# How scientific papers mention grey literature: a scientometric study based on Scopus data

Joachim Schöpfel, Hélène Prost

# How scientific papers mention grey literature: a scientometric study based on Scopus data

Joachim Schöpfel

Hélène Prost

## Abstract

Purpose: The purpose of the paper is to provide empirical insight into the use of the term and concept of grey literature in recent scientific papers.

Design/methodology/approach: We conducted a scientometric analysis in the Scopus database, on 1,606 papers dealing with grey (or gray) literature and published in 2018. Additionally, we performed a content analysis on a random subsample of 70 papers in open access.

Findings: The part of papers dealing with grey literature is low (0.05%) but steadily rising. They are from over 100 countries and a long tail of institutions, covering namely medical and health sciences and related topics. The dominant document type is systematic reviews, defining grey literature generally thought of as "unpublished", "not peer reviewed" and "not in databases" and meaning most of the time all kinds of reports and conference papers. A large variety of sources and options how to retrieve grey literature is mentioned, including Google and Google Scholar, specialized digital libraries, relevant websites, handsearching in bibliographic references and contact with experts in the field.

Research limitations/implications: The study is limited to papers indexed in the Scopus database, mainly journals, written in English, with a bias in favor medical and life sciences.

Originality/value: There is no recent study on the real usage of the term of grey literature in a large sample of academic papers.

## Key words
Grey literature, systematic reviews, academic publishing, scientometrics.

## Classification
Research paper

## Introduction

Some decades ago, nobody spoke about grey literature because the concept simply did not exist. It is historical (Schöpfel 2011). Even today, the meaning of the term "grey (or gray) literature" is uncertain and depends on the context and the people. Launched in 1992, the international grey literature network service, GreyNet International[1], promotes the New York definition of grey literature: "*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, i.e., where publishing is not the primary activity of the producing body*" (Schöpfel & Farace, 2010), together with its updated version, the so-called Prague definition: "*Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body*" (Schöpfel 2011).

---

[1] http://www.greynet.org/

However, outside of this community of experts, the meaning of grey literature is widely unknown or more or less misunderstood. An analysis of articles, conference papers and reports about grey literature reveals a large diversity of how the term is applied and interpreted, with sometimes rather simplistic ideas about what grey literature means and covers (Schöpfel 2011). The understanding of the concept depends largely on who is speaking about it. Initially, the concept of grey literature was created and promoted by information professionals, especially by acquisition librarians working in the sector of scientific and technical information (Auger 1989) and by people working in the field of scientific or economic intelligence. But there are many more people using the term of grey literature, and others speaking about relevant objects, like theses and dissertations, reports, working papers etc., apparently without knowledge of the concept (Prost & Schöpfel 2014).

A recent study assesses the way how LibGuides, a content management and library knowledge-sharing system used by several libraries worldwide, inform about grey literature, and it recommends some minimal requirements in order to increase the awareness and understanding of grey literature, including for instance a reference definition, a typology of grey items and helpful resources and tools for the discovery and the retrieval of grey literature (Carlson et al. 2017).

This paper adds new empirical evidence about the real impact of the concept of grey literature in the academic world. The leading research questions are:

- What is the part of published academic papers dealing with grey literature in an explicit way?
- Who are the authors of these papers, in terms of country, discipline etc.?
- How are they defining grey literature?
- What are the types of grey literature they mention, which examples do they provide, what are the resources and tools for the discovery and retrieval of grey literature?

## Methodology

The scientometric study is based on the Scopus database. The analysis and the export of the sample corpus are dated 7 October 2019. Statistical analyses and figures are performed with MS Excel.

The basic research question in the Scopus database is "grey literature" OR "gray literature" in "All fields", including metadata (title, keywords, abstract) and the full text, and limited to 20 years (1999-2018).

The scientometric assessment is performed with the Scopus discovery tools, on the following variables:

For the whole period (1999-2018):

- Number of publications per year

For the year 2018:

- Authors
- Affiliations
- Countries
- Document type
- Source title
- Source type
- Subject

The content analysis of the definition of grey literature is limited to the freely available documents (open access), with a random sample of 70 items (10%). The assessed criteria are, cited definitions (references), semantics (description, characteristics, examples), types, sources.

## Results

### Evolution 1999-2018

Scopus contains 8,853 publications mentioning grey (or gray) literature, published between 1999 and 2018 (figure 1). The average annual increase is 28%, with a sharp rise from 26 (1999) to 1,606 publications (2018).

[Insert Figure 1]

The annual increase is well above the overall growth of the Scopus database (average annual growth = 5%). Yet, the part of grey literature is low, 0.02%, compared to the overall number of publications in Scopus for the same period (45,059,595). However, this part is rising, from 0.01% or less before 2010 to 0.05% in 2018.

In 2018, 95% of the items were published in academic journals, mainly as articles (46%) and reviews (44%). 4% are conference papers published in proceedings; 1% are book chapters.

Out of the 1,606 papers published in 2018, 704 papers are freely available, corresponding to 44% open access.

### Authors 2018

The authors of the 1,606 papers mentioning grey literature are from 113 different countries. Three countries – United Kingdom, USA and Canada – represent half of the published items; they are followed by Australia, Brazil, Italy, Germany, The Netherlands, France, Switzerland, Spain and China (together 80%).

There is a long tail of affiliated institutions. In terms of published papers, the ten highest ranked institutions are mainly from Canada and Australia: University of Toronto, University of Alberta, McMaster University, University of Oxford, University of Ottawa, University of Queensland, Monash University, University of Calgary, University of Sydney, and University of Melbourne. Also, nearly all papers are published in English.

### Research fields 2018

Four out of five papers are published in journals covering medical science and health, including nursing, pharmacology, dentistry and veterinary (80%, see figure 2).

[Insert Figure 2]

30% of the papers are from social sciences, humanities, arts or economics, 29% are from natural sciences, mathematics etc. 9% are from engineering sciences.

Again, we can observe a long tail of many hundreds of academic journals where these papers are published. The most important journals, in terms of the number of published papers, are from medical and life sciences, such as *British Medical Journal Open*, *PLoS One*, *Cochrane Database of Systematic Reviews*, *Systematic Reviews*, *BioMed Central Public Health*, *BioMed Central Health Services Research* and *Medicine*. However, among the ten most highly ranked journals are also *The Grey Journal* and the *International Conference Series on Grey Literature*, both from Textrelease Amsterdam (affiliated with GreyNet International, the organizer of the international conferences on grey literature), and *Environmental Evidence*, a journal covering the field of environmental studies.

## Definitions, semantics

How do these papers define grey literature? Do they define the term? In the subsample of 70 freely available papers published in 2018 (10% of the 2018 papers in open access), only five papers supply an explicit definition of grey literature, such as the New York definition from GreyNet International (Farace & Schöpfel 2010) or the older definition from Peter Auger (Auger 1989).

In spite of defining grey literature, many papers provide descriptive elements (attributes) or a negative definition, by exclusion, leaving it to the reader to guess what grey literature is or is not. The most widely used attributes are "*unpublished*" and "*not peer reviewed*". Not peer review generally means of doubtful quality; one paper explicitly states that "*access quality and heterogeneity of grey literature are stumbling blocks to its inclusion in a systematic review*" but admits, nevertheless, that "*exclusion of grey literature was a limitation of this review*".

Other authors describe grey literature not only as ephemeral items, with "*internal dissemination*" or "*difficult to locate if they have survived*", inconsistent, unconfirmed, "*without statistically significant findings*", but also as "*ongoing studies*". One study on librarian-mediated literature searches for systematic reviews characterizes grey literature as "*time consuming*" (Bramer et al. 2018). Another refers to grey literature as "*publicly accessible Internet resources*".

Often, authors mention that they have been searching for grey literature, i.e. "*documents not in bibliographic (electronic) databases*", especially Scopus, Web of Science or Medline. Grey literature is described as "*complement to databases*", not in journals, or even more simply, "*not articles*". These are negative definitions: they do not say anything about what grey literature really is; they only say roughly what it is not. One paper opposes grey literature and scientific literature, as if grey would mean non-scientific.

## Types and examples

Even if they do not define the term of grey literature, many papers provide some indications of what they mean by it, or simply, of what they found when searching for grey literature. In terms of document types or categories, grey literature often means some kind of report or conference papers:

- published reports
- technical reports
- governmental and institutional reports
- official police or medical expert reports
- policy papers (strategy)
- conference proceedings
- meeting papers
- presentations

Other categories are less often mentioned, such as:

- white papers
- conference abstracts
- posters
- theses and dissertations
- fact sheets
- standards
- technical notes
- guidelines

- guidance notes

Yet, some authors also consider "*primary material*", unpublished observational studies and clinical trials, opinion pieces, editorials, letters to the editor, blogs, a forum, "*research objects*" or a governmental "*gazette*" as grey literature.

When the papers, especially the systematic reviews, provide statistics about the search results, the part of grey literature is low, somewhere between 0.1% and 1%. However, in other studies, grey literature represents nearly the half of the retrieved items, for instance 3,000 or more references.

The part of grey literature in systematic reviews is even lower after the selection of relevant documents. One study for instance found 52 grey documents, of which only four were used. Other systematic reviews even seem to keep none of the retrieved grey documents.

## Sources

How do authors find grey literature? How do they discover relevant items for their systematic reviews or research studies? Following the papers' methodology sections, we can distinguish five different options:

- General or academic search engines: Google Search (advanced), Google Scholar.
- Specialized databases and digital libraries: this may be international databases like SIGLE (discontinued), OpenGrey, DissOnline.de, or the ProQuest Dissertation and Theses Database, but also PubMed, Cochrane Library, the Latin American and Carribean Health Sciences Literature cooperative database (LILACS) and Scopus; institutional platforms such as the Digital Library of the Federal University of Minas Gerais (UFMG); national discovery tools like the thesis and dissertations bank of the Brazilian Coordination of Improvement of Personnel of Higher Education (CAPES), the Brazilian Digital Library of Theses and Dissertations; or disciplinary servers such as governmental or international trial registers, the AIDSFree Resource Library, or the USAID Development Experience Clearinghouse. A couple of studies also mention the New York Academy of Medicine's Grey Literature Report as a valuable source of relevant grey literature in the field of medical sciences.
- Reference screening: handsearch or "*snowballing*" in articles' bibliography sections is a third option to get references of grey literature: "*We also screened the reference lists of included studies and relevant systematic reviews*".
- Relevant websites: the manual search in relevant websites (i.e. websites of relevant institutions) depends on the expertise of the authors who know which websites or organizations are relevant or not or, by default, on the advice from other experts (see below). Some examples: World Health Organization (WHO), World Bank, Joseph Rowntree Foundation, Age UK, Alzheimer's Association, InterGen, Manchester Institute for Collaborative Research on Ageing, the Andalusian Health Service, BioBran Research Foundation, etc.
- Experts: the contact with "*key experts*", "*authors of conference proceedings that are of potential relevance*", "*most productive researchers in the field*" or simply "*key organizations*" in the area of interest is another way to identify potentially relevant grey items and get information on any ongoing or unpublished research in the field.

One study warns of a potential problem with the Google search engines (Piasecki et al. 2018), a problem that would add uncertainty and doubt and lack of quality to grey literature. In particular, Piasecki et al. criticize the lack of transparency of the Google search algorithms and the bias induced by the personalized search. Especially uncontrolled personalization bears the risk of non-replicable systematic reviews.

## Discussion

### Limitation

The study analyzes papers indexed in the Scopus database. Scopus is the largest academic scientometric database worldwide. However, this limitation to Scopus introduces a systematic bias in favor of journal articles, of research papers published in English, and of medical and life sciences. In other words, our study takes into account a large part of the mainstream academic publishing, especially in the fields of science, technology and medical sciences but neglects non-English papers, reports, dissertations and unpublished conference papers, and underestimates the scientific production in social sciences and humanities.

### Real impact

Twenty years ago, a couple of studies questioned the future of grey literature in the environment of new information and communication technologies (Schöpfel 2006). The main argument was that the web would increase the findability and the accessibility of dissertations, reports, working papers etc. in a way that would make the concept of grey literature meaningless. The scientometric assessment of the Scopus database shows that the concept is still in use, even more than before, with a steadily increasing number of papers mentioning or dealing with grey literature in one way or the other.

However, the scientometric results reveal too that the part of scientific papers that mention grey literature is very low, below 0.1%. In 2018, only one academic paper out of 2,000 contained the term of grey (or gray) literature. Insofar as most of the papers include some kind of state of the art or literature overview, and (at least potentially) cite other types of documents than books or journals; this percentage seems rather low, too low for a concept that is defined in order to characterize a significant part of scientific publishing (Schöpfel & Farace 2010).

### Disciplines

The content of the former European grey literature database SIGLE, today OpenGrey[2], confirms that grey literature exists in all scientific disciplines, a fact that has been confirmed by other studies (see Schöpfel & Farace 2010). However, the results of this new study show another reality. Most of the relevant papers are published in the field of medical and life sciences, and only a smaller part covers other subjects.

The reason is probably not a deficit of grey literature in social sciences and humanities, physics, chemistry, mathematics, informatics, agriculture and so on. The explanation is rather a lack of awareness not of the existence of the resources themselves but of the global concept of grey literature. Medical and life sciences seem the only large field where grey literature is an explicit part and even core element of scientific methodology. Grey literature "*may form a vital component of evidence reviews such as systematic reviews and systematic maps, rapid evidence assessments and synopses (...) The inclusion of grey literature is a central tenet of systematic review methodology, which aims to include all available documented evidence and reduce susceptibility to bias*" (Haddaway et al. 2015). The widely accepted PRISMA guidelines[3] for systematic reviews make the search for grey literature a mandatory part of the procedure (Moher et al. 2009). Also, the Cochrane methodological guidelines for systematic reviews consider the search for grey literature (e.g. "*reports/dissertations/theses databases and databases of conference abstracts*") as "*highly desirable*", in order to "*reduce the risk of publication bias and to identify as much relevant evidence as possible*" (Higgins et al. 2019). Therefore, many papers presenting results from a systematic review describe the search for grey

---

[2] http://www.opengrey.eu/
[3] Preferred Reporting Items for Systematic Reviews and Meta-Analyses http://www.prisma-statement.org/

literature, even if the results may not always be convincing: few documents only, and not always of doubtless quality.

The inclusion of grey literature in a standard procedure of scientific work is the reason for a couple of methodological studies on grey literature: how to search it and where to look, how to deal with it, how to increase search efficiency and avoid biased results, and so on (see Haddaway et al. 2015, Piasecki et al. 2018). This too seems specific to the field of medical and life sciences; also, the only widely known alert service for grey literature, i.e. the Grey Literature Report[4] (discontinued), was produced in the field of medical sciences by the New York Academy of Medicine. However, there are similar papers outside of medical sciences, such as Augusto et al. (2010) in soil science.

## Definition

Finally, the results of the content analysis reveal a kind of paradox: a general and shared understanding of grey literature as unpublished, non-peer reviewed papers, not or less indexed by the usual academic databases; and at the same time, the lack of a reference definition and of a common and accepted knowledge if what the concept of grey literature means.

There may be different reasons for this situation. One reason may be the fact that the definition of grey literature is the result of professional experience and academic research in the field of library and information sciences, especially by academic librarians, with a limited impact on the scientific work of researchers. This is not specific to grey literature; researchers most often ignore other library concepts, such as catalog formats or metadata which does not mean that they do not use them – yet, they do not speak of them in their papers.

Another reason may be that while the importance and relevance of grey literature depend on the subject and methodology of scientific work, in a general way, it is less important and less relevant than journal and book publishing, especially for academic assessment (evaluation) and ranking. This may explain why researchers pay less attention to this part of scientific information and why they do not need a precise and explicit understanding of it.

A last reason may be related to the general understanding of grey literature as not-peer reviewed. The quality and value of grey literature has always been a major argument for acquisition and collection building; in the same time, it was a central element of criticism and suspicion, as if the lack of peer review means deficit of trustworthiness: "*Grey information may be at high risk of bias*". Such latent suspicion can be read between the lines of many papers mentioning grey literature, especially when they distinguish between journal articles indexed in the bibliographic databases (trustworthy, because peer reviewed) and other documents which need more attention, care and control, because untrustworthy, because not peer reviewed. It is too time consuming for an uncertain result. This is another reason not to spend too much mental effort for a deeper understanding of this part of scientific information.

This latent suspicion of untrustworthiness distinguishes (and separates) the scientific communities from the small group of librarians and scientists working in the field of grey literature who know better about the value of grey literature and who are dedicated to the improvement of its quality. Yet, some scientists of the survey sample are very aware of the potential of grey literature; for instance, one paper states that publication bias across studies was minimized by including grey literature and consulting with experts. Another example: Baines & Regan de Bere (2018) explain that grey literature "*was included on the basis that it would help validate the results of published literature searches, identify the most up-to-date information and respond to criticisms of existing literature operating in*

---

[4] https://www.nyam.org/library/collections-and-resources/grey-literature-report/

*silos (…) such literature often includes the most up-to-date information from those 'working' on the ground, and thus, in turn, presents alternative perceptions to those available in peer-reviewed literature*". Here, grey literature is considered not at risk of bias but on the contrary, useful for reducing bias.

## Conclusion

The purpose of the paper is to provide empirical insight into the use of the term and concept of grey literature in recent scientific papers. A scientometric analysis of the Scopus database, completed by a content analysis of a subsample of journal articles in open access, provides empirical evidence about the real impact of the concept of grey literature in the academic world:

- What is the part of published academic papers dealing with grey literature in an explicit way? We identified 8,853 relevant papers for the 20-years-period from 1999 to 2018, corresponding to 0.02% of the total number of academic papers. The number of papers mentioning grey literature rises from 26 (1999) to 1,606 (2018), with an average annual growth of 28%, well above the general growth of the Scopus database. In 2018, 46% of the papers are regular journal articles, 44% are reviews.
- Who are the authors of these papers, in terms of country, discipline etc.? The authors of the 1,606 papers published in 2018 come from 113 different countries, with three countries – United Kingdom, USA and Canada – representing half of the published items. There is a long tail of affiliated institutions; the highest ranked institutions are mainly from Canada and Australia. Nearly all papers are published in English. 80% of the papers are published in journals covering medical science and health, including nursing, pharmacology, dentistry and veterinary.
- How are the authors defining grey literature? Very few papers provide an explicit and precise definition. Many papers describe some general attributes, especially "*unpublished*", "*not peer reviewed*" and "*not in databases*". The search for grey literature is considered as time-consuming, with most only few or no result.
- Which are the types of grey literature they mention, which examples do they provide, what resources and tools are used for the discovery and retrieval of grey literature? The typical "grey items" are all kinds of reports, policy papers, theses and dissertations and conference papers. Many methods, tools and sources are described on how and where to search for grey literature, including specialized digital libraries and databases, relevant websites, reference screening and expert advice.

Grey literature is (still) part of academic research and communication, even more than before. It is still a matter of attention and investigation (and doubts, too). Searching for relevant grey literature outside of the usual discovery tools is (still) not as simple and easy as the mainstream journal articles and books, and it requires time, domain-specific knowledge and networking. "*Considerable efforts are typically required within systematic reviews to search for grey literature* (…)" (Haddaway et al. 2015). The results – ongoing research, unpublished results, non-academic research, policy papers etc. – generally give good reason for this effort. However, while the need for expertise, time and so on still seems the same as before (see Auger 1989), the digital transformation of the research environment and the scientific information landscape shifts much if not all of the required effort from the library to the scientists: for the description, dissemination and conservation as well as for the discovery and retrieval of grey literature.

## Bibliography

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

Augusto, L. et al. (2010), "Is 'grey literature' a reliable source of data to characterize soils at the scale of a region? A case study in a maritime pine forest in southwestern France, *European Journal of Soil Science*, Vol. 61 No. 6, pp. 807–822. https://doi.org/10.1111/j.1365-2389.2010.01286.x

Baines, R. L. and Regan de Bere, S. (2018), "Optimizing patient and public involvement (PPI): Identifying its 'essential' and 'desirable' principles using a systematic review and modified Delphi methodology", *Health Expectations*, Vol. 21 No. 1, pp. 327–335. https://doi.org/10.1111/hex.12618

Bramer, W. M., Rethlefsen, M. L., Mast, F. and Kleijnen, J. (2018), "Evaluation of a new method for librarian-mediated literature searches for systematic reviews", *Research Synthesis Methods*, Vol. 9 No. 4, pp. 510–520. https://doi.org/10.1002/jrsm.1279

Carlson, K. N., Schöpfel, J. and Vaska, M. (2017), "Grey Literature LibGuides or LibGuides about Grey Literature: A Two Continent Environmental Scan of Common Themes & Trends", poster presented at *GL19 International Conference on Grey Literature*, 23-24 October 2017, Rome (Italy).

Haddaway, N. R., Collins, A. M., Coughlin, D., and Kirk, S. (2015), "The role of google scholar in evidence reviews and its applicability to grey literature searching", *PLoS ONE*, Vol. 10 No. 9, e0138237. https://doi.org/10.1371/journal.pone.0138237

Higgins, J., Lasserson, T., Chandler, J., Tovey, D., Thomas, J., Flemyng, E., and Churchill, R. (2019), *Methodological Expectations of Cochrane Intervention Reviews (MECIR)*. Retrieved from https://community.cochrane.org/mecir-manual

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (2009), "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement", *Annals of Internal Medicine*, Vol. 151 No. 4, pp. 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Piasecki, J., Waligora, M. and Dranseika, V. (2018), "Google Search as an Additional Source in Systematic Reviews", *Science and Engineering Ethics*, Vol. 24 No. 2, pp. 809–810. https://doi.org/10.1007/s11948-017-0010-4

Prost, H. and Schöpfel, J. (2014), "Grey communities: a scientometric approach to grey literature, in and outside of GreyNet", *The Grey Journal*, Vol. 10 No. 1, pp. 38-49. https://hal.archives-ouvertes.fr/hal-01226212

Schöpfel, J. (2006), "Observations on the Future of Grey Literature", *The Grey Journal*, Vol. 2 No. 2, pp. 67–76. https://hal.archives-ouvertes.fr/sic_00168998

Schöpfel, J. (2011), "Towards a Prague definition of grey literature", *The Grey Journal*, Vol. 7 No. 1, pp. 5–18. https://hal.archives-ouvertes.fr/sic_00581570

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