

Grey Literature and Professional Knowledge Making

Joachim Schöpfel

▶ To cite this version:

Joachim Schöpfel. Grey Literature and Professional Knowledge Making. Börjesson, Lisa; Huvila, Isto. Research Outside The Academy. Professional Knowledge-Making in the Digital Age, Palgrave Macmillan, pp.137-153, 2018, 978-3-319-94176-9. 10.1007/978-3-319-94177-6_8. hal-01872353

HAL Id: hal-01872353 https://hal.science/hal-01872353v1

Submitted on 19 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Grey literature and professional knowledge making

Joachim Schöpfel

University of Lille, GERiiCO laboratory

Abstract

What does grey literature mean? What role does it play in the production and dissemination of practitioner knowledge? How do reports, presentations and communications, working papers and other un-published material contribute to professional, extra-academic knowledge making? The following paper tries to provide some elements for a better understanding of grey literature, with examples from different collections and disciplines. Moreover, it puts the focus on critical issues like standards, identifiers and quality, and it discusses the impact of open science, i.e. the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society, amateur or professional.

Keywords

Grey literature, academic publishing, scientific information, professional publishing, intermediation, open science

About grey literature

Grey literature is a concept invented some decades ago by acquisition librarians looking for specific categories of documents difficult to get – in particular documents a library cannot buy or subscribe to via vendors. At the beginning, the term covered principally scientific and technical reports (Auger 1989) produced by research and industry, often from foreign countries. These documents are not classified, confidential or protected someway ("black literature") but open source material. Often, the information professional knows that these items exist, that they have been disseminated (often in small numbers) and are available – but the problem is how to get them because they cannot be acquired through the usual commercial channels of the scientific information market, like books or journals ("white literature").

In spite of many different and sometimes obscure and misleading definitions, grey literature is still a library concept, generated and conditioned by acquisition policy and collection building. 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. Four aspects are essential for the understanding of grey literature: it consists of text documents ("literature"); it is the work of the mind, protected by intellectual property; it is of interest to a user community, i.e. it has a minimum quality level; and it is conditioned by (inter)mediation (acquisition/selection for a collection). Therefore grey literature can be defined as follows (Schöpfel 2010): "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."

On categories and boundaries

Grey literature is not limited to scientific and technical information (STI) produced by academia, i.e. Higher Education and public research organizations. It is a broad term, an umbrella concept for resources issued by many other information sources, such as governmental agencies, local authorities, administrations, corporate companies, NGOs, unions, international organizations, consulting offices... Academia is only one source of grey literature. Many grey items, like reports,

white papers, presentations or minutes, are produced by "all levels of government, (...) business and industry".

However, not all of them can be considered as research output, many are business, political or administrative documents without any scientific ambition. Not all are results of extra-academic research, i.e. research conducted in institutional settings other than universities and other research performing organizations (Börjesson 2016a). Yet, they may be interesting for science. The boundaries are not always clear. While parliamentary records, governmental green and white papers or annual reports from companies are clearly non-scientific literature, PhD dissertations, final project reports or working papers produced by research teams are part of scientific information. But, what can we say about marketing studies, reports by investment management firms or papers delivered by business consultants? Some grey legal information resources produced as part of the three branches of the US government clearly have scientific character, e.g. expert reports, environmental impact statements & assessments (National Environmental Policy Act), permits (Clean Air Act, Clean Water Act, Endangered Species Act) or inspection reports (Occupational Health and Safety Act) (Rucinski 2015). A study on NGOs in several areas of health-related activities showed that about 25% of their documents can be considered as scientific grey literature (papers, factsheets...) (Rudasill 2009).

Also, the line between academia and non- or extra-academic research is not easy to draw. If extra-academic research is defined as research conducted by professionals, not by scientists, what can we say about the research in library and information sciences conducted by academic librarians and data officers? How do we categorize the results produced by mixed research teams in digital humanities, often composed by scientists (historians, linguists, philologists...), computer experts and engineers? And, finally, how do we deal with the great number of documents based on R&D projects conducted with corporate companies?

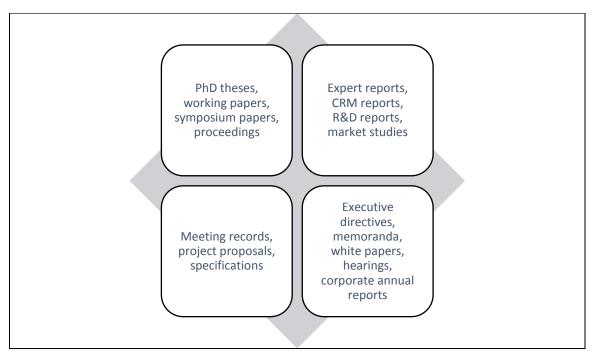


Figure 1: Categories of grey literature (upside research, downside professional; left side academic, right side extra-academic)

Figure 1 illustrates the different categories of grey literature, with some typical and representative items. The figure however shows clear separations between the four categories while the reality is more complex, with more or less blurred boundaries. Also, as mentioned above, the greyish nature of a particular document is *not* so much determined by its typology as it is by its dissemination and

acquisition mode in the value chain of scientific information. This makes it difficult to estimate the part of grey literature as a whole and of its different categories. In the following, we will provide some examples of extra-academic grey literature and discuss critical issues of this specific way of professional knowledge making.

Grey literature from extra-academic research – some examples

As said before, grey literature is a library concept, determined by acquisition and collection. This means that any assessment of grey literature should start at the point of mediation, i.e. the holdings (catalogues), databases, portals, repositories etc. In fact, there are two questions: do libraries (and other STI services) collect this kind of material, and can this material be reliably identified as such? Traditionally, such a study would (and still does) include library collections (holdings) and their catalogues and records. Today, the focus is larger and embraces digital libraries, gateways (portals) and above all institutional repositories, with a special focus on their metadata.

Collections and gateways

First observation: yes, the major scientific libraries and information centres do collect this literature as a service for their academic and professional customers and patrons. Sometimes, this acquisition policy, collection building and gateway function is explicitly identified as "B2B", as business to and for business, like for instance the German National Library of Science and Technology (TIB) which provides access to their rich national and international report collection and catalogue via a special search space called "TIB for Research and Business". Here are some examples.

The National Technical Reports Library (NTRL)¹ run by the US Department of Commerce contains nearly 1.9m technical reports, many of them from universities and research institutions but others are issued by extra-academic structures, like federal administrations, US agencies and offices, even if it appears impossible to make a reliable estimation of their number or relative importance. Also the boundaries are not clear-cut, as it seems that many of these structures host academics and produce academic grey literature, in particular PhD dissertations.

The Educational Research Information Center (ERIC)² indexes a large variety of grey literature, and their list of approved sources includes work produced or funded by federal departments, state or local agencies, policy organizations, non-profit organizations, federal technical assistance providers, professional associations, and international or foreign organizations; along with university affiliated programs, research organizations, and state and district research offices. But again, no estimation of this extra-academic part of indexed grey materials seems available or possible.

Other libraries with large report collections are the British Library, the aforementioned German TIB and the Korean Institute of Scientific and Technical Information (KISTI). Many of the more than 330,000 German reports of the TIB holdings are from authors affiliated to universities or research organization but others are produced by governmental or regional agencies, ministries or foundations, outside of the usual academic landscape. The KISTI collection of Korean R&D programs (NTIS) holds thousands of reports produced by extra-academic institutions; the Small and Medium Business Administrations alone represents 18% of the total R&D programs (Choi et al. 2015).

The National Repository of Grey Literature (NRGL)³ of the Czech National Library of Technology (NTK) receives deposits and records (more than 400,000) from various institutions from the Czech Republic, such as grant agencies, state administration, and the commercial sphere. Nevertheless the part of

¹ https://ntrl.ntis.gov/NTRL/

² https://eric.ed.gov/?nonjournals

³ http://nrgl.techlib.cz/

universities and research organizations is very high (with 88% BA, Master and PhD theses and dissertations), and only a small percentage of items with scientific character (<5%) comes from the extra-academic sector.

The OpenGrey repository⁴, heir of the former System for Information on Grey Literature in Europe (SIGLE) contains more than 1m records from the major European countries. Especially in the fields of economics, business and political science, searching the database reveals many extra-academic institutions and organizations which have produced grey research documents such as country reports, case studies and investigations from NGOs, economic intelligence units, foundations, agencies, the Institution of Chartered Accountants in England and Wales or the UK Cabinet Office.

For all these examples, it is often easier to browse through the list of affiliated or networking institutions than to evaluate the scientific nature of their documents or estimate their number. The few public available figures lead us to think that the extra-academic scientific output may be significant but remains a minority issue. In a second time, we'll present some examples from specific disciplines and research domains.

Disciplines and domains

Health: The health sector is a significant "consumer" of grey literature, especially for systematic reviews of specific treatments, procedures, syndromes etc. The New York Academy of Medicine Grey Literature Reports⁵ exploit resources from 2,100 publishers; only few (<5%) are clearly identifiable as universities or colleges but this does not mean that others are not affiliated with academia. Yet, many sources are professional, extra-academic, and crucial for health-related knowledge. NGOs are another primary source of health-related grey literature, in particular for researching healthcare information for developing countries (Crowe et al. 2010). This literature is rich and varied and includes field notes, surveys, newsletters, annual reports and so on. Perhaps the most impressive example is the grey literature database Northern Light⁶, a discovery portal for life science conference abstracts, with more than 2.5m conference abstracts and posters from over 3,500 conferences held around the world since 2010, many of them produced by corporate companies and organizations outside of academia.

Nuclear energy: Launched by the International Atom Energy Agency (IAEA) nearly fifty years ago, the International Nuclear Information System (INIS) hosts one of the world's largest collections of published information on the peaceful uses of nuclear science and technology⁷. INIS is produced by the IAEA in partnership with a network of more than 150 member organizations, ranging from university institutes, public research laboratories and organizations to national agencies and corporate R&D centres. In 2016, the INIS collections contain nearly 4m items on nuclear materials, nuclear physics, engineering and instrumentations, nuclear power and safety etc. - mainly books and articles published in scientific and professional journals and magazines, but also non-conventional (grey) literature, i.e. reports and miscellaneous. This literature is research output produced by the extra-academic sector; its cumulative part in INIS is 27% (figure 2), and 70% of these more than 1m items are available from the INIS holdings in Vienna, Austria.

⁴ http://www.opengrey.eu/

⁵ http://www.greylit.org/

⁶ https://northernlight.com/

⁷ https://www.iaea.org/inis/index.html

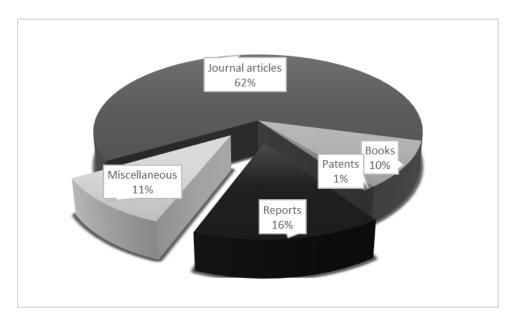


Figure 2: Content of INIS 1970-2016, in % (total 3,993,008 items)

Aeronautics: OpenGrey (see above) contains about 7,000 records in the fields of aeronautics, on aerodynamics, aircraft components and safety, air traffic control, airport operations etc. These documents – many reports and theses but also miscellaneous and communications – have been produced not only by public research organizations like the French ONERA or the German DFVLR and universities like Cranfield or Karlsruhe, but also by Civil Aviation Authorities and corporate R&D, like Rolls Royce or the German MBB – together these two companies represent 9% of all records on aeronautics in OpenGrey, mainly reports (57%), followed by communications (36%). All these items have been collected by the main European STI centres like The British Library, the German TIB or the French INIST.

Economics: The decentralized bibliographic database RePEc⁸ of working papers, journal articles, books, books chapters and software components, provides among many other services a directory (EDIRC)⁹ with 13,698 economics departments, institutes and research centres in 231 countries and territories all around the world. Whenever possible, EDIRC links to the organizations' records in the database. Of course, EDIRC has many economics departments, research centres and institutes in universities, but also finance ministries, statistical offices, central banks, monetary authorities, think tanks, and other non-profit institutions which are clearly extra-academic. It is difficult to estimate their part of the grey literature in economics, but their preferred grey document type is the working paper.

Archaeology: In the words of Börjesson, professional archaeology reports "balance on the boundary between academia and government" (2015, p.139). A large number of practitioners in the US engaged in cultural resource management, i.e. a range of activities tied to national heritage preservation, producing reports on surface surveys and excavation projects. How many? Difficult to say, but probably more than 0.5m (cf. Fagan & Durrani 2016). The archives of the UK Archaeology Data Service at the University of York contain more than 40,000 unpublished fieldwork reports, many of them issued by extra-academic contractors, like consultants, architects, county councils, planning offices, geophysical survey contracting and other companies¹⁰.

.

⁸ http://repec.org/

⁹ https://edirc.repec.org/

¹⁰ http://archaeologydataservice.ac.uk/archives/view/greylit/az.cfm

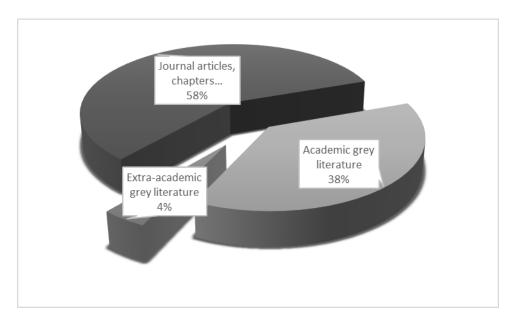


Figure 3: Content of ArchiveSIC, in % (total 3,934 items)

Library and information sciences: It is well known that the research output in library and information sciences (LIS) is in parts produced by information professionals (librarians, archivists etc.) or in "joint ventures" between scientists and professionals. The French open repository ArchiveSIC¹¹ contains nearly 4,000 documents, many journal articles and book chapters but also survey reports, communications, working papers and other studies. This grey literature represents 42% of the whole repository content (figure 3). 4% of this grey literature is extra-academic output, produced by non-academic institutions and authors non-affiliated to universities or public research centres. Another example is the special case of the US and UK library annual reports, with their rich statistical, narrative and visual information. Not only are they an invaluable source for research on library development and professional knowledge since the 19th century, but they also inform about innovative facilities, survey results etc. (Searing 2011). This clearly is extra-academic scientific grey literature.

Challenges

This neither exhaustive nor representative overview illustrates how much extra-academic grey literature is a reality, with more or less impact and importance depending on research domains and holdings. Like other grey literature, extra-academic research output faces four major challenges.

Monitoring: As a matter of fact, extra-academic scientific grey literature is a reality. Any information professional in charge of non-conventional documentation can confirm this. But it is quite difficult if not impossible to give a good and reliable estimate of the number and importance of these items. Often, catalogues, databases or repositories do not provide sufficient information (records, metadata) about producers (suppliers) and document types which makes it impossible to distinguish and analyse this material. Often, too, OPAC and repositories do not allow this kind of research in their records. In the past, this problem was called "lack of bibliographic control".

Standardization: Missing standards for the production and dissemination of reports, conference and working papers has always been an issue for grey literature. But at least, research communities created their own mode of communication and debate, with some more or less standard and formalized features. We cannot observe the same tendency outside of academia; on the contrary,

_

¹¹ https://archivesic.ccsd.cnrs.fr/

each company, each agency, administration or organization tends to create and apply its own rules, different from others, with different formats, metadata, structures etc. A particular problem are unique identifiers — while academia starts adopting DOI, ORCID and institutional identifiers, no comparable development can be observed in extra-academic research. This makes it also improbable that extra-academic research output will be compliant in a foreseeable future with dissemination principles like the FAIR approach (Wilkinson et al. 2016).

Data: While academia and publicly funded research is part and stakeholder of the open data policy of the major European and North American countries, the situation in extra-academic research appears much more unclear and confused. While some public agencies and administrations outside of academia are involved and engaged in open data policy, other organizations (ONGs etc.) and especially the corporate sector pursue quite different and heterogeneous policies regarding data which often have more in common with secrecy and protection of interests than with open data or open science. The same observation applies to the potential reuse of extra-academic research output for text and data mining.

Preservation: The long-term conservation has always been an issue for grey literature. For this reason grey literature has sometimes been described as ephemeral or fugitive. In the context of web technologies and open access, universities and research organizations have started to implement institutional strategies for the preservation and availability of their research output via institutional repositories, data archives etc. For the moment, their success may remain limited. But while Crowe et al (2010) argue in favour of repositories for NGO reports etc., in general we do not observe any significant effort for long-term conservation of extra-academic grey research literature. If this situation does not change, extra-academic research output risks becoming or remaining a kind of "fast food documentation", for fast consumption without any guarantee for sustainable availability.

Quality issues

Sometimes, grey literature is erroneously considered as "un-published" or "not peer-reviewed" scientific documents, inferring less or missing quality control. The reality is different insofar as much if not most of the grey literature is published and disseminated by organizations with some kind of validation, labelling or quality control, e.g. by scientific committees (conferences), juries (theses and dissertations), institutions or peers (reports, working papers). All these procedures guarantee a minimum level of quality, even in the absence of a formalized, full-scale double-blind peer review. But as the debate on open peer review, post-review etc. shows, even peer review is not able to guarantee 100% quality and absence of falsified or fake data, scientific fraud etc. And in the era of gold open access and article processing charges, predatory journal publishing with bad quality has become a critical issue (Beall 2012).

Quality does not depend on the colour of the document. For the user (and the librarian), the crucial question is *not* the type of document but the quality of the information (content), its relevance, reliability and validity. The broad term of grey literature embraces many different situations and items, and it is not as easy for journal articles to make general statements about quality control and quality assurance. But information professionals especially when working in the field of competitive or strategic intelligence know very well how to rate the reliability and credibility of open source information (Crowe & Davidson 2009). The problem with grey resources may be, nevertheless, that necessary data for the rating like author, date of publication, publisher (organization), intended audience etc. may sometimes be more difficult to obtain because of missing or incomplete metadata.

Comparing grey literature with journal publishing, Farace (2011) considered that grey literature requires alternative models for peer review, that "community sourcing lends itself to the review of

grey literature", and that "grey literature focuses more on the review process than the end product" (p.30). Also, he highlights a specific aspect which is equally relevant for extra-academic research output in general, i.e. that commercial and grey publishing share more in common with one another than with self-publishing – in the overwhelming number of grey items, the institutional character is predominant, and this means that quality is not only a characteristics of a single resource but must be seen in the context, as a holistic concept including process as well as institutional variables (Farace). Or in other words, regarding development-led archaeology reports, "although report authors are responsible for the report content, the control of what goes into a report is distributed between regional authorities and report authors" (Börjesson 2016).

Beyond boundaries

Our last issue is open science. In 2016, the Competiveness Council of the European Union defined open science as a new "friendly regulatory environment for research and innovation (which) breaks down the barriers around universities and ensures that society benefits as much as possible from all scientific insights (maximizing) the input of researchers, universities and knowledge institutions"¹². What does maximizing the input of academia mean for scientific information in general and grey literature in particular?

For scientific information in general, this new research policy probably means

- an acceleration of the open access movement: in order to increase the immediate availability
 of research results, governments and scientific authorities will foster and speed up the
 transition to open access, via institutional repositories, academic networks and, above all, via
 "gold" and hybrid journals;
- a broadening of the scope of scientific information: with open science, the focus of scientific information includes corporate companies, start-ups, NGOs etc., unaffiliated knowledge workers (Brown 2016) and people involved in citizen science, i.e. public participation in scientific research conducted at least partly by amateur or nonprofessional scientists (Hand 2010);
- an abandon of usual patterns of scientific research: the "breakdown" of traditional boundaries will contribute to less formalized communities, with less institutionalization, other values, behaviours and traditions; and
- the development of new vectors of communication and dissemination: wiki, blogs and social networks have become, in a few years, key sources to identify trends, debates and experts; some of them have all characteristics of grey documentation, i.e. originality, freshness, textual nature, content richness, interest for intelligence and information products, but also large diversity, ephemeral character and lack or deficit of standardization.

Open science, certainly, will increase the part and impact of extra-academic research; it will also contribute to an environment favourable to new, non-conventional scientific communication. So, are there good times in sight for grey literature? Yes, theoretically. In fact, there are two challenges – at the end, will there still be literature *stricto sensu*? And will librarians and other information professionals still collect and curate this kind of material?

Uncertain future

Insofar as professional knowledge making continues generating information outside of the usual commercial and academic dissemination vectors, extra-academic grey literature will remain part of

¹² Council of the European Union 26-27 May 2016 http://www.consilium.europa.eu/en/meetings/compet/2016/05/26-27/

the scientific and technical information and challenge the information professionals. However, we can identify two major threats for the development and future of professional grey literature, i.e. transformation and disintermediation.

Transformation: Because of the diversification and enrichment of contents and functionalities, the boundaries of grey literature become blurred. Traditional grey literature like laboratory notebooks or conference papers have a strong transgressive potential, and it is becoming increasingly difficult to distinguish documents and digital tools or services. Scientific and technical reports, just like theses and dissertations, are research output but also and moreover, "containers" of data and metadata on projects, organizations, experts, resources, funding etc. This evolution challenges the nature of documents and literature, but not the informative value of these items for evaluation and intelligence.

Disintermediation: Per definition, grey literature depends on library intermediation. We cannot think of grey literature without libraries, information centres or intelligence. The concept has been invented by professionals in charge of acquisition of "special" documents for customers and patrons; it is a professional business. Thus, the major challenge of grey literature is the challenge of libraries. Some people think that grey literature is at risk because of the Internet and increased availability of documents. This analysis is wrong – some documents may be more available than before, but many others remain difficult to identify, find, and access. No, the real threat of grey literature is the weakening intermediation function of libraries and information centres. In some way, grey literature is like a canary in coal mines – its existence and wellbeing witness the intact function of intermediation. The moment information professionals, for which reason whatsoever, stop search, acquisition and dissemination of this kind of documents, grey literature will disappear. The moment information professionals, like their customers and patrons, start to consider Google, Facebook and Amazon as sufficient for information needs, the concept of grey literature stops making sense.

Acknowledgments to Hélène Prost, Taryn L. Rucinski and Dobrica Savic for helpful information and advice.

Suggestions for further research

Further research could be conducted in the field of grey literature, especially in archaeology and extra-academic research, and on new forms of scientific communication in the environment of open science and citizen science. Some references for further reading:

Börjesson, L. (2016). Beyond information policy: conflicting documentation ideals in extra-academic knowledge making practices. *Journal of Documentation*, 72(4):674-695.

Brooks, L. and Fitz, G. (2015). Grey matter(s): Embracing the publisher within. *The Foundation Review*, 7(2):38-50.

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

Groom, Q., Weatherdon, L. and Geijzendorffer, I. R. (2016). Is citizen science an open science in the case of biodiversity observations? *Journal of Applied Ecology*.

Harlan, M. (2010). Black and white literature, grey scholars. Archaeologies, 6(2):270-288.

Seymour, D. J., editor (2009). Archaeology and grey literature (special issue). *The Grey Journal*, 5(2):63-110.

Urquhart, C. and Dunn, S. (2013). A bibliometric approach demonstrates the impact of a social care data set on research and policy. *Health Information & Libraries Journal*, *30*(4):294-302.

References

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

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, 489(7415):179.

Börjesson, L. (2015). An attempt to nuance the understanding of professional reports in archaeology. *The Grey Journal*, 11(3):137-144.

Börjesson, L. (2016a). Research outside academia? - An analysis of resources in extra-academic report writing. In *ASIST 2016, October 14-18, 2016, Copenhagen, Denmark*.

Brown, D. J. (2016). *Access to scientific research. Challenges facing communications in STM*. De Gruyter Saur, Munich.

Choi, K., Chae, C.-J., Yae, Y.-H. and Shin, Y. J. (2015). Analysis of collection and management of the Korea national R&D report. *The Grey Journal*, 11(2):115-120.

Crowe, J. and Davidson, T. S. (2009). The 'grey' intersection of open source information and intelligence. *The Grey Journal*, 5(3):123-133.

Crowe, J., Hodge, G. and Redmond, D. (2010). Grey literature repositories: Tools for NGOs involved in public health activities in developing countries. In Farace, D. and Schöpfel, J., editors, *Grey Literature in Library and Information Studies*, pages 199-214. De Gruyter Saur, Munich.

Fagan, B. M. and Durrani, N. (2016). Archaeology: a brief introduction. Routledge, New York.

Farace, D. (2011). Peering through the review process: Towards transparency in grey literature. *The Grey Journal*, 7(1):25-31.

Hand, E. (2010). Citizen science: People power. *Nature*, 466(7307):685-687.

Rucinski, T. L. (2015). The elephant in the room: Toward a definition of grey legal literature. Law Library Journal, 107(4):543-559.

Rudasill, L. M. (2009). Grey literature and development: the non-governmental organization in action. *The Grey Journal*, 5(3):117-121.

Schöpfel, J. (2010). Towards a Prague definition of grey literature. In 12th International Conference on Grey Literature. Prague, 6-7 December 2010.

Searing, S. E. (2011). Grey literature past and present: the evolution of library annual reports. *The Grey Journal*, 7(3):147-152.

Wilkinson, M. D., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018+.