcing of data that gives public & private data for eSearchers information in multiple options of delivery offering infographics
- Trifacta[35] – uses machine learning technology to find, present, and suggest types of data that might be useful to see and explore
- Paxata[36] – automates data preparation for analysis
- MapReduce[37] – a software framework that allows developers to write programs that process massive amounts of unstructured data in parallel across a distributed cluster of processors or stand-alone computers
- Data Conversion Laboratory (DCL)[38] – been in the marketplace since 1981; uses optical character recognition (OCR) to integrate data from multiple sources such as scanning, MS documents,

Each of these can be applied to social science data, medical, environmental, science, and interdisciplinary contexts that blend and create new understandings and lend to new knowledge generation. Data and digital conversion, storage, archiving, description and analysis are all the more possible due to "automation and technology that lets us transform things without needing to build things anew every time."[39]

Credibility of Data

"Big data may mean more information, but it also means more false information," and even when the information may not be false the problem is "that the needle comes in an increasingly larger haystack."[40] That is the problem most researchers and scholars face.

Contemporary studies of mindfulness and how it contributes to creativity and innovation have been in both the scholarly and popular media. "There are some who believe the increasing power of Big Data (using powerful computers to sift through and find patterns in massive amounts of information) is going to rival the human consciousness at some point. But there's also growing skepticism about how effective Big Data is at solving problems."[41]

All data are not equal – "experts" may "trust" certain data over others. Source credibility is one big factor – who is compiling the data and is that institution's work reliable? Are its methods transparent? And is there enough documentation to allow a subsequent user to understand how the original data was collected, arranged, described, massaged, and cleaned. This has led to the concept of apply a data seal of approval to data archives – for example, ICPSR, the major social science data repository at the University of Michigan, has obtained a "Data Seal of Approval" for its work.

Researchers are concerned about relationships and conclusions that their findings and data suggests. Spurious correlation, defined by Pearson in 1897 describes the correlation between ratios of absolute measurements that arises as a consequence of using ratios rather than because of any actual correlations between measurements[42] and is one of the main motives for the field of compositional data analysis which deals with the analysis of variable that carry only relative information, such as proportions, percentages and parts-per-million.[43] Spurious data, relationships and correlations may offer respite and humor but the seriousness of when they enter mainstream media or are introduced erroneously is problematic. These examples of relationships of total revenue generated by computer arcades correlating with the number of computer science doctorates awarded provides mixed up interpretations as the computer gaming industry matures. Another example is the number of oil imports correlating with consumption of high fructose corn syrup.[44] Pearson's definition is obviously not to be confused with misconceptions about correlation and causality.

Data Publishing & Publications

New information products issued by a variety of sources globally support individual researchers and functions of work conducted in different sectors. Specific resources developed by the academic community with which we are most familiar include the Data Management Plan created by the California Digital Library (CDL), and other widely respected efforts generated at Purdue, in the UK and at many institutions.[45]

The University of California has also created DASH, a data sharing platform for researchers at the 10 different campuses.[46] These resources are open source and can be used by institutional adoption at other institutions. Each of them contributes in unique ways to data sharing and ultimately to new knowledge generation.

Funding agencies requirements vary regarding the retention periods for data generated from sponsored research, depending on discipline, and form of data. Data longevity thus may vary by different indicators. This implies that data can be "weeded" – a heresy to some data archivists whose archives had missions to preserve and retain data in perpetuity. Before deciding to weed data, attention needs to be given to whether the data can be repurposed to meet current research needs, and whether format migration is feasible. The retention vs de-selection debate remains controversial and problematic, with tools and economic consequences allowing for a

longer lifespan and less expensive storage of data with ease of providing better metadata describing the data.

Looking Forward: The next horizon for grey data

Recent research and scholarship has shown a large wave of interest in data management in this accelerating digital age. We must continue to be vigilant about the "validity of research data, standards that do not keep pace with the high rate of innovation; restrictions on data sharing that reduce the ability of research to verify results and build on previous research; and huge increases in the amount of data being generated, creating severe challenges in preserving that data for long term use."[47]

Reflecting on the Internet Manifesto, the current trajectory and the landscape we think data, especially grey data instills, we can't forget the diverse nature of library work and the opportunities we now have with diverse technologies and new skills individuals must develop and realize that it too, will continue to be part of an evolution of services and resources in libraries. As Steven Bell writes, "But what if, in addition to other core values instilled in LIS programs such as protecting patron privacy and defending intellectual freedom, LIS students learned and practiced methods for tackling tough problems and developing thoughtful solutions – in any situation. Academic libraries need LIS graduates who can assess situations and resources, identify the source and nature of a problem, and then craft an appropriate solution. In other words, educate [the next generation] to identify, frame, think through and solve problems the way designers do."[48] Perhaps we will have strategists, technologists, designers and curators in the role of librarians to manage data.

Conclusions: Maligned or Malignant?

If scholars think that data is maligned it may be because of this "liberal notion" we possess regarding how evidence-based decision-making is superior to one without evidence. The specialty field of logistics and decision sciences points to this. Let's be clear that not all evidence contains data and not all data constitutes evidence concerning the issue at hand. Data may be relevant or it may not be. Hence the "wrong" or misappropriate date that is selected should be worthy of being maligned. Data is often not presented in ways that lets us judge if it is worthwhile or contains relevant data. Hence the dilemma – is it maligned or malignant? How much malaise or data fatigue is there in the work in which we engage? How do we plan accordingly? Strategically, for policy-makers, data is seen through preconceived notions, if it matches one's preconceptions, it is considered "good" data; otherwise it may be maligned.

Marcus Banks writes in 2010 how his interest in grey literature has shifted over the previous five years.[49] He speculates that grey data is compatible with grey literature due to the Web 2.0 directives and new opportunities to disseminate via many new avenues such as blogs, social media, tweeting. The end product or tweets has become "grey data" but we urge our audience and readers to consider the research by-products of both formal science and the social content or new methods to communicate it as worthy of retention and preservation. Libraries are currently trying to influence governments and large enterprises to save, authenticate and preserve older runs or versions of data for future generations to compare, study and align with recent or current datasets, but that continues to be a challenge based on available resources, financial and human.

Malignant data can be construed as being misunderstood, clumsy or insufficiently analyzed – a sense of false positives, incomplete data, or a set of data that cannot be replicated using the same techniques and instruments for whatever reasons. This is increasingly problematic in the data sharing world. Funding sources may not want to consider supporting multiple efforts to study or engage in similar projects, however, it is the challenging outcome that may be providing clarity and a definitive analysis, not the original source. Building on previous work is the hallmark of research and taking it to the next level with more learned resourcefulness, an open and innovative mind makes it ultimately successful. Malignant data may still be vigorous but it is sick, perhaps suffering from malaise and fatigue and needs a new push to find its path.

Just two weeks ago, Pope Francis addressed the European Parliament with a grimly somber diagnosis about the state of Europe, that he described as a "continent's malaise," referring to

how he perceives it to have lost its way, its energies sapped by economic crisis and a remote technocratic bureaucracy. It is increasingly a bystander in a world that has become "less and less Eurocentric," and that frequently looks at the Continent "with aloofness, mistrust and even at times, suspicion."[50] He went on to describe the aging product as "grandmotherly, no longer fertile and vibrant…and we cannot allow the Mediterranean to become a vast cemetery" and concluded another speech to the Council of Europe in Strasbourg on this same trip, by expressing more optimism, "It is my profound hope that the foundations will be laid for a new social and economic cooperation."[51]

We concur with O'Neil and Schutt that we would like to "encourage the "next-gen" data scientists to become problem solvers and question askers, to think deeply about appropriate design and process, and to use data responsibly and make the world better, not worse."[52] And to echo Pope Francis with his optimistic pledge for cooperation and not despair. That also makes data less maligned, malignant and grey.

References

[1] Auger, C.P., ed (1989). *Information Sources in Grey Literature, 2d ed.* London: Bowker-Saur.

[2] "Luxembourg Definition" of Grey Literature (http://opensigle.inist.fr/handle/10068/697932)

[3] Schopfel, J, Farace, D.J. (2010), "Grey Literature." In Bates, M.J., Maack, M.N. eds., *Encyclopedia of Library and Information Sciences, 3rd ed*.. Boca Raton, FL: CRC Press, 2029.

[4] OMB Circular 110 (http://www.whitehouse.gov/omb/circulars_a110#36)

[5] Data Science Initiative, University of California, Irvine (http://datascience.uci.edu)

[6] See http://datascience.uci.edu/about/

[7] Adapted from Defining Research Data (http://library.uoregon.edu/datamanagement/datdefinned.html) by the University of Oregon Libraries.

[8] Clarke-Stewart, K.A. (1989). "Infant Day Care: Maligned or Malignant?" *American Psychologist* 44, #2, 266-273.

[9] *Ibid*, 266.

[10] Data Paper (http://www.gbif.org/publishingdata/datapapers)

[11] Dominus, S. (2014). "You Can Affect Billions of People," *New York Times Magazine*, November 2: 14.

[12] Fienberg, Sharing Research Data (1985)

[13] Committee on Intellectual Property Rights and the Emerging Information Infrastructure, Computer Science and Telecommunication's Board Commission on Physical Sciences, Mathematics and Applications, and the National Research Council *(2000). Digital Dilemma: Intellectual property in the information age.* Washington, DC: National Academies Press. http://uclibs.org/PID/31331

[14] Internet Manifesto 2014. See http://www.ifla.org/publications/node/224

[15] eMail announcement from Martyn Wade, November 25, 2014.

[16] Elliott, S. (2014). "At Ogilvy, New Unit Will Mine Data," *New York Times*, October 21, 2014: B6.

[17] Lanser, J. (2014). "A Whole Lot of Data," *Orange County Register,* October 12. Real Estate Section: 6. 14.

[18] Westerman, G., Bonnet, D., and McAffee, A., (2014). *Leading Digital: Turning Technology into Business Transformation*. Boston, MA: Harvard Business Review Press, 113.

[19] See current job descriptions for Data Management positions such as at the University of Michigan (http://umjobs.org/job_detail/103342/research_data_curation_librarian or at the Pennsylvania State University Libraries (http://www.libraries.psu.edu/psul/jobs/facjobs/sdl.html) or at McGill University (http://www.mcgill.ca/library/files/library/14-al9904-01-coordinator-dat-curation-scholarly-communications.pdf) or at the University of Texas at Arlington (https://uta.service-now.com/jobs/recruiting_job_posting_detail.do?sysparm_document_key=u_recruiting_job_posting,a8f9b7706f6c75009f15cccc5d3ee4c3)

[20] Lohr,S. (2014). "For Big-Data Scientists, "Janitor Work' is Key Hurdle to Insights," *New York Times*, August 18: B4.

[21] See http://digitalcommons.iwu.edu/

[22] Davis-Kahl, S and Helmsley, M., eds. (2013). Common Ground at the Nexus of Information Literacy and Scholarly Communication. Chicago, IL: ACRL. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/commonground_oa.pdf

23 Coalition for Networked Information, Spring 2014 meeting, St. Louis, Mo. (http://www.cni.org/events/membership-meetings/past-meetings/spring-2014/)

[24] Data Science Initiative, University of California, Irvine (2014). (http://datascience.uci.edu/). Forum held October 24.

[25] Suber, P. (2008). Open-Access Timeline (http://legacy.earlham.edu/~peters/fos/timeline.htm

[26] O'Neil, C. and Schutt, R., (2013). Doing Data Science. Sebastopol, CA: O'Reilly Media.

[27] White House (2014) Big Data: Seizing Opportunities, Preserving Value (http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf)

[28] *Ibid.* Fact Sheet, http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf

[29] Hurwitz, J., Nugent, A., Halper, F., Kaufman, M. (2013). *Big Data for Dummies*. Hoboken, NJ: Wiley: 280.

[30] *Ibid*. Fact Sheet.

[31] *Ibid*. Fact Sheet.

[32] Uprichard, E. (2014), "Big Data Doubts," *Chronicle of Higher Education*, October 17: B14-15.

[33] http://www.clearstorydata.com/

[34] http://www.statista.com/

[35] http://www.trifacta.com/platform/

[36] http://www.paxata.com/

[37] http://hortonworks.com/hadoop/mapreduce/

[38] http://www.dclab.com/

[39] Gross, M. (2014). "Data Conversion in the 21st Century," *Information Today*, October: 26.

[40] Taleb, N. (2013). "Beware the Big Errors of 'Big Data.'" *Wired*, February 8. (http://www.wired.com/2013/02/big-data-means-big-errors-people/)

[41] Huffington, A., (2014). *Thrive: The Third Metric to Redefining Success and Creating a Happier Life*. New York: Random House: 140.

[42] Pearson, K (1897). "Mathematical Contributions to the Theory of Evolution – On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs," Proceedings of the Royal Society of London, 60:489-498. DOI: 10.1098/rspl.1896.0076

Proceedings of the Royal Society of London **60**: 489–498. doi: 10.1098/rspl.1896.0076.

[43] Pawlowsky-Glahn, V. and Buccianti, A., eds. (2011). Compositional Data Analysis: Theory and Applications. Hoboken, NJ: Wiley.

[44] Spurious Correlations. http://www.tylervigen.com/

[45] https://www.lib.purdue.edu/researchdata

[46] http://cdluc3.github.io/dash/

[47] Ensuring the Integrity, Accessibility and Stewardship of Research Data in the Digital Age (2009). Washington, DC: National Academies Press.(http://www.nap.edu/penbook.php?record_id=12615)

[48] Bell, S. (2014). "MLD: Masters in Library Design, Not Science: From the Bell Tower." Library Journal, November 20, 2014. http://lj.libraryjournal.com/2014/11/opinion/steven-bell/mld-masters-in-library-design-not-science-from-the-bell-tower/

[49] Banks, M. (2010). "Blog Posts and Tweets: The Next Frontier for Grey Literature." In Farace, D., and Schopfel, J., eds., *Grey Literature in Library and Information Studies*. Berlin: DeGruyter Saur: 222-223

[50] Higgins, A. (2014). "At European Parliament, Pope Bluntly Critiques a Continent's Malaise," *NY Times,* November 26: A4, A10.

[51] *Ibid*.

[52] O'Neil and Schutt, chapter. 16.

Additional Resources:

Bessis, N. and Dobre, C., eds., (2014) Big Data and Internet of Things: A Roadmap for Smart Environments. Studies in Computational Intelligence, v.546. Cham: Springer.

Big Data Now: Current Perspectives from O'Reilly Media (2011). Sebastopol, CA: O'Reilly.

Chen, M., Mao, S., Zhang, Y., Leung, V, (2014). Big Data: Related Technologies, Challenges and Future Prospects. Cham: Springer.

Gauntner Witte, G. (2014). "Content Generation and Social Network Interaction within Academic Library Facebook Pages," Journal of Electronic Resources Librarianship 26 (2): 89-100.

Krier, L. and Strasser, CA (2014). Data Management for Libraries: A LITA Guide. Chicago, IL ALA TechSource.

Manning, P. (2013). Big Data in History. New York: Palgrave Macmillan.

Marcum, D, and George, G. (2010). The Data Deluge: Can Libraries Cope with eScience? Littleton, CO: Libraries Unlimited.

Pryor, G., Jones, S., and Whyte, A., eds. (2014). Delivering Research Data Management Services: Fundamentals of Good Practice. London: Facet.

https://doi.org/10.26069/greynet-2025-000.542-tgj

Open data, grey literature and disciplinary differences – Perspectives from a Dutch data archive*

Marnix van Berchum (Netherlands)

Abstract:
Since 2005 Data Archiving and Networked Services (DANS) promotes sustained access to digital research data. DANS offers several services to support this, including the online archiving system EASY, the Dutch Dataverse Network and the portal NARCIS. In this paper these services will be presented, including the differences we encounter between the disciplines using these services. Within the disciplines served by DANS – mostly belonging to the Social Sciences and Humanities – differences can be discerned regarding the 'openness' of the data. As case study the archeology datasets in EASY will be discussed. Many of them contain reports of archeological excavations done in the Netherlands. Should we consider these as research data or grey literature? And, should we open up the access to these datasets? And if not, why not? These, and other questions, will be addressed, providing a view on how DANS currently deals with open data, grey literature and disciplinary differences.

About DANS

DANS, Data Archiving and Networked Services[1], is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW)[2] and the Netherlands Organisation for Scientific Research (NWO)[3]. DANS is founded in 2005, as a 'merger' of existing institutions dealing with research data in the Netherlands. Taking into account these predecessors – the Steinmetz Archive (founded 1964) and the Historical Data Archive (founded 1989) – DANS builds on fifty years of experience in the storage and management of research data. DANS has as its mission to promote sustained access to digital research data. 'Digital research data' is understood to mean: research information, research data (such as databases, spreadsheets, text, images, audio, video, and multimedia) and digital publications (including preprints, reports). Although DANS is not primarily involved with Grey Literature, its services do hold a substantial amount of grey material, and as this paper will show, DANS supports the ideas of this community regarding the optimal availability of research data.

The Services of DANS

To accomplish its mission of the promotion of sustained access to research data, DANS offers three applications for the management, storage and dissemination of research data, and several non-technical services (e.g. training and consultancy, and support on the certification of digital repositories). Below a short introduction on the data services is given.

EASY: Certified Long Term Preservation of Research Data

EASY (Electronic Archiving SYstem) is the system of DANS for the long-term preservation of research data. At the time of writing it hosts a rough 35 thousand datasets, mainly related to the disciplines Humanities and Social Sciences. EASY is a Data Seal of Approval (DSA) certified repository[4], suitable for the storage of data as required by e.g. the Dutch governmental funder NWO[5]. Individual users are able to manually upload their data, including the appropriate metadata. Furthermore, EASY has several more automated processes of ingesting content into the system, connected to other systems (e.g. local data repositories or Virtual Research Environments, hosted at DANS' client institutions).

EASY offers six deposit procedures, tailored to specific disciplinary requirements. These may include additional metadata, like mandatory metadata fields for the Dutch Archeology community[6] (see also below) or the CMDI metadata format of the CLARIN research infrastructure for language and linguistics data[7].

Dutch Dataverse Network: storing data during the research processes

The Dutch Dataverse Network (DDN) is a collaboration of DANS and nine Dutch institutions of Higher Education[8]. DDN started as a project at Utrecht University, in which the university library wanted to explore the possibilities of hosting and managing research data at the institutional level. Several other university libraries joined this initiative, and in a close cooperation between

* First published in the GL16 Conference Proceedings, March 2015.

the two institutions, Dutch Dataverse Network was migrated from Utrecht University to DANS in Spring 2014.
In recent years the Research Data Netherlands (RDNL) consortium[9] introduced the so-called *Front Office | Back Office model*, in which the different functions and roles in the data management process are divided. The RDNL partners supply the Back Office functionalities, like long-term preservation of data and training and consultancy for the Front Offices. The latter, e.g. the university libraries, support the researchers and host data application(s) for the duration of the research and mid term storage. DDN is well suited to take this role, offering the Front Offices a tool with which they can control their institutional research data management requirements.

NARCIS: the gateway to scholarly information in the Netherlands
While EASY and DDN deal with research *data*, the NARCIS portal has a broader scope. NARCIS provides access to different types of scholarly output, including (open access) publications from the repositories of all the Dutch universities, KNAW, NWO and a number of research institutes, and descriptions of research projects, researchers and research institutes. All datasets from EASY can also be found via NARCIS, as well as datasets from other Dutch data archives (like 3TU.datacentrum and The Language Archive). NARCIS harvest the metadata of the abovementioned resources from the institutions, and is at the same time the hub for harvesting the aggregated metadata by service like Google Scholar and OpenAIRE[10].

Open Access at DANS
DANS – as well as its host institution KNAW and NWO – supports the Open Access principles, and stimulates researchers to make their scholarly output available as wide as possible. Acknowledging that data cannot always be open, e.g. in the case of sensitive data containing personal information, the motto of DANS is: "*Open if possible, restricted if necessary!"*
All metadata of the data sets in EASY are Open Access available to all users, and are also harvestable via an OAI-PMH endpoint. In 2014 DANS changed the access categories in EASY, for the first time offering a full Open Access license to the data files. The current categories, with which a dataset can be deposited, are[11]:

1. Open access
2. Open access for registered users
3. Restricted access
4. Other access

The first option is a *CC0 No Rights Reserved* licence[12], making it possible for human and machine users to freely use and re-use the data. Although the CC0 licence does not require an attribution to the creator of the data when re-using, like the related CC BY licences do, this should not keep the scholarly user from properly citing the data she uses[13]. To this end DANS supports data citation with the DataCite Digital Object Identifier (DOI)[14].
Since the implementation of the new CC0 option, just a few datasets have this access category. DANS is actively promoting depositors to deposit with this new category, or reassign existing datasets to this category.

Grey at DANS
DANS is in several ways connected to grey literature. First of all, DANS hosts the datasets related to the Grey Literature Conference Proceedings[15]. All authors of papers at the Greynet Literature Conference are allowed, free of cost, to deposit their related data or supplementary material in EASY. Furthermore, the data documentation files and publications deposited together with the data files can be considered to be grey literature as well.
NARCIS, as national aggregator of the Dutch higher education institutions, contains a substantial amount of metadata referring to grey literature. An estimated 33% of the contents of NARCIS can be classified as grey, existing of reports, unpublished doctoral theses, conference papers etc. present in the institutional repositories.

The archeological datasets
A special type of dataset in the EASY archive has a shade of grey as well: the datasets deposited by the archeological community. All archeologists in the Netherlands are legally obliged to digitally deposit the results of their excavations results in EASY. Special ingest procedures are devised, in which an archeologist can deposit her material at DANS and, at the same time,

register the excavation at the governmental institution involved[16]. Due to this obligation, archeology is a well-represented discipline in EASY, occupying a rough 80% of all data.
A typical archeological dataset consist of the excavation report, pictures taken and geographical data on the location. Disciplinary additional metadata fields are needed and include e.g. references to the excavation number[17]. To preserve the historical buildings and archaeological sites, and perform the archeological research as effectively as possible, most archeological datasets are available under the 'Restricted Access – Archeology Group'. This means only archeologists registered at DANS are able to access the files present in the dataset. As mentioned earlier, all metadata is freely available to all, making these grey resources available to a global audience.

References

DANS Licence Agreement, 2014, http://dans.knaw.nl/en/about/organisation-and-policy/legal-information/DANSlicenceagreementUK.pdf

Dillo, I., and Leeuw, L. de, *Data Seal of Approval: Certification for sustainable and trusted data repositories*, 2014 Data Archiving and Networked Services (DANS), http://www.datasealofapproval.org/media/filer_public/2014/10/03/20141003_dsa_overview_defweb.pdf.

A federated data infrastructure for the Netherlands: the front-office - back-office model, 2014 Research Data Netherlands (RDNL), http://www.researchdata.nl/fileadmin/content/RDNL_algemeen/Documenten/RDNL_FOBOmodel-UK-web.pdf.

Hollander, H., 'The e-Depot for Dutch Archaeology: Archiving and publication of archaeological data', in *Proceedings of the 18th International Conference on Cultural Heritage and New Technologies 2013*, 2014 Museen der Stadt Wien – Stadtarchäologie, http://www.chnt.at/wp-content/uploads/Hollander_2014.pdf.

Riding the wave: How Europe can gain from the rising tide of scientific data, 2010 High Level Expert Group on Scientific Data, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf.

Sustained access to digital research data: Summary of the DANS Strategy Policy 2011-2015, 2011 Data Archiving and Networked Services (DANS), http://www.dans.knaw.nl/en/about/organisation-and-policy/information-material/DANS-strategy-policy-20112015-summaryUK.pdf.

Wiersma, P., *Podium Plus: Onderzoek naar de mogelijkheden tot landelijke samenwerking bij de opslag van onderzoeksdata [Podium Plus: Exploration into the possibilities of a national, cooperative storage of research data],* 2011 SURF, https://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2011/Podium+Plus+-+definitief.pdf.

Notes

[1] www.dans.knaw.nl.
[2] www.knaw.nl.
[3] www.nwo.nl.
[4] For more information on the Data Seal of Approval see www.datasealofapproval.org; the assessment document of the DANS-EASY seal can be found at https://assessment.datasealofapproval.org/assessment_101/seal/pdf/.
[5] See http://www.dans.knaw.nl/en/deposit/information-about-depositing-data/data-contract/data-contract?set_language=en and http://www.nwo.nl/en/news-and-events/dossiers/datamanagement.
[6] https://easy.dans.knaw.nl/doc/DANSarchaeologicalmetadataUK.pdf.
[7] http://www.clarin.eu/content/component-metadata.
[8] At the time of writing these are the three technical universities of the 3TU consortium (Eindhoven University of Technology, University of Twente and TU Delft), Erasmus University Rotterdam, Maastricht University, NIOO – Netherlands Institute of Ecology, Tilburg University, University of Groningen, and Utrecht University.
[9] www.researchdata.nl. RDNL is a collaboration of DANS, 3TU.datacentrum and SURFsara.
[10] See http://www.narcis.nl/background/Language/en for an overview of the content of NARCIS. For a full list of organizations or services that harvest NARCIS, see http://www.narcis.nl/dataprovider/Language/en.
[11] See the Appendix, for a full description of the categories available.
[12] https://creativecommons.org/about/cc0.
[13] Cf. *The Netherlands Code of Conduct for Academic Practice*, as published by the Association of universities in the Netherlands (VSNU), at www.vsnu.nl/files/documenten/Domeinen/Onderzoek/The%20Netherlands%20Code%20of%20Conduct%20for%20Academic%20Practice%202004%20%28version%202014%29.pdf.
[14] https://www.datacite.org. The DOI identifier will be implemented in DANS-EASY in the first half of 2015.
[15] http://www.greynet.org/dansdataarchive.html.
[16] This collaboration of DANS with the Cultural Heritage Agency (RCE) is formally known as the e-Depot for Dutch Archaeology (EDNA), www.edna.nl.
[17] See https://easy.dans.knaw.nl/doc/DEP_ArchaeologyUK.pdf and https://easy.dans.knaw.nl/doc/DANSarchaeologicalmetadataUK.pdf.

Appendix – description of the Access rights available in EASY

[The text below is taken from the help text in EASY (visited 22 January 2015)]

1. Open access
The files are accessible to all users of EASY.

2. Open access for registered users
The files are accessible to all registered users of EASY.

3. Restricted access
The access to the files deposited is limited to a smaller group of users. EASY currently has the two following forms of limited access:

- Restricted - archaeology group.
Please note that this option can only be selected for archaeological deposits.
The datasets are only accessible to registered users of EASY who belong to a specific group. In principle, this option is used as default for all archaeological datasets. Just like Archis is an information system for and by archaeologists, the archaeological files in EASY are also only accessible to other, registered archaeologists. We share the available information on archaeological sites, without this information being immediately available on the Internet so that we can preserve the historical buildings and/or archaeological sites or perform scientific research as effectively as possible.

- Restricted - request permission.
With this option, you - as the person depositing the dataset - can personally manage the accessibility to the files deposited by you. The datasets are accessible only if another registered user, not necessarily an archaeologist, has submitted a ‘permission request'. You, the person depositing the dataset, will be informed of this request and you will personally be able to grant or reject this request for permission to access the dataset.

4. Other access
It is not possible to access the datasets through EASY. DANS is only intended for sustainable, digital archiving behind the scenes. Another possibility is that the data have been stored elsewhere and that only the metadata have been stored in EASY. You can only choose this option if the digital data are accessible, for example, through a different data repository, digital archive, databank or website.

https://doi.org/10.26069/greynet-2025-000.542-tgj

Pisa Declaration on Policy Development for Grey Literature Resources

Introduction

A wealth of knowledge and information is produced by organizations, governments and industry, covering a wide range of subject areas and professional fields, not controlled by commercial publishing. These publications, data and other materials known as grey literature, are an essential resource in scholarly communication, research, and policy making for business, industry, professional practice, and civil society. Grey literature is recognized as a key source of evidence, argument, innovation, and understanding in many disciplines including science, engineering, health, social sciences, education, the arts and humanities.

Grey literature document types in print or electronic formats include among others: research and technical reports, briefings and reviews, evaluations, working papers, conference papers, theses, and multimedia content, representing an important and valuable part of research and information.
In order to realize the benefits of research and information for scholarship, government, civil society, education and the economy, We, the signatories to this declaration, call for increased recognition of grey literature's role and value by governments, academics and all stakeholders, particularly its importance for open access to research, open science, innovation, evidence-based policy, and knowledge transfer.

May 16, 2014

To achieve the full benefits of grey literature for local, national and global communities we call for and encourage the following:

Organizational

1. Greater commitment to open access by governments and organizations.
2. Greater cooperation and coordination among organizations engaged in the production, use, collection and management of grey literature.
3. The use of persistent identifiers and open metadata standards for grey literature.

Research/Educational

4. New forms of recognition and reward for quality grey literature materials by governments, universities and other institutions.
5. Improved standards in the production and bibliographic control of grey literature.
6. Development and implementation of interoperable standards in the management of grey literature.
7. Development of good practice guides for the production, dissemination, and evaluation of grey literature.

Legal

8. Changes to legal deposit and copyright law that enhance the capacities of libraries, collecting services and educational institutions and programs to collect and provide access to grey literature, particularly non-commercial public interest materials.
9. Addressing legal obstacles to the dissemination of grey literature.
10. Further strides in licensing grey content for both commercial and non-commercial purposes.

Financial/Sustainable

11. Identifying available funding for research involving grey literature.
12. Increased support for collection development and long term preservation of grey literature.
13. Increased investment in infrastructure and new technologies for accessing and using print and digital grey literature.

Technical

14. Strategies to tackle link rot and enhance the stability and accessibility of online content.
15. Systems for linking data and other non-textual content to their grey literature publications together with interoperability standards for sharing grey literature.

To endorse the Pisa Declaration:

http://greyguide.isti.cnr.it

Contact
info@greynet.org

GreyNet
Grey Literature Network Service

https://doi.org/10.26069/greynet-2025-000.542-tgj

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

https://doi.org/10.26069/greynet-2025-000.542-tgj

The Intersect of Cultural Studies with other focus areas in the HDIAC Database*

June Crowe and Crystal Sherline (United States)

Our bibliometric research will examine a select set of documents in the Homeland Defense and Security Information Analysis Center's (HDIAC) collection. The focus will be on documents in the "Cultural Studies" focus area. There are over 210,000 documents in the HDIAC collection, and much of it is grey literature. Staff use a template that includes bibliographic information, keywords, task areas, and other descriptive information to catalog items for inclusion in HDIAC's database. In staff discussions, it was discovered that some focus areas intersect with other focus areas. For example, Cultural Studies overlaps with the Medical, Alternative Energy, Critical Infrastructure Protection, and Homeland Defense & Security focus areas. The purpose of this research is to examine the intersects that Cultural Studies has with the other HDIAC focus areas by examining the keywords and task area terms that include the term "Cultural Studies." This process will assist in clarifying the subject key words used to identify cultural studies and facilitate the tagging, acquisition, and discovery processes. It will also give staff a model they can use to gain a deeper understanding of how culture intersects with other disciplines. A bibliometric study will be conducted to evaluate the HDIAC database, Scientific and Technical Analysis and Research Tool (START), and quantify the Activity Index (AI) of the HDIAC database.

$$AI_1 = \frac{Total\ Number\ of\ Publications\ in\ START}{Total\ Cultural\ Studies\ Publications\ in\ START}$$

and

$$AI_2 = \frac{Total\ Number\ of\ Publications\ in\ START}{Total\ Cultural\ Studies\ Cross\ \ Publications\ in\ START}$$

This bibliometric research will allow HDIAC staff to quantify the multidisciplinary nature of Cultural Studies and provide a vantage point for a qualitative study that examines the intersect with other focus areas in HDIAC.

Keywords: Scientific and technical data, Gap Analysis, Grey Literature, Data Discovery, Taxonomy, Bibliometrics

Introduction

The Homeland Defense and Security Information Analysis Center (HDIAC) is located in Oak Ridge, Tennessee and is a knowledge center of excellence for the Department of Defense (DoD) Homeland Defense and Security (HLDS) scientific and technical information (STI). Previously known as the Chemical, Biological, Radiological and Nuclear Information Analysis Center (CBRNIAC), the HDIAC has eight focus areas: Homeland Defense & Security, Critical Infrastructure Protection (CIP), Weapons of Mass Destruction (WMD), CBRN Defense, Biometrics, Medical, Cultural Studies, and Alternative Energy. The last four topic areas are new to HDIAC. Information resources in these areas are collected, analyzed, synthesized, and disseminated to the DoD community. Additionally, the HDIAC produces technical reports, journal articles, and web content, as required in these areas. HDIAC's goals are to: use the best knowledge and technical expertise from the government, industry, and academia to solve difficult scientific and technical problems in HLDS; serve as a ready tool for strategic, operation, and tactical organizations in DoD and the broader homeland defense community to resolve issues; build a community of homeland defense SMEs and provide long-term STI corporate memory for DoD; and build networks to reduce duplicate information holdings in the research and engineering communities (STI –reuse).

* First published in the GL16 Conference Proceedings, March 2015.

Research Method

To examine the HDIAC collection and aid in the development of a richer literature taxonomy, the researchers examined Task Keywords, i.e., the category of the HDIAC scope areas (Alternative Energy, Cultural Studies, WMD, etc.). This assessment allows for quantification of the juxtaposition of the Cultural Studies literature with literature in the other scope areas in the collection, giving an Activity Index (AI).

$$AI_1 = \frac{Total\ Number\ of\ Publications\ in\ START}{Total\ Cultural\ Studies\ Publications\ in\ START}$$

In order to review the literature, an evaluation of the HDIAC START database was performed, and metrics were examined. First, the START database metrics were retrieved by searching Task Keywords for the eight HDIAC scope areas. Once metrics were retrieved for the individual scope areas, metrics were collected for areas where Cultural Studies and other scope areas specifically intersected. In order to obtain these metrics, a search in START was formulated by searching Task Keywords "Cultural Studies" and "Alternative Energy." This was also done for all the other six scope areas, giving an AI_2.

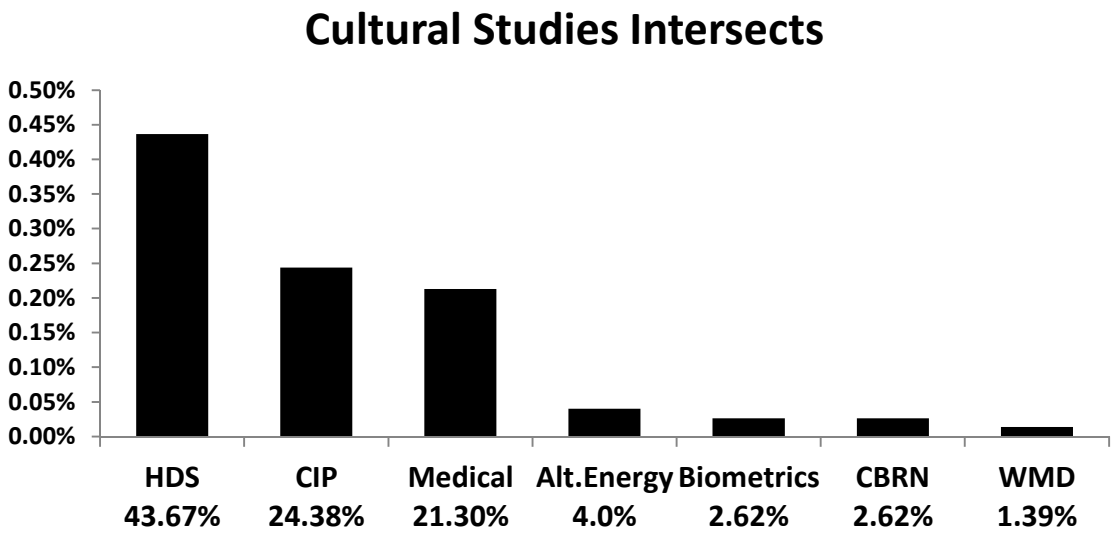
$$AI_2 = \frac{Total\ Number\ of\ Publications\ in\ START}{Total\ Cultural\ Studies\ Cross\ \ Publications\ in\ START}$$

Bibliometric Research Analysis

The database that HDIAC inherited from the previous contractor contained very few records, if any, on the four new focus areas (Cultural Studies, Alternative Energy, Biometrics, and Medical). The original database focused on chemical, biological, radiological, and nuclear subject areas. In essence, there were almost no documents on the new focus areas. HDIAC staff developed a new taxonomy from the performance work statement that provided an overarching framework for collecting information on the various subject domains at a first and second level. The current need is for a more in-depth taxonomy for each focus area. Additionally, our research will assist with gap analysis in collection development in the eight focus areas. Currently, the collection consists of grey literature, journal articles, technical reports, maps, books, and data on various topics. Since November 11, 2013, HDIAC has added 8,880 records related to the focus areas.

Cultural Studies documents comprise 14.82% of the additions or 1,316 items. Cultural Studies is a cross disciplinary domain that intersects with Alternative Energy, Critical Infrastructure Protection (CIP), Homeland Defense and Security (HLDS), Biometrics, Medical, and Weapons of Mass Destruction (WMD). For example, "megacities" is a category under Cultural Studies that also intersects with CIP, HLDS, Alternative Energy, Medical, and Biometrics. Depending on the document, this topic could also easily intersect with CBRN and WMD. Figure 1 shows how Cultural Studies intersects with the other domains. As can be seen in Figure 1, HLDS, CIP, and Medical are the top three focus areas that have intersects with Cultural Studies. The total percentage of Cultural Studies documents with cross over task areas at this point in time is 49.24%, almost half of the content that has been cataloged.

Figure 1: Cultural Studies Intersects with other Focus Areas



In addition to task area terms, the authors plan to examine the subject keywords that describe the eight focus areas in an attempt to standardize and develop a more detailed subject thesaurus for the HDIAC collection. The HDIAC START system's template for cataloging includes bibliographic information, a distribution statement, copyright information, abstract, document type, corporate authorship, task keywords, and subject keywords. The task keywords are in effect the subject domain terms or HDIAC scope area. The subject keywords that have been used have been taken directly out of the statement of work pertaining to this contract. Subject key wording has become more important in search efforts, as the collection has grown in both breadth and depth. The subject keywords provided by the contract no longer encapsulate the documents being collected. Therefore, is has become imperative that a more accurate, homogenous method to categorize the information is created.

Our research will further develop the existing taxonomy for each of the domains to facilitate catalog tagging and searching. For example, under the heading "Behavior," there is a subcategory of "human adaptation and response to perturbations," which will have further subcategories that may include climate change, adaptation to food stress, or civil unrest.

Multidisciplinary Example

Figure 2 is a document cataloged in HDIAC's START database. This document was chosen as an example because it does not fit into just one of the HDIAC's scope areas. The document is cataloged as having the following task keywords: "Cultural Studies," "Alternative Energy," and "Critical Infrastructure Protection."

Figure 2. Multidisciplinary Example Document

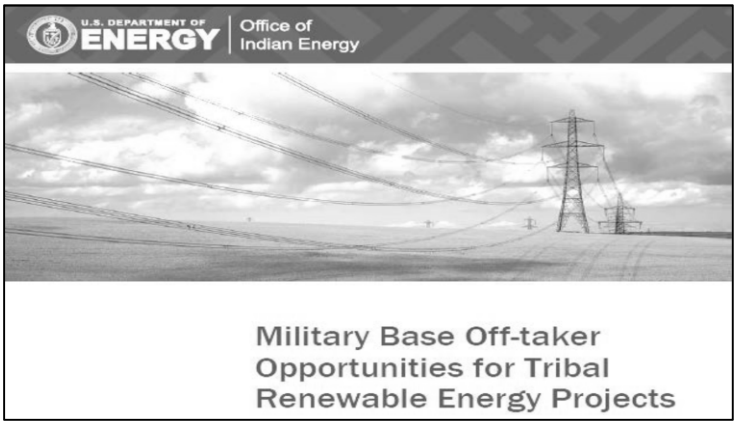


Figure 3 is the document's cataloging output in the START database. The document is a multidisciplinary document and has been cataloged with three different task key words (HDIAC scope areas), yet the subject keywords are vague and only contain secondary subject headings related to some of the task keywords. Ideally, the subject headings for this document would include: diplomacy, communication, land ownership, socio-economics, and environmental resources.

Figure 3. HDIAC Catalog Record

IAC Number: HD-224134
AD Number:
Date Received: 2014-03-31
Title: Military Base Off-taker Opportunities for Tribal Renewable Energy Projects.
Site Holding: HD
SiteHolding Number:
Media Type: PDF
CB Collection: UA
Author(s): Nangle, J.
Corp Source Code: 424651
Corp Author Name: NATIONAL RENEWABLE ENERGY LAB GOLDEN CO
Descriptive Note: White Paper
Publish Date: 2013-05-01
Page Count: 19
Country: US
Contract Number:
Project Number:
Report Number: DOE/IE-0018
Distribution Code: PUBLIC
Distribution Statement: Approved for Public Release; Distribution Unlimited.
Abstract: Tribal lands represent approximately 5% of the national renewable energy technical potential (Doris, Lopez, and Beckley 2013), and their close proximity to many DOD installations makes the Tribes potential partners in helping the DOD meet its renewable energy goals. The benefits to the Tribes include additional income streams and workforce development in green jobs. This white paper surveys DOD installations that could have an interest in the purchase of energy from renewable energy projects on tribal lands. Tribes, in the right context, could supply local renewable energy to meet military demands, which would provide additional sources of energy to the installations, helping them meet federal requirements for renewable energy purchases and greenhouse gas reduction. Tribal members could benefit from additional income from land leases and energy sales, as well as workforce development and maintaining the renewable energy projects. Identification of likely purchasers of renewable energy is a first step in the energy project development process, and this paper aims to identify likely DOD electricity customers that tribal commercial-scale projects could serve.
Supplemental Notes:
Subject Keywords: BIOMASS; DEPARTMENT OF ENERGY (DOE); GEOTHERMAL; HYDROPOWER; INSTALLATION ENERGY POTENTIAL; PHOTOVOLTAIC (PV); PV-RURAL; PV-URBAN; TRIBAL LAND
Task Keywords: ALTERNATIVE ENERGY CRITICAL INFRASTRUCTURE PROTECTION CULTURAL STUDIES OTHER GOVERNMENT AGENCY TEMS-WS 2014-07-29

Taxonomy Development

Since Cultural Studies is one of the four new focus areas, our basic overarching taxonomy was devised using the performance work statement. It was quickly recognized that a more detailed and standardized taxonomy was needed for both cataloging and search purposes. Some preliminary research into the use of other existing taxonomies was conducted. "For example, the Defense Technical Information Center (DTIC) has a technology taxonomy while the Standard Subject Identification Code (SSIC) is the standard for all DoD information, including memorandums and records management. Similarly, the Library of Congress Classification (LOCC) is a commonly used general purpose system. However, taxonomies inevitably have a central theme that guides how the tree structure is arranged."[1]

Our first problem was to determine what subject key terms had already been used in conjunction with the Cultural Studies task tag. An SQL query was performed to find those terms and rank them by frequency of use. Since Cultural Studies is multidisciplinary, this query can provide crossover terms that may have been used from any other task areas. Secondly, DTIC's taxonomy,[2] which is available online at http://dtic.mil/dtic/stresources/techreports/dticSearchTools/thesaurus_desc.html, needed to be evaluated. DTIC's taxonomy did not provide the level of detail needed for the scope of Cultural Studies. The Library of Congress' "Subject Heading List"[3] for Culture, Anthropology, Social Science, and Human Geography, and "Military Culture Factors" was also consulted. We are in the process of evaluating all of the relevant cultural terms from these two sources. Other taxonomies will also be researched. After relevant terms from all these sources are selected, we will look into possible programs that could integrate those terms with the broad overarching terms. Figure 4 is an

example of what the tree may look like. The program will be run periodically to update the taxonomy. After the Cultural Studies taxonomy is completed, the same gap analysis and integration of new terms will be conducted for the other focus areas.

Figure 4. Example Taxonomy Tree

Dominant Values & Beliefs (Broad Term)
- Social Values & Beliefs (second level of Broad Term)
- Religion (broad term from Military taxonomy)
  - Major tenets (more specific terms from military taxonomy)

Conclusion

After reviewing and quantifying the AI of the HDIAC's database, START, several deductions can be made. Firstly, the HDIAC scope area of Cultural Studies is a multidisciplinary area. The AI index shows that nearly half of the Cultural Studies documents cataloged in the database intersect with at least one other scope area. Because of the multidisciplinary status of the Cultural Studies literature being collected, it has become increasingly more important to develop a living, growing taxonomy specific to the HDIAC collection. Existing taxonomies within the Library of Congress and the DoD, although useful as a stepping stone, are not specific enough to delineate the HDIAC collection. Secondly, with projected collection growth of HDIAC literature at 5% per year, it is imperative a working framework for scope area taxonomies, which will create standards not just for collection but for search and retrieval, be created. The evolving taxonomies also benefit the end user by ensuring a more accurate return. Lastly, developing an organic taxonomy for HDIAC's scope areas allows literature gaps to be recognized and addressed, which in turn ensures a more robust collection.

Notes

[1] Dr. Geoffrey P Malafsky and Newman, Brain. Organizing Knowledge with Ontologies and Taxonomies, TECHi2 , Fairfax, VA. http://techi2.com/download/Malafsky%20KM%20taxonomy_ontology.pdf

[2] DTIC Online. http://dtic.mil/dtic/stresources/techreports/dticSearchTools/thesaurus_desc.html

[3] Library of Congress Subject Headings. The Library of Congress. 18 October 2014: http://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html

https://doi.org/10.26069/greynet-2025-000.542-tgj

Slovak Centre of Scientific and Technical Information
SCSTI

Achieve your goals with us

INFORMATION SUPPORT OF SLOVAK SCIENCE

SCIENTIFIC LIBRARY AND INFORMATION SERVICES

- technology and selected areas of natural and economic sciences
- electronic information sources and remote access
- depository library of OECD, EBRD and WIPO

SUPPORT IN MANAGEMENT AND EVALUATION OF SCIENCE

- Central Registry of Publication Activities
- Central Registry of Art Works and Performance
- Central Registry of Theses and Dissertations and Antiplagiarism system
- Central information portal for research, development and innovation - CIP RDI >>>
- Slovak Current Research Information System

SUPPORT OF TECHNOLOGY TRANSFER

- Technology Transfer Centre at SCSTI
- PATLIB centre

POPULARISATION OF SCIENCE AND TECHNOLOGY

- National Centre for Popularisation of Science and Technology in Society

IMPLEMENTATION OF PROJECTS

- National Information System Promoting Research and Development in Slovakia – Access to electronic information resources - NISPEZ
- Infrastructure for Research and Development - the Data Centre for Research and Development - DC VaV
- National Infrastructure for Supporting Technology Transfer in Slovakia - NITT SK
- Fostering Continuous Research and Technology Application - FORT
- Boosting innovation through capacity building and networking of science centres in the SEE region - SEE Science

www.cvtisr.sk
Lamačská cesta 8/A, Bratislava

Into the grey: A modified approach to citation analysis to maximize understanding of research impact*

Shannon L. Sibbald, Jennifer C. D. MacGregor, Marisa Surmacz, and C. Nadine Wathen (Canada)

Introduction

Academic authors and funders often want to know the "impact" of their publications, and this impact is generally judged by how and where the paper is cited in other academic works. This limited appraisal has been expanded in recent years as many are beginning to argue that non-academic publishing venues should be included in assessing the impact of academic publications. This is an issue of particular concern with the growing emphasis on 'knowledge translation' from the scientific literature to policy and practice applications [1],[2],[3] and to sources other than the traditional peer reviewed and indexed venues – i.e., translation into the "gray literature" [4].

In this paper, we describe the process of developing and applying a 'modified citation analysis' that builds on existing methods of examining a research paper's impact in two key ways: 1) by deliberately including gray literature in the citation analysis search process, and 2) by including quantitative and qualitative methods of analysis to gain a better understanding of how a research paper was used. By broadening the search and deepening the level of analysis, we suggest this new approach can better assess the impact of a given research paper – both within and outside of traditional peer reviewed venues. We begin with a review of gray literature, and then describe current methods for analyzing the impact of a research paper. Finally, we use a specific example to describe our new approach, highlight its potential for evolving the field of citation and impact analysis, and discuss future refinements and evaluation.

Gray Literature

Gray literature is an important aspect of scientific evidence as it "produc[es] and distribut[es] the seeds of new knowledge" [5]. Current definitions of gray literature are overly broad and lack clear lines of distinction between different producers, users and dissemination techniques. A commonly cited definition is: "information produced on all levels of government, academia, 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" (ICGL Luxembourg definition, 1997 - Expanded in New York, 2004, [6]). Another definition proposed by the *International Journal on Grey Literature* is: "the information and resources that do not categorically fall into what is available via standard, traditional or commercial publishing channels" [7]. Even these well-respected definitions provide only a vague understanding of what is classified as gray literature.

The distribution of gray literature in a multitude of mediums by individual scholars and scientists, research institutes, community and non-profit organizations, has become widespread since the advent of online repositories and other open access venues [8], [9], [10]. Organizations produce policy briefs, issue briefs or technical reports on specific content areas (e.g., health, environment, poverty) meant to inform decision-making that may not enter formal publication venues. This can be attributed to a bias towards publishing larger studies with specific results (i.e., positive, novel, generalizable), to the considerable time lag between research production and publication [11], or the production of gray literature for internal organization purposes.

Gray Literature as a Key Dissemination Strategy

Gray literature remains underused and undervalued. However, its importance and relevance is becoming apparent in areas such as policy development and research [12], [13], [14]. In the medical field, the well-regarded Cochrane systematic evidence review process requires the inclusion of any relevant gray literature as part of the evidence-base [15].

The exclusion of gray literature can skew the results of research syntheses that may have important ramifications for the accumulation and dissemination of scientific knowledge. For example, McAuley, Pham, Tugwell and Moher [16] and Hopewell, McDonald, Clarke, and Egger [4] argue that the exclusion of gray literature can influence the results of meta-analyses. In one example, by extending a search to include literature outside 'mainstream sources,' an estimated 29.2% more sources were found [17]. Gray literature is also important in many fields of research in both

* Originally published in the Journal of the Medical Library Association, 103 (1), pp. 49-54. – ISSN 1558-9439

understanding the theoretical underpinnings of findings [18], as well as identifying research gaps which produces research questions [19]. This is especially true in health research, where findings published in the gray literature often inform the development of clinical trials [16].

Gray literature is extremely valuable in providing timely communication on complex issues often using simple, actionable and comprehensive language, making it a key resource for stakeholders outside of academia [20], [21]. These works often include important information on context, policy decisions, and public interest that are of particular value to decision-makers [22]. It is apparent that while the amount and influence of gray literature is dramatically increasing, quality assurance, search methods, and impact measures for gray literature have not kept pace.

Quality of gray literature.

The sheer volume and accessibly of gray literature also raises the issue of the quality of these resources [23]. With the general absence of peer review, the only way to ensure the quality of gray literature is through critical assessment. Although the producers of gray literature may be experienced and knowledgeable in their field, there is essentially no formal quality control. There have been suggestions to improve the quality of the more 'traditional' types of gray literature such as technical and evaluation reports, strategic plans, policy briefs and others. An example is to include standard information in documents such as: (1) the process of creating the product, (2) the review process (if any) and (3) any conflicts of interest [24]. The Grey Literature International Steering Committee provides guidelines on what scientific and technical reports should look like. They should include a peer review process, ethical considerations, and publishing and editorial recommendations [25]. However, given the unregulated space that gray literature inhabits, it is unrealistic to expect all who create it to conform to such standards, leaving the end user responsible to assess its quality.

Accessing gray literature.

Searching for gray literature is challenging and can be time-consuming. There is currently no accepted strategy or protocol for doing so. While some gray literature becomes incorporated in the published literature and vice versa, much gray literature remains "fugitive" [26] and difficult to capture. Since gray literature falls under the broad definition of 'materials not published in a traditional way,' it can be found in numerous places, and requires many different search strategies to collect it. Although some of the commercially available databases, such as *Web of Science* and *Scopus*, contain gray literature (e.g., dissertations, conference proceedings), this is not their primary function and they likely capture only a fraction of existing material. Google Scholar is a good adjunct to these, but similarly does not have a primary focus on gray literature. As its search algorithms are unknown, it is difficult to understand and assess what has been included and excluded from these searches.

Another challenge with gray literature is the lack of bibliographic control or classification [27]. Unlike scholarly works such as journal articles, most gray literature is not indexed in databases. Although some databases do include gray literature, poor or inconsistent choice of titles or keywords can make such documents difficult to source, catalogue and store [28]. The 'scattered' and disorganized nature of gray literature presents particular difficulty for those wishing to incorporate it into a systematic and comprehensive search strategy, as is required for many types of literature review.

Current Methods for Analyzing Citation Impact

Several methods of analyzing individual study citations are currently used to determine impact in terms of penetration into the broader academic literature. These include methods such as *citation analysis*, *co-citation analysis, bibliographic coupling*, and a co-citation bibliographic coupling hybrid. In general, the purpose of these types of analysis are to track citations of publications and, in some methods, group similar papers together. Although these methods of impact analysis are widely used, they fail to capture gray literature and all are based solely on counts or quantitative comparisons of citations. That is, none delve deeper to explore *how* (as opposed to simply *how often*) a research paper has been used.

A relatively new method of citation analysis is the Becker Medical Library Model for Assessment of Research (*Becker Model*), which analyzes research impact through indicators such as "contribution to the knowledge base" and "change in practice" by tracking diffusion of research output and activities [29]. The Becker Model suggests locating gray literature materials such as non-peer-reviewed journal articles, trade publications and supplemental materials but does not assist in locating these items, and it still relies solely on quantitative counts of research outputs, albeit of more types. (For more information on this approach visit https://becker.wustl.edu/impact-assessment).

Modified Citation Analysis

We were interested in tracking the impact of a specific research article published in a high impact, peer reviewed journal [30]. This was the first randomized controlled trial (RCT) of universal screening in health care settings for exposure to intimate partner violence (IPV). We sought to examine how this new evidence was taken up in scholarly literature, as well as in broader policy, practice and advocacy. (For a description of the findings of that analysis, see [31]). After reviewing current methods for citation analysis (as above) we determined that none of these could effectively track the influence of a single paper through both commercially published and gray literature. For this reason, we developed a *modified citation analysis* method that would capture both traditional and gray literature and allow both quantitative and qualitative interpretive analysis on extracted data. Such an approach was important because (1) we knew that the content of the paper was of interest, and contentious within the academic realm and beyond, therefore limiting our search to only commercially-published sources would not provide an accurate portrayal of how the findings were used, and (2) we were interested, in *how* the paper was interpreted and used, therefore citation counts alone would be insufficient.

We searched the scholarly/academic and gray literature for sources with high potential to influence policy, practice or further research. Since we were primarily interested in the various uses of the study findings, we did not assess source quality or perform any formal results synthesis of the included materials. Preliminary scans of the literature helped determine the most effective subject headings and keywords/phrases to use for these searches. These terms were then applied using the appropriate database or other online search tools to retrieve articles, documents, media reports or other items that specifically cited or discussed the MacMillan et al. [30] paper. All retrieved results meeting inclusion criteria were stored in a reference management program.

The gray literature was searched using multiple resources, such as multidisciplinary databases, websites and point of care tools (Scopus, Medline Plus, MDConsult, UpToDate). Major healthcare associations and professional organizations likely to include related content were identified and their websites were individually searched. News media reports were searched by using news databases including Factiva, Lexis Nexis, Google News, and Proquest Canadian Major Dailies.

'Traditional' citation analysis statistics were computed using SPSS 20.0. These included the number of sources citing the MacMillan et al.[30] paper (by search method), and the number of MacMillan et al. [30] citations per source. Data analysis took place in three steps according to the specific research questions guiding the analysis: (1) content analysis of text extractions; (2) categorization of sources regarding their position on IPV screening; (3) coding of sources regarding how they defined IPV screening.

Adapting this approach

This method is theoretically adaptable to other contexts where tracing the uptake and interpretation of new knowledge would benefit from more than quantitative citation counts and inter-linkages. The gray literature search strategy would need to be specifically tailored to the area of interest and could involve more extensive searching depending on the discipline and topic area. For this reason, we recommend that experts in the field be involved in the process and the searches be undertaken by trained information professionals.

In addition to a tailored search strategy, the quantitative and qualitative coding of the text citing the original document will also depend on the specific questions being addressed. Some codes may apply to any paper; for example, anyone might code whether the paper is cited for a methodological or theoretical point, or might examine whether the citation appears in the Introduction, Method or Discussion section. Most codes, however, will be highly specific to the goals of the analysis.

Discussion

Our modified citation analysis has several advantages over more mainstream methods of citation analysis. By expanding our search strategy, we were able to find more citing sources. Specifically, using search techniques designed to track citations in the commercially published academic literature (e.g., Google Scholar 'cited by' tool, Web of Science) we found 80 sources citing the MacMillan et al. [30] paper. The gray literature searches (e.g., Google, news databases and association websites) found an additional 29 sources. Second, by moving beyond basic citation counts and coding *how* the paper was cited we were able to provide a richer understanding of its impact on subsequent research, as well as policy and practice guidance regarding intimate partner violence screening. It also provided a way to assess the evolution of the main debates in this area (for example, confusion caused by varying definitions of "screening").

This method requires additional resources. The much broader range of potential search venues demands more time and expertise. Delving into gray literature is a challenging task and requires planning and coordination, including consideration specific inclusion/exclusion search. Unlike database searching, common nomenclatures rarely exist for searching diverse gray literature sources, therefore the concept of consistency in search terms across sources is difficult to achieve. It is important to document all search strategies used in each source to attempt reproducibility in the ever-changing nature of online resources. Similarly, the addition of a qualitative analysis component necessitates multiple researchers to undertake coding, analysis, and interpretation that is rigorous and adheres to qualitative methodology standards [32]. More research using this method is required to assess its validity and utility across different applications.

Conclusion

The modified citation analysis approach appears well-suited to report how new research evidence is taken up, interpreted and used in published material, a quality that is valuable to those attempting to assess the effectiveness of knowledge translation strategies [3], [33]. We believe it is a useful extension of traditional citation analysis methods and addresses some of the gaps in the proposed Becker Medical Library Model for Assessment of Research [29]. Since there are no validated methods for conducting this method of citation identification, we created a new search and analysis strategy, building on the growing importance and acknowledgment of searching gray literature [34], [35].

This new approach adds to the ongoing discussion regarding the inclusion and analysis of gray literature resources in searching for different types of "evidence" to include in both academic research, practice and policy contexts. We hope that it encourages a more holistic approach to assessing "impact" of new research evidence.

References

1. Davis D, Evans M, Jadad A, Perrier L, Rath D, Ryan D, Sibbald G, Straus S, Rappolt S, Wowk M, Zwarentstein M. The case for knowledge translation: shortening the journey from evidence to effect. British Medical Journal. 2003;327(7405):33–35.
2. Grol R, Grimshaw J. From best evidence to best practice: Effective implementation of change in patients' care. Lancet. 2003;362(9391):1225–1230.
3. Kothari A, Wathen CN. A critical second look at integrated knowledge translation [Internet]. Health Policy. 2013;109:189-191.[cited 1Jan 2014]. <http://dx.doi.org/10.1016/j.healthpol.2012.11.004>.
4. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. Cochrane Database of Systematic Reviews. 2008 2, MR000010.
5. Marzi C, Pardelli G, Sassi M. A terminology based re-definition of grey literature. The Grey Journal: An International Journal on Grey Literature. 2011; 7(1):19-23.
6. Boekhorst A, Farace D, Frantzen J. Grey Literature Survey 2004 - A research project tracking developments in the field of grey literature. Paper presented at the Sixth International Conference on Grey Literature: Work on Grey in Progress, GL 2004 Conference Proceedings [Internet]; 2004 Dec; New York, US [cited 1 Jan 2014]. < http://www.open grey.eu/item/display/10068/697831>.
7. Emerald. International journal on grey literature [Internet]. 2000 [cited 1 Jan 2014]. <http://www.emeraldinsight.com/journals.htm?issn=1466-6189&>.
8. Farace DJ. Rise of the phoenix: a review of new forms and exploitations of grey literature. *Pub ResQ*. 1997;13(2):69-76.
9. Schopfel J, Farace D. Grey Literature. In Bates MJ & Maack MN. (Eds.) Encyclopedia of Library and Information Sciences, (3rd ed., pp. 2029-2039) Taylor and Francis. New York, NY: CRC Press York; 2011.
10. Artus HM. Old Wine in New Bottles? Developments in electronic information and communication: Structural change and functional inertia. The Grey Journal: An International Journal on Grey Literature. 2005;1(1):9-16.
11. Pappas C, Williams I. Grey literature: Its emerging importance. Journal Hospital Librarianship. 2011;11(3):228- 234.
12. Benzies K, Premji S, Hayden KA, Serrett K.. State-of-the-evidence reviews: advantages and challenges of including grey literature. Worldviews on Evidence-Based Nursing. 2006;3:55-61.
13. Turner A, Liddy E, Bradley J, Wheatley J. Modeling public health interventions for improved access to the gray literature. Journal Medical Library Association. 2005; 93(4):487-494.
14. Luzi D. Trends and evolution in the development of grey literature: a review. International Journal on Grey Literature. 2000;1(3):106 – 117.
15. Higgins JPT, Green S (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions* (Version 5.1.0). The Cochrane Collaboration [Internet]. 2011 [cited 1 Jan 2014]. < www.cochrane-handbook.org>.
16. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? The Lancet. 2000;356(9237):1228-1231.
17. Savoie I, Helmer D, Green CJ, Kazanjian A. Beyond medline: reducing bias through extended systematic review search. International Journal of Technology Assessment in Health Care. 2003;19(1):168-78.
18. Malina D, Nutt D. *Grey literature is a feminist issue: Women's knowledge and the Net*. International Journal on Grey Literature. 2000;1:18-27.

19. McKimmie T, Szurmak J. Beyond grey literature: How grey questions can drive research. Journal of Agricultural & Food Information. 2002;4(2):71-79.
20. Outten C. Gray literature [Internet]. 2012 [cited 1 Jan 2014]. <http://csulb.libguides.com/print_content.php?pid=120664&sid=1038841&mode=g>
21. Gardner P, Fong A, Huang R. Measuring the impact of knowledge transfer from public research organisations: A comparison of metrics used around the world. International Journal of Learning and Intellectual Capital. 2010;7(3-4):318-327.
22. Jacobs JA, Jones E, Gabella BA, Spring B, Brownson RC. Tools for implementing an evidence-based approach in public health practice. Prev Chronic Dis. 2012;9(11):1-9. DOI: http://dx.doi.org/10.5888/pcd9.110324.
23. Martin JLR, Pérez V, Sacristán M. Is grey literature essential for a better control of publication bias in psychiatry? An example from three meta-analyses of schizophrenia. European psychiatry. 2005; 20(8): 550-3.
24. Giustini D, Thompson D. Finding hard to finds: searching for grey literature. University of British Columbia. [Internet]. 2006 [cited 13 Aug 2014].
25. Grey Literature International Steering Committee. Guidelines for the production of scientific and technical reports: how to write and distribute grey literature [Internet]. 2006 [cited 1 Jan 2014]. <http://www.glisc.info/#2.3>.
26. Denda K.. Fugitive literature in the cross hairs: an examination of bibliographic Control and Access. Collection Management. 2002;27(2):75-86.
27. Ranger S. Grey Literature in special libraries: Access and use. Publishing Research Quarterly. 2005; 21(1):53-63.
28. Mason MK. Grey Literature: Its History, Definition, Acquisition, and Cataloguing [Internet]. 2013 [cited 1 Jan 2014]. <http://www.moyak.com/papers/grey-technical-literature.html>.
29. Sarli C, Dubinsky E, Holmes K. Beyond citation analysis: a model for assessment of research impact. Journal of the Medical Library Association. 2010;98(1):17-23.
30. MacMillan HL, Wathen CN, Jamieson E, Boyle MH, Shannon HS, Ford-Gilboe M, Worster A, Lent B, Coben JH, Campbell JC, McNutt LA, McMaster Violence Against Women Research Group. Screening for intimate partner violence in health care settings: A randomized trial. JAMA. 2009;302(5):493-501.
31. Wathen CN, Macgregor JC, Sibbald SL, Macmillan HL. (2013). Exploring the uptake and framing of research evidence on universal screening for intimate partner violence against women: A knowledge translation case study. Health Research Policy and Systems. 2013;11:13.
32. Bitsch V. Qualitative Research: a grounded theory example and evaluation criteria. Journal of Agribusiness. 2005;23(1):75-91.
33. Dembe AE, Lynch MS, Gugiu PC, Jackson RD. The translational research impact scale: development, construct validity, and reliability testing. Evaluation & the Health Professions. 2013.
34. Hanney SR, Gonzalez-Block MA, Buxton MJ, Kogan M. The utilisation of health research in policy-making: concepts, examples and methods of assessment. Health Research Policy System. 2003;1(1): 2.
35. Jones TH, Donovan C, Hanney S. Tracing the wider impacts of biomedical research: A literature search to develop a novel citation categorisation technique. Scientometrics. 2012; 93:125-134.

https://doi.org/10.26069/greynet-2025-000.542-tgj
TGJ Volume 11, Number 1 2015 On the News Front

Seventeenth International Conference on Grey Literature

'A New Wave of Textual and Non-Textual Grey Literature'

Royal Netherlands Academy of Arts and Sciences
Amsterdam, Netherlands on December 1-2, 2015

Announcement

As the internet becomes increasingly grey and every cloud now has a grey lining, there arises the need to address a new and challenging wave of textual and non-textual grey literature. GL17 will examine a number of new types of textual grey literature both web-based and submerged in the sea of social networks. No less attention will be drawn to the expanding quantity of non-textual grey literature accessible in visual, audio, and diverse data formats and frequencies.

Actually, in order to grasp this new wave of grey literature it may be even more advantageous to look at the convergence of these new types of textual and non-textual content rather than focus separately on each. The problems textual grey literature faced and addressed over the past quarter century are to a certain extent very similar to what non-textual grey literature faces today. The wide range of graphics, photographs, and other data-intensive grey literature is obscure, hard to find, and often short lived because it lacks proper indexing and sustained access. Such non-textual grey literature requires interpretation and documentation, which can in part be achieved by linking and crosslinking to their related textual counterparts. In this way, grey literature becomes leveraged and its value and return on investment made transparent.

While bridging textual and non-textual content is technically possible, it also requires an information policy in place that supports these new digital assets. Likewise, information professionals and practitioners must be able to (re)appropriate human resources and streamline their workflow in innovative ways. These should allow for content and feedback generated in social networks and in particular the information communities served.

The Seventeenth International Conference on Grey Literature welcomes authors and researchers to share their experience and vision on how to channel this new wave of grey literature by responding to the GL17 Call for Papers, http://www.textrelease.com/gl17callforpapers.html.

Conference Related Topics

New Types of Grey Literature Documents
New Forms of Non-Textual Grey Literature
The Convergence of Textual and Non-Textual Grey Literature
The Influence of Social Networks on Grey Literature
Innovative ways in Leveraging Grey Literature

NOTE To keep conference delegates and participants informed on the progress of the Seventeenth International Conference on Grey Literature, a quarterly newsletter, conference calendar, and further documentation will be made available on the conference site, www.textrelease.com. Please direct all correspondence to conference@textrelease.com.

TextRelease
GL17 Program and Conference Bureau

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

https://doi.org/10.26069/greynet-2025-000.542-tgj
TGJ Volume 11, Number 1 2015 *On the News Front*

Seventeenth International Conference on Grey Literature

'A New Wave of Textual and Non-Textual Grey Literature'

Royal Netherlands Academy of Arts and Sciences
Amsterdam, Netherlands on December 1-2, 2015

Call for Papers

Title of Paper:	**Related Topic(s):**
Author Name(s):	**Telephone:**
Organization(s):	**Fax:**
Postal Address:	**Email:**
Post/Zip Code - City - Country:	**URL:**

Guidelines for Abstracts

Participants who seek to present a paper at GL17 are invited to submit an English abstract between 350-400 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, topic(s) 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 the 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.

Conference Related Topics

- ☐ **New Types of Grey Literature Documents**
- ☐ **New Forms of Non-Textual Grey Literature**
- ☐ **The Convergence of Textual and Non-Textual Grey Literature**
- ☐ **The Influence of Social Networks on Grey Literature**
- ☐ **Innovative ways in Leveraging Grey Literature**
- ☐ **Other Conference Topic**

Due Date and Format for Submission

Abstracts in MS Word must be emailed to conference@textrelease.com on or before **April 24th 2015**. The author will receive verification upon its receipt. Shortly after the Program Committee meets in mid-May, authors will be notified of their place on the conference program. This notice will be accompanied by further guidelines for submission of full text papers, accompanying research data, PowerPoint slides, and required Author Registration.

TextRelease
GL17 Program and Conference Bureau

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

https://doi.org/10.26069/greynet-2025-000.542-tgj

Korea Institute of Science and Technology Information (KISTI)

English version - http://en.kisti.re.kr/

*** Vision**

World-class information research institute creating values for customers

*** Main functions**

*** Management and service of Korean R&D reports**

KISTI exclusively manages, preserves, and serves Korean R&D reports for citizens and government officials. It provides Korean R&D reports and their information with National science & Technology Information Service (NTIS) and National Discovery for Science Leaders (NDSL).

***Contact information**

KISTI email address: hcpark@kisti.re.kr

Headquarters: Tel : +82-42-869-1004, 1234 Fax: +82-42-869-0969

Author Information

Berchum, Marnix van **41**
Marnix van Berchum is Head of Data Services at DANS, responsible for the coordination of the services DANS offers (including EASY, the narcis.nl portal, and the Dutch Dataverse Network). After his studies in Musicology at Utrecht University, he started working at Utrecht University Library. In the years 2010-2012 he also worked at SURF. At both employers he was mostly involved with projects related to Open Access and innovations in scholarly communications (e.g. OpenAIRE). Marnix combines his work at DANS with a PhD trajectory at Utrecht University, in which he wants to apply the concepts and methods of network theory on the dissemination of music in the 16th century. Marnix is Associate Director of the CMME Project.
Email: marnix.van.berchum@dans.knaw.nl

Biagioni, Stefania **17**
Stefania Biagioni graduated in Italian Language and Literature at the University of Pisa and specialized in data processing. She is currently a member of the research staff at the Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI), an institute of the Italian National Research Council (CNR) located in Pisa. She is head librarian and member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She has been the responsible of ERCIM Technical Reference Digital Library (ETRDL) and currently of the PUMA (PUblication MAnagement) & MetaPub, a service oriented and user focused infrastructure for institutional and thematic Open Access repositories looking at the DRIVER vision, http://puma.isti.cnr.it. She has coauthored a number of publications dealing with digital libraries. Her activities include integration of grey literature into library collections and web access to the library's digital resources, including electronic journals and databases. She is a member of GreyNet since 2005. Email: stefania.biagioni@isti.cnr.it

Crowe, June **47**
June Crowe is the Senior Researcher and Group Manager, Open Source Research 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. Her current position, subject matter expert cultural studies, is with the Homeland Defense and Security Information Analysis Center (HDIAC) in Oak Ridge, TN. Her primary interests are Grey Literature research, socio-cultural and S&T research, digital repositories, and open source intelligence tools.
Email: jcrowe@iiaweb.com

Gelfand, Julia M. **29**
Julia Gelfand has served as Applied Sciences and Engineering Librarian at the University of California, Irvine Libraries since 1981, where she was recognized as a Distinguished Librarian. She has presented and written extensively on various issues related to grey literature, collection development, scholarly communication, library vendor relationships, emerging technologies, scholarly publishing, social media, and organizational change in academic libraries. She is active in many professional organizations and currently serves on the Board of Directors of the Association of College and Research Libraries (ACRL) Division of the American Library Association (ALA), as Secretary of Section T for Information, Computers & Communication of the American Association for the Advancement of Science (AAAS), and is a member of the Acquisitions & Collection Development Section of the International Federation of Library Associations (IFLA).
Email: jgelfand@uci.edu

Giannini, Silvia **17**
Silvia Giannini graduated and specialized in library sciences. Since 1987 she has been working in Pisa at the Institute for the Science and Technologies of Information “A. Faedo” of the Italian National Council of Research (ISTI-CNR) as a librarian. She is a member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She is responsible of the library automation software “Libero” in use at the CNR Research Area in Pisa and coordinates the bibliographic and managing activities of the ISTI library team. She cooperates in the design and development of the PUMA (PUblication MAnagement) & MetaPub, an infrastructure software for institutional and thematic Open Access repositories of published and grey literature produced by CNR.
Email: silvia.giannini@isti.cnr.it

Goggi, Sara **17**
Sara Goggi is a technologist at CNR-ILC. She started working at ILC in 1996 working on the EC project LE-PAROLE for creating the Italian reference corpus; afterwards she began dealing with the management of several European projects and nowadays she is involved with organisational and managerial activities mainly concerning international relationships and dissemination as well as organization of events (e.g. LREC conference series). Currently one of her permanent activities is the editorial work for the international ISI Journal Language Resources and Evaluation, being its Assistant Editor. Since many years (from 2004) she also carries on research on terminology: this field of research starting intersecting with work on Grey Literature in 2011 at GL13.
Email: sara.goggi@ilc.cnr.it

MacDonald, Bertrum H. **7**
Bertrum H. MacDonald is a Professor of Information Management in the School of Information Management at Dalhousie University, Halifax, Canada. With a background in science (BSc, Biology), history of science (MA), and information science (MLS, PhD), he pursues research that investigates the dissemination and use of scientific information in historical and contemporary contexts. He pursues interdisciplinary research, particularly within the Environmental Information: Use and Influence initiative (www.eiui.ca), since this work tackles large questions from the point of view of several relevant disciplines. He has been Director of the School of Information Management and Associate Dean (Research) in the Faculty of Management at Dalhousie University. He can be seen speaking about research projects at local, national, and international levels, and he holds executive positions with national and international associations. In 2004, he won the International GreyNet Award with his research colleagues, Ruth Cordes and Peter Wells. He is the recipient of the Marie Tremaine Medal, the highest award of the Bibliographical Society of Canada, and he was awarded a Dibner Research Fellowship at the Smithsonian Institution in 2001. Email: bertrum.macdonald@dal.ca

Pardelli, Gabriella **17**
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

Sherline, Crystal **47**
Crystal Sherline is a Technical Information Analyst for Homeland Defense and Security Information Analysis Center (HDIAC), a researcher at Information International Associates (IIA), and doctoral candidate at the University of Tennessee, Knoxville. She received a Master's degree from the University of Maryland, and an MLIS from Florida State University, and will be receiving her Ph.D. from the University of Tennessee in July 2014. She has experience in operations of both public and academic libraries. Her interests are varied, but consist of primarily research, including Network Analysis, Grey Literature, HCI, Social Science, and data sharing. Email: csherline@iiaweb.com

Soomai, Suzuette S. **7**
Suzuette S. Soomai is an interdisciplinary doctoral student in the Faculty of Graduate Studies. Her research focuses on the role of fisheries information, published extensively by national and international governmental organisations, in policy making for fisheries management. She holds a Master of Marine Management from Dalhousie University and an MPhil (Zoology/Aquatic Ecology) and BSc from the University of the West Indies, Trinidad and Tobago. Prior to her studies at Dalhousie University, Suzuette was a Fisheries Officer with the Ministry of Agriculture, Land and Marine Resources in Trinidad and Tobago and was a member of fisheries scientific working groups led by the United Nations Food and Agriculture Organisation and the Caribbean Regional Fisheries Mechanism. She has special expertise in fisheries resource and coastal zone management and has completed national and regional fish stock assessments while interacting with a diverse range of stakeholders in the Caribbean. In 2012, she was awarded a major doctoral fellowship from the Social Sciences and Humanities Research Council of Canada.
Email: suzuette.soomai@dal.ca

Tsang, Daniel C. **29**
Daniel C. Tsang has served as Social Sciences Bibliographer and Data Librarian at the University of California, Irvine Libraries since 1986 and was named Distinguished Librarian. He is a well published author in mainstream and alternative media and invited speaker concentrating on anti-surveillance and social movements. Recognized for his scholarship and contributions to the documentation of small and alternative presses, and for his distinguished service to data librarianship, he was awarded the William F. Flanigan Award for Distinguished Service to ICPSR in 2013. He has graduate degrees from the University of Michigan. Email: dtsang@uci.edu

Wells, Peter G. **7**
Peter G. Wells is an Adjunct Professor in the School for Resource and Environmental Studies and the Marine Affairs Program, Faculty of Management, and a Senior Research Fellow, International Ocean Institute, at Dalhousie University, Halifax, Canada. He holds a BSc (Biology) from McGill University, a MSc (Zoology) from the University of Toronto, and a PhD (Zoology/Aquatic Toxicology) from the University of Guelph. After over 30 years of public service, he took early retirement from Environment Canada, in 2006, to focus on his research. He concluded his work with Environment Canada as Head, Coastal and Water Science, and Senior Research Scientist, Coastal Ecosystems. He served the United Nations Joint Group of Scientific Experts on Marine Environmental Protection in various capacities for 14 years, and taught an international marine pollution course in Bermuda for 16 years. His current research includes choosing indicators for coastal ecosystem health, utilizing blue mussels for monitoring chemical contaminants, and evaluating the use and influence of marine environmental information in environmental policies and decision making. He was elected Fellow of the American Association for the Advancement of Science, and is a recipient of Dalhousie's highest award for teaching excellence by part-time faculty. oceans2@ns.sympatico.ca

https://doi.org/10.26069/greynet-2025-000.542-tgj

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.

https://doi.org/10.26069/greynet-2025-000.542-tgj
TGJ Volume 11, Number 1 2015

An International Journal on Grey Literature

'Public Awareness To Grey Literature'

TGJ Volume 11, Number 1 2015