

Towards a Prague Definition of Grey Literature

Joachim Schöpfel

Charles-de-Gaulle University Lille 3

Abstract

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

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

Results: Based on the state of the art and the survey data, we make a proposal for a new definition of grey literature ("Prague definition") with four new essential attributes: "Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body." The attributes and challenges are discussed.

Note on the author

Joachim Schöpfel is head of the department of information and communication sciences at the Charles de Gaulle University of Lille 3 and researcher at the GERiCO laboratory. He is interested in scientific information, academic publishing, open repositories, grey literature and usage statistics. He is member of GreyNet and EuroCRIS.

Université Charles-de-Gaulle Lille 3, UFR IDIST, BP 60149, 59653 Villeneuve d'Ascq Cedex, France.

joachim.schopfel@univ-lille3.fr

1. Introduction

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

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

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

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

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

Material that "may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents" (U.S. Interagency Gray Literature Working Group): this concept meets Mackenzie Owen's observation that "grey does not

imply any qualification (but) is merely a characterization of the distribution mode" (1997).

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

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

2. Methodology

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

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

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

The corpus consists of 32 documents selected from 219 GL conference communications published on the *OpenSIGLE* website¹ (sampling = 15%), through a content analysis of titles, abstracts and full texts (Fig. 1).

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

Figure 1: Corpus of GL communications

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

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

Between 2005 and 2010, *The Grey Journal* published 101 articles referenced in the online RefDoc database². From these articles, we selected three original articles (not published in GL proceedings) with substantial debate on grey literature (sampling= 3%) and added them to our GL corpus (Fig. 2).

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

² <http://www.refdoc.fr>

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

Figure 2: Corpus of TGJ articles

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

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

2.2. Empirical evidence: online survey

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

The survey was carried out in October 2010. The questionnaire was made available online³. The survey population included 1390 information specialists and scientists from GreyNet's distribution list. Promotion was also done on Twitter and through the social networks LinkedIn, Viadeo and Facebook.

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

3. Results

3.1. Content analysis of papers on grey literature

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

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

3.1.1. Essential attributes of the definition of grey literature

Which are the main features mentioned in the sample corpus? Two-thirds of the studies insist on **dissemination** as the central characteristic

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

of grey literature, e.g. the unconventional or unusual mode of distribution through non-commercial channels (see Figure 3).

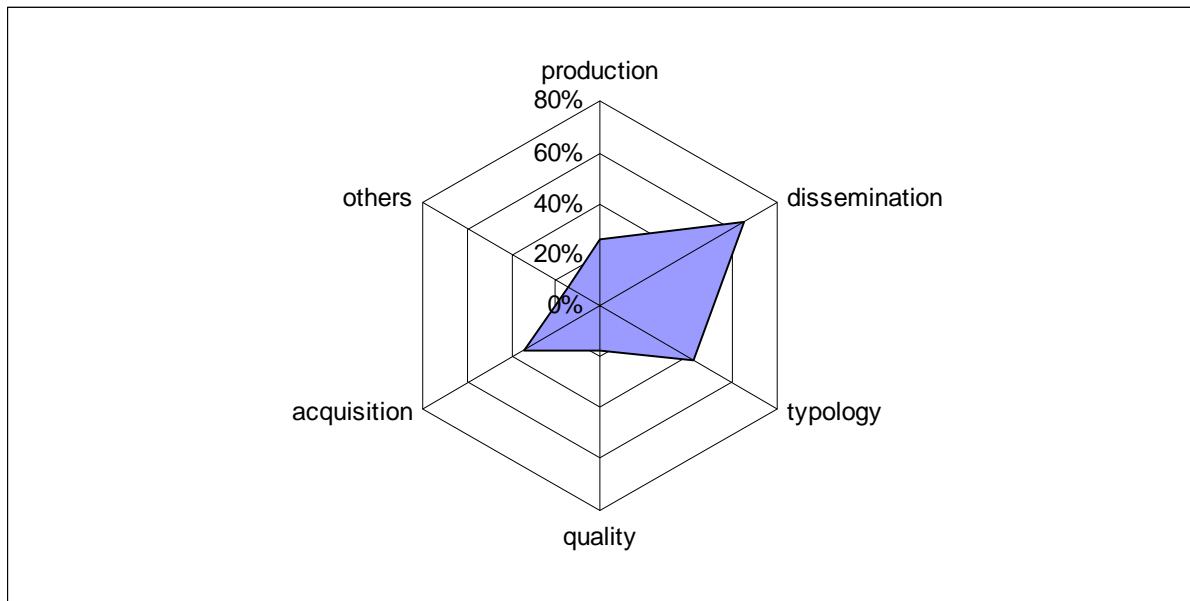


Figure 3: Main topics of GL definitions in corpus

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

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

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

About 40% studies adopt a **typological approach**. In such a definition the operative issue is which type of document belongs to grey literature, which doesn't? Librarians mostly agree that theses and dissertations, conference proceedings, reports and working papers are grey. But what

about patents and preprints, blogs, datasets, and tweets? Grey literature “embraces such things as non-conventional literature, archival material, fugitive material, non-book material and unpublished documents” (Kufa, 1993). Luzi (1995), Luzi et al. (2003) and Ranger (2004) worked on new forms of scientific information, such as electronic conferences, protocols, websites or digital datasets. Stock & Schöpfel (2008) evaluated the presence of more traditional items – theses, reports, working papers etc. – in open archives. Sulouff et al. (2005) provide a cross-disciplinary comparison of different types of grey literature associated with academic departments and disciplines, derived from survey data.

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

Fewer authors raise the question of **quality**, and they do so in a controversial way. Erwin (2006) observes, “quality (of grey literature) continues to be suspect even among researchers” and is not surprised that “because of the range of quality in grey literature (...) grey literature continues to be absent from most formal academic collection development policies”. In contrast, Wessels (1997) argues that “much grey literature is published by prestigious organizations whose names are a guarantee for quality” and puts forward its uniqueness⁴.

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

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

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

Open source: Crowe & Davidson (2008) place grey literature into the intersection “of open source and intelligence” and describe the way how

⁴ The controversy about quality is reminiscent of the more recent IPCC Amazongate debate. See for instance, <http://www.guardian.co.uk/environment/georgemonbiot/2010/jul/02/ipcc-amazongate-george-monbiot>

information that is openly available and is lawfully obtained may become classified and become a source of intelligence.

3.1.2. Functions of grey literature

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

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

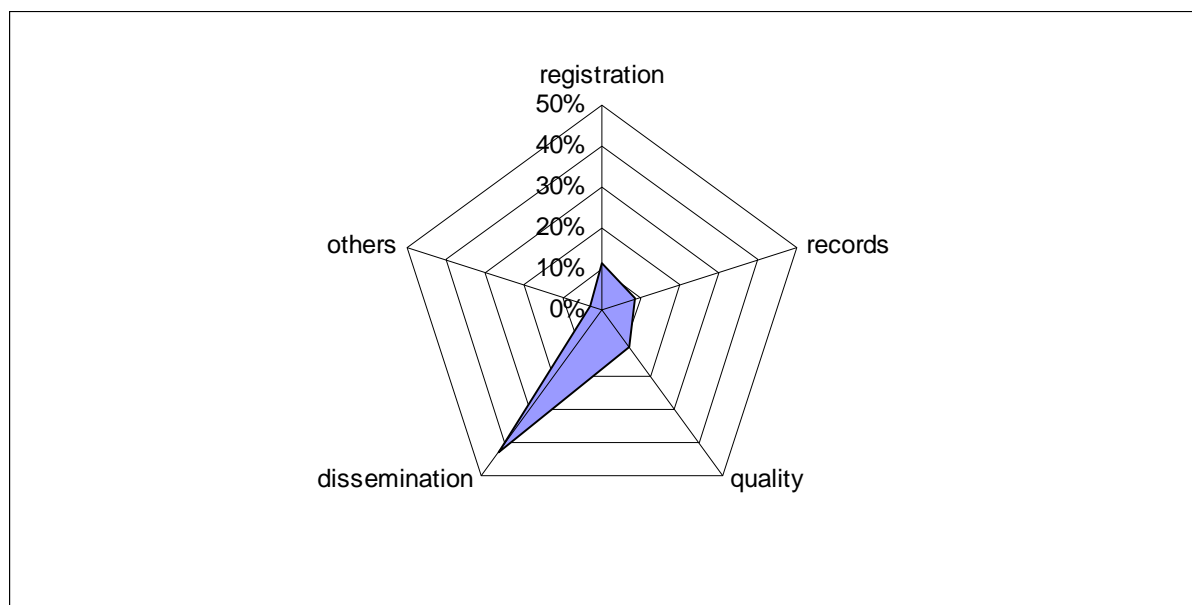


Figure 4: Main functions of GL in corpus

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

Kufa (1993) identified an interesting specific role for grey literature in African countries insofar as it may be more relevant to local needs and conditions⁵. Following Kufa, non-commercial publishing may be useful to communicate and access locally produced materials that are not published elsewhere.

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

3.2. Survey results

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

3.2.1. Attitudes towards the current definition

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

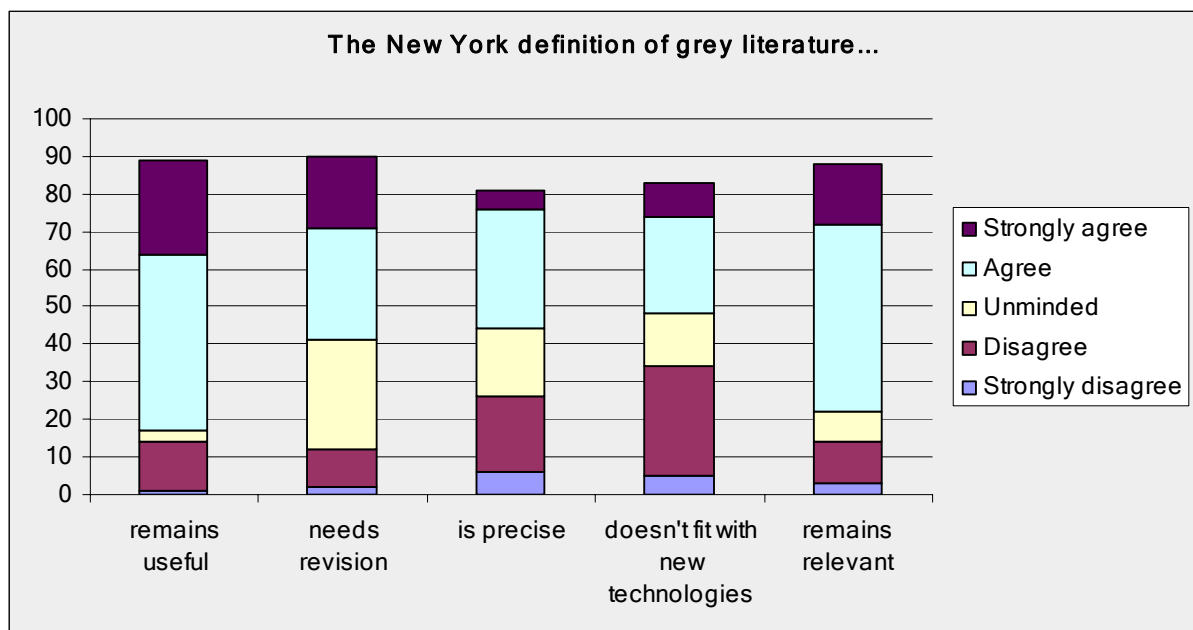


Figure 5: Opinions on the current NY definition

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

3.2.2. Elements for a revised definition

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

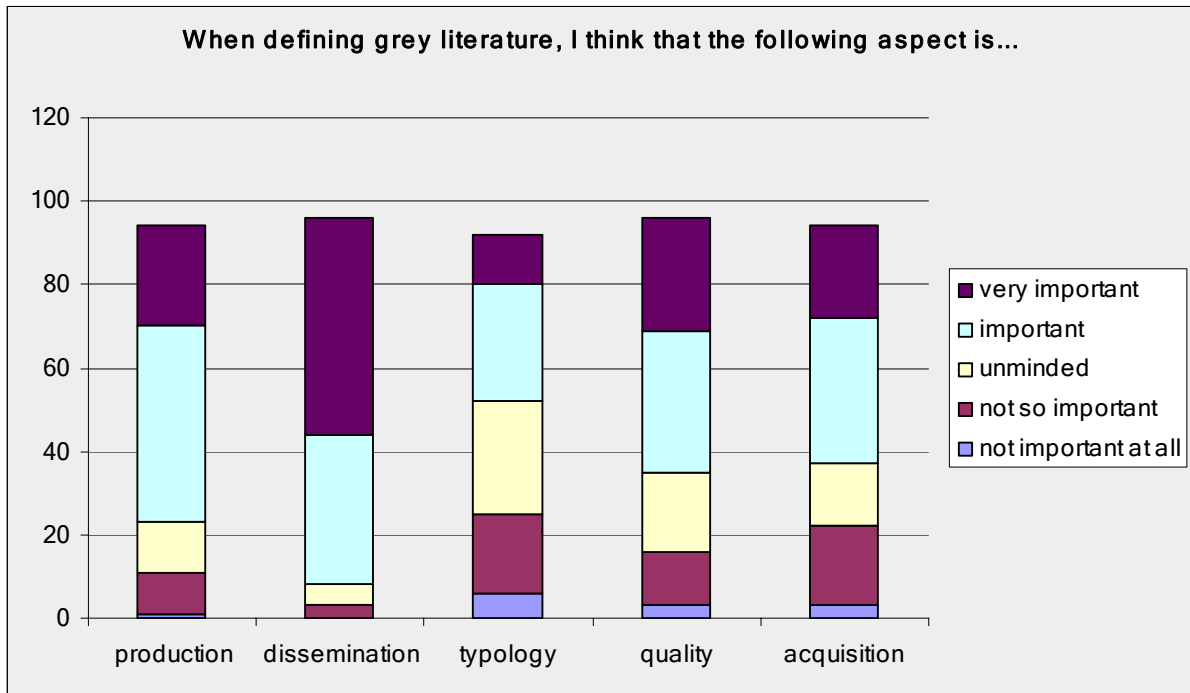


Figure 6: Opinions on essential attributes of a GL definition

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

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

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

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

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

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

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

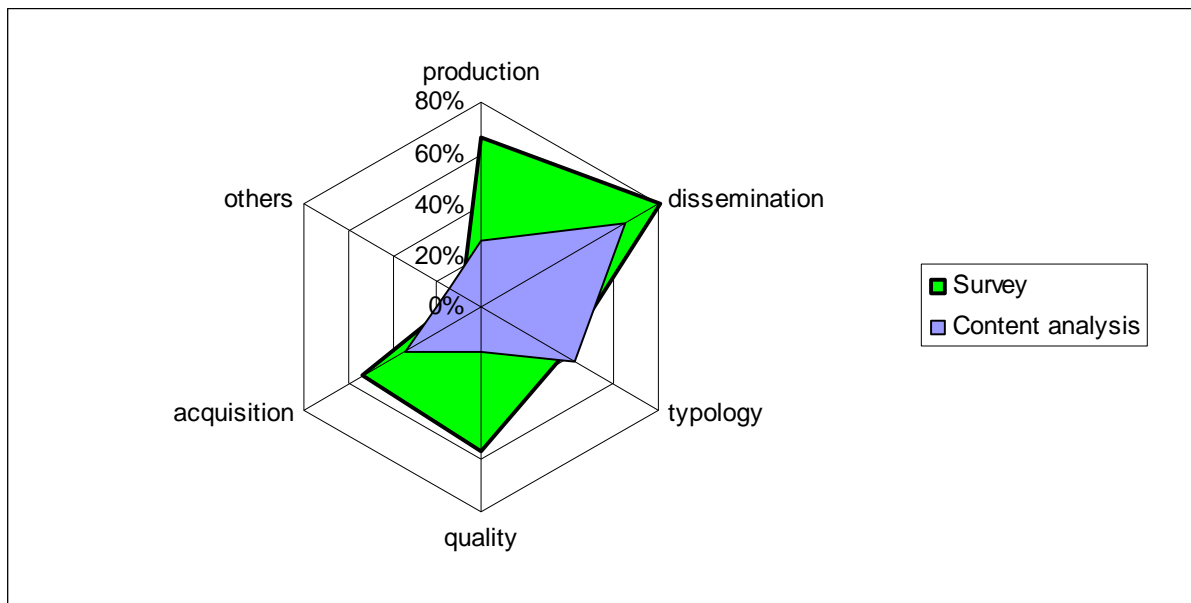


Figure 8: Main topics of GL definitions in corpus (content analysis) and survey

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

3.2.3. Perceived functions

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

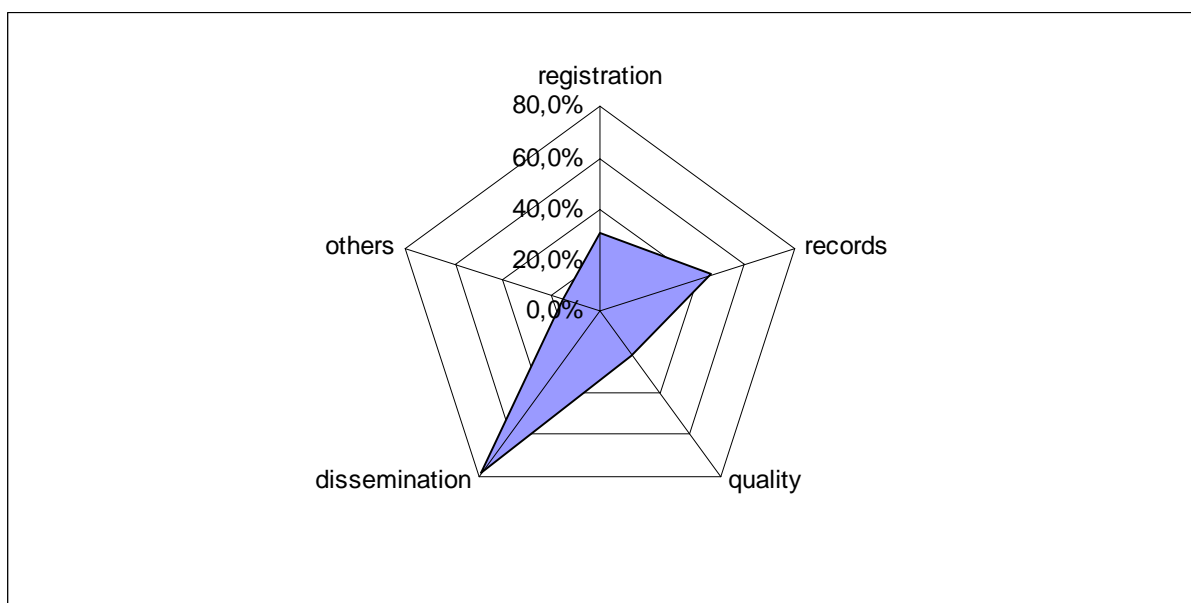


Figure 9: Main functions of GL in survey

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

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

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

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

3.2.4. A world without grey literature?

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

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

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

- Can you imagine a world with only white literature?

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

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

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

4. Discussion

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

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

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

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

4.1. Fallacies of the New York definition

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

The New York definition is over-broad. It does not allow for a clear distinction between grey literature and other “objects”, and it applies to items that are not part of the extension of the definition. Too many papers based on the common definition fail to differentiate between grey and other literature and/or include “stuff” that may be grey but clearly is NOT literature.⁶

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

The New York definition bears a risk of circularity. This means that it partially assumes a prior understanding of what grey literature is. For instance, how is it possible to define the meaning of “where publishing is not the primary activity of the producing body”? Simply spoken, the New York definition allows for interpretations such as “grey is that which is not white (nor black)”⁷. This circularity may be useful for community cohesion and acceptance but not for furthering understanding.

⁶ With the words of one respondent of our survey: “Keeping our focus. Now that GL has become 'sexy' - people expand the boundaries beyond what makes sense. Raw data is not GL. Ephemera is not GL. The entire Web is not GL. I think losing the focus and overexpanding runs the risk of losing credibility when making the arguments for the value of GL - especially in the sciences, when up against the elitism of more formal publication (...)”

⁷ “Grey literature can only be defined in its relation to formally published or ‘white’ literature (...)” (Artus, 2003). But then, what is white, and where does it begin?

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

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

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

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

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

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

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

Are digital objects grey, such as datasets, emails, technical copies, wikis, blogs, tweets or even Websites⁸, etc., or are they even literature?

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

"Defining grey literature as the 'opposite' to commercial publishing is wrong. (That is, if commercial publishing means publishing for a profit). It is wrong because many non-profit organisations publish on exactly the same basis as commercial publishers (e.g. American Chemical Society) - and their publications are not grey. To my mind, grey literature is all scholarly work that is published without a formal peer-review (or equivalent) process outside the traditional journal and book channels. Therefore, grey literature includes data outputs, analytical tools, software as well as the more usual 'classic' pre-print or working paper outputs."

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

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

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

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

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

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

4.3. From typology to quality

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

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

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

Quality means quality assurance and control and intrinsic scientific value. A recent synthesis on reports states that "it is (...) a misconception

that grey literature is not peer reviewed (...) reports often proceed through several levels of internal and agency review before final printing" (Seymour, 2010). The 2005 initiative GLISC⁹ or "Nancy group" was expected to establish a label of quality to reports. Other grey items such as theses, dissertations or communications undergo procedures of labelling and/or validation that guarantee a minimum quality level. Our own surveys on grey literature in France and Belgium reveal up to 60% items with some kind of quality control.

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

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

4.4. Grey literature without library?

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

In fact, the real customer of grey literature, since the invention of the concept, is and always has been the librarian, the LIS professional who tries to collect it. This is a striking point when going through the papers written on grey literature in the last two decades. "Searching and accessing the grey literature (...) have always been a great challenge *for librarians and documentalists*" (Natarajan, 2006)¹⁰. This is a problem not for scientists, students or other patrons but for LIS professionals – this is the point.

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

⁹ <http://www.glisc.info>

¹⁰ Italics added by JS

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

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

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

Today, institutional repositories have started to take over some of the traditional roles of library holdings. In terms of function, they bear some equivalency with grey literature itself, as their main role consists in dissemination and, to a lesser extent, preservation. This may explain the growing interest of the grey literature community for this special type of open archives¹¹. A new definition should or could mention this.

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

5. Conclusion

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

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

¹¹ See the recent study of Luzi (2010).

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

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

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

"Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body."

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

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

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

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

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

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

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

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

(3) The so-called *Fourth Paradigm*¹², e.g. data-intensive science and the access to datasets that together generate a trend to transform and/or marginalise literature (documents).

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

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

6. References

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

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

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

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

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

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

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

¹² See Hey et al. (2009).

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

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

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

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

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

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

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

Annexe A - Corpus

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

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

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

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

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

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

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

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

- J. Crowe & T. S. Davidson (2008). 'The 'Grey' Intersection of Open Source information and Intelligence'. In *Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society*, 8-9 December 2008.
- R. Di Cesare & C. Sala (1995). 'The use of grey literature in the agricultural economics field : a quantitative analysis'. In *Second International Conference on Grey Literature: Grey Exploitations in the 21st Century*, 2-3 November 1995.
- P. A. Erwin (2006). 'The Messy World of Grey Literature in Cyber Security'. In *Eighth International Conference on Grey Literature: Harnessing the Power of Grey*, 4-5 December 2006.
- J. Gelfand (1993). 'Academic libraries and collection development implications for grey literature'. In *First International Conference on Grey Literature: Weinberg Report 2000*, 13-15 December 1993.
- J. Gelfand (1999). 'Grey Literature: Plausible Applications for Distance Education Practices'. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature*, 4-5 October 1999.
- P. Gokhale (1997). 'Grey Literature Varieties - Definitional Problems'. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information*, 13-14 November 1997.
- H. von Hofe (2004). 'Towards a Genealogy of Grey Literature via Newton's Journals'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress*, 6-7 December 2004.
- K. G. Jeffery (1999). 'An Architecture for Grey Literature in a R&D Context'. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature*, 4-5 October 1999.
- K. G. Jeffery & A. Asserson (2006). 'Hyperactive Grey Objects'. In *Eighth International Conference on Grey Literature: Harnessing the Power of Grey*, 4-5 December 2006.
- J. C. Kufa (1993). 'Grey literature in academic research in Developing Countries : Problems and Prospects'. In *First International Conference on Grey Literature: Weinberg Report 2000*, 13-15 December 1993.
- D. Luzi (1995). 'Internet as a new distribution channel of scientific grey Literature: the case of Italian WWW servers'. In *Second International Conference on Grey Literature: Grey Exploitations in the 21st Century*, 2-3 November 1995.
- D. Luzi, et al. (2003). 'The communication flow of research projects results'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information*, 4-5 December 2003.
- B. H. MacDonald, et al. (2003). 'Grey Literature in the Life of GESAMP, an International Marine Scientific Advisory Body'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information*, 4-5 December 2003.
- J. Mackenzie Owen (1997). 'Expanding the Horizon of Grey Literature'. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information*, 13-14 November 1997.
- J. C. McDermott (1995). 'Defining grey literature for the 21st Century'. In *Second International Conference on Grey Literature: Grey Exploitations in the 21st Century*, 2-3 November 1995.

- M. Nahotko (2007). 'Some Types of Grey Literature: A Polish Context'. In *Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, 10-11 December 2007*.
- M. Natarajan (2006). 'Grey literature: problems and prospects for collection development in e-environment'. *The Grey Journal* 2(2):100–105.
- L. P. Pavlov (2003). 'The Commercialization of Research Findings Documented in Grey Literature'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information, 4-5 December 2003*.
- L. P. Pavlov (2004). 'The Problems of Grey in the Context of Postmodern Science and Culture'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, 6-7 December 2004*.
- S. L. Ranger (2004). 'Grey Literature in Special Libraries: Access and Use'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, 6-7 December 2004*.
- J. Schöpfel (2006). 'Observations on the Future of Grey Literature'. *The Grey Journal* 2(2):67–76.
- M. S. Seadle (1999). 'Grey Copyrights for Grey Literature: National Assumptions, International Rights'. In *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature, 4-5 October 1999*.
- D. B. Simpson (1995). 'Grey Literature: the challenges for an increasingly important body of research literature'. In *Second International Conference on Grey Literature: Grey Exploitations in the 21st Century, 2-3 November 1995*.
- C. Stock & J. Schöpfel (2003). 'Grey Literature in an Open Context: From Certainty to New Challenges'. In D. Farace & J. Frantzen (eds.), *GL5 Conference Proceedings. Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information.*, pp. 199–202. TextRelease.
- C. Stock & J. Schöpfel (2008). 'Grey literature in French digital repositories: a survey'. In *Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society, 8-9 December 2008*.
- P. Sulouff, et al. (2005). 'Learning about grey literature by interviewing subject librarians'. *The Grey Journal* 1(3):131–136.
- R. H. A. Wessels (1997). 'An EAGLE's eye-view of grey literature research'. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information, 13-14 November 1997*.

Annexe B - Survey

During previous conferences in the GL-Series, several authors have raised questions concerning the (re)definition of the term grey literature.

In this exploratory survey, an effort is made to assess attitudes towards the currently held definition and to gather elements that may lead to a redefinition of the term grey literature.

This online survey will take about ten minutes of your time to complete. Your response is greatly appreciated.

Dr. Joachim Schöpfel

Head, Department of Information Sciences

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

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

2. The 2004 amended definition of the term grey literature reads as follows:

“Information produced and distributed on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishers i.e. where publishing is not the primary activity of the producing body”.

What do you think about this definition?

The current definition of grey literature

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

Strongly agree/Agree/Disagree/No Comment

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

When defining grey literature, I think that the following aspect is...

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

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

- other (please specify)

4. What do you think about the following statements?

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

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

5. Do you agree with the following statements?

Grey literature will continue to play a significant role alongside commercial publishing.

The borderline between "grey" and "white" (commercial) literature will become increasingly indistinct.

The proportion of "grey" documents published on the Web will continue to increase.

The Internet will encourage a greater diversity in the types of "grey" resources available.

Bibliographic control of grey literature will remain problematic.

Open archives will offer more appropriate services and functions for at least some segments of grey literature.

Some organizations will develop tools and services to aid in the efficient exploitation of grey resources on the Web.

Searching and collecting grey literature will become as straightforward as it is for journals and books.

New tools for collecting, depositing, and archiving will make grey literature less ephemeral and volatile than in the past.

6. Can you imagine your world without grey literature?

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

8. You are a...

Scientist

Scholar

Information professional

Student

Other: