

Are electronic theses and dissertations (still) grey literature in a digital age? A FAIR debate

Joachim Schöpfel, Behrooz Rasuli

► To cite this version:

Joachim Schöpfel, Behrooz Rasuli. Are electronic theses and dissertations (still) grey literature in a digital age? A FAIR debate. *Electronic Library, Emerald*, 2018, 36 (2), pp.208-219. 10.1108/EL-02-2017-0039 . hal-01701276

HAL Id: hal-01701276

<https://hal.univ-lille3.fr/hal-01701276>

Submitted on 5 Oct 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.

Are electronic theses and dissertations (still) grey literature in the digital age? A FAIR debate

Joachim Schopfel
GERiiCO, University of Lille, Villeneuve d'Ascq, France, and
Behrooz Rasuli
*Department of Scientometrics and Data Analysis,
Iranian Research Institute for Information Science and Technology (IranDoc),
Tehran, Iran*

Abstract

Purpose – While distribution channels of theses and dissertations have changed significantly in the digital age, they are generally still considered grey literature. This paper aims to argue the applicability of the concept of grey to electronic theses and dissertations (ETDs).

Design/methodology/approach – The paper is presented as a debate between two contradictory opinions on the application of the grey literature concept to ETDs.

Findings – The paper provides a definition of grey literature and then discusses its application to electronic dissertations and theses. In particular, it assesses the aspects of acquisition, quality, access and preservation. Some arguments highlight the “grey nature” of ETDs, such as the limited access via institutional and other repositories. Other arguments (e.g. the development of ETD infrastructures and the quality of ETDs) question this grey approach to ETDs. The paper concludes that “greyness” remains a challenge for ETDs, a problem waiting for solution on the way to open science through the application of the FAIR (findability, accessibility, interoperability reusability) principles.

Research limitations implications – Library and information science (LIS) professionals and scientists should be careful about using the concept of grey literature. The debate will help academic librarians and LIS researchers to better understand the nature of grey literature and its coverage, here in the field of ETDs.

Originality/value – Some definitions from the print age may not be applicable to the digital age. The contradictory character of the debate helps clarify the similitudes and differences of grey literature and ETDs and highlights the challenge of ETDs, in particular, their accessibility and findability.

Keywords Open access, Academic publishing, Grey literature, Electronic theses and dissertations, FAIR principles

Paper type Viewpoint

Introduction

The debate started on the campus of the University of Lille, during the nineteenth International Symposium on Electronic Theses and Dissertations in July 2016. Are electronic theses and dissertations (ETDs) still grey literature? Yes, they are still grey, said one of the co-authors, at least in part, because they are difficult to identify, to preserve and to access. No, replied the other co-author, as, in the digital age, PhD theses are no longer “grey”,

The authors would like to acknowledge Gail McMillan from Virginia Tech for helpful comments and advice.



because they can be made freely available through institutional repositories and ETD infrastructures. The time was too short to close the debate, and so the word was given to continue the arguments in a public forum but in a written form. The authors feel that not only can this debate contribute to a better understanding of the term *grey literature* but also it can shed new light on some significant developments in the field of academic librarianship and publishing. Furthermore, this is an attempt to study the effects of the digital age on grey literature's definition.

About grey literature

Grey literature is a concept born in the domain of library and information science (LIS) in the second half of the twentieth century – a modern concept, therefore, but still marked by the Gutenberg era and its large print collections (Schöpfel and Farace, 2010). It was invented by acquisition librarians looking for specific categories of documents difficult to get. At the beginning, the term covered principally reports (Chillag, 1993) and meeting papers from different fields, such as aeronautics, engineering, defence, economics, atomic energy, or agriculture and produced by governments, research laboratories, and business or as trade literature (Auger, 1989). These documents were not classified or protected (“black literature”) but were open-source. Often, the information professional knew that these items existed and that they had been disseminated. Yet, the problem was how to get them, especially when they were published in small numbers or in foreign countries. Other than books or journals (“white literature”), these documents could not be acquired through the usual market channels. A specific knowledge of networks, information sources and dissemination vectors was needed. Grey literature was a challenge for information professionals.

At the beginning, the “hot topics” of grey literature were special library acquisitions, material, microfilms and microfiches, document supply, bibliographic control, standards and organizations with a significant output of scientific and technical reports, such as NASA and the US National Technical Information Service. A specific aspect was that a good part of these documents was produced outside of academia, by government agencies, corporate companies, international structures, non-governmental organizations, etc. Dissertations, unlike reports and meeting papers, were not part of the initial concept of grey literature. But with the extension of the concept and the development of acquisition policies and infrastructures, including the System for Information on Grey Literature in Europe (SIGLE) (Wood and Smith, 1993), a networked database for European scientific and technical grey literature, theses and dissertations became a central element of the concept. Therefore, as a result of this larger approach to grey literature, 53 per cent of the resources of the former SIGLE database were dissertations produced by 20 different countries (Juznic, 2010).

In systematic reviews and library guidelines, grey literature is often defined as unpublished, that is, not available via traditional publishing, unconventional, with little distribution and not peer-reviewed. The US Interagency Gray Literature Working Group, in its *Gray Information Functional Plan* of 1995, described grey literature as:

[...] foreign or domestic open source material that usually is available through specialized channels and may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents (www.dtic.mil/dtic/tr/fulltext/u2/b300928.pdf).

The International Network of Grey Literature (GreyNet, see www.greynet.org) defines grey literature as “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". This "Luxembourg definition", approved at the Third International Conference on Grey Literature in 1997, was extended seven years later at the Sixth Conference in New York City, where a postscript was added: grey literature is "[...] not controlled by commercial publishers, i.e. where publishing is not the primary activity of the producing body" (Schöpfel and Farace, 2010). This definition has since been used extensively and is widely accepted.

Grey literature is 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 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 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.

Because this definition is inclusive and based on the former GreyNet concepts, it will be applied to the following analysis.

The question

A discussion about grey literature is necessarily a discussion about different topics in LIS, in particular, about acquisition policy and special collections, and also about documents and document typologies, taxonomies, publications and dissemination, business models, the information market, availability, preservation and mediation. GreyNet and its conference series were, and still are, the venue of several discussions on larger or narrower interpretations of grey literature and between typology-based and mediation-based approaches.

Usually, especially in systematic reviews in medical and life sciences, grey literature is considered as "unpublished material" that has not been peer-reviewed. This term normally covers reports, conference papers, dissertations and working papers, but some studies define the term as much larger, including several types of information resources, such as websites, maps, software, datasets, exam topics and leaflets (Pejšová *et al.*, 2011). ETDs are part of the GreyNet coverage, and the GreyNet international conference series received several papers on theses and dissertations, raising awareness not only on the transition from print to digital formats, special collections and holdings and their availability and deposit in open repositories but also on persistent problems with bibliographic control, preservation and access (see the conference papers at www.opengrey.eu).

On the other hand, the main international organization dedicated to promoting the adoption, creation, use, dissemination and preservation of ETDs, the Networked Digital Library of Theses and Dissertations (NDLTD), generally does not use the term grey literature for ETDs (McMillan, 1999). The reason may be that their focus is on formats, workflows, dissemination and access, not on acquisition. Their objective is to achieve the digital transformation of theses and dissertations and to foster their findability and availability on the Web. The NDLTD Global ETD Search portal contains metadata on 4.5

million items from over 150 sources worldwide (catalogues, repositories, portals, etc.), harvested by the NDLTD Union Archive. The DART-Europe E-Theses Portal gives access to 715,208 open-access research theses from 601 universities in 28 European countries. In this new world of digital repositories and open access, is there still a place for grey literature?

So far, the “grey community” (Prost and Schöpfel, 2014) never really questioned the fundamentally “grey character” of ETDs. However, the landscape of ETDs is more complex and cannot be reduced to one colour – a small but significant part of dissertations are confidential, hidden and/or embargoed in academic libraries and ETD infrastructures (“black”), while other dissertations are published as printed or e-books by academic publishers (“white”). Moreover, Web-based technologies and infrastructures, in particular, open repositories and social networks, bear the potential of gratis and libre/free access to theses and dissertations (Schöpfel, 2013). Does this mean that dissertations have become less grey or that the distinction between different types of literature – that is, white and grey – has become obsolete (Artus, 2003)? What about mediation of ETDs, about their acquisition, collection, preservation and dissemination; are they no longer a challenge for academic librarians? Are ETDs (still) grey literature?

Yes, they are (still) grey literature (pros)

Applying the six criteria of the Prague definition (Schöpfel, 2010), the answer to this question is affirmative: yes, ETDs are still grey literature, for several reasons. In detail:

“Grey literature stands for manifold document types [...]”. There is no doubt about the document character of dissertations, in print and digital formats. However, the digital format allows links to research results and to data sets in the dissertation itself and, moreover, to exploit the dissertation with content mining tools as if it were itself a data file (Schöpfel *et al.*, 2015a). Yet, up to now, prescriptive rules and academic habits keep intact and preserve the document character of dissertations.

“[...] produced on all levels of government, academics, business and industry [...]”; ETDs are considered as genuine scientific output of universities, via graduate schools, academic departments and/or research laboratories (Larivière, 2012). As the metadata in the NDLTD and DART portals show, one part of them is “co-produced” by corporate R&D, hospitals, or public administrations.

“[...] in print and electronic formats [...]”; after centuries of print theses and dissertations, today, many universities accept digital deposits or have made them mandatory (Lippincott and Lynch, 2008; Reeves *et al.*, 2006). However, one can observe very different solutions for the processing of ETDs, regarding metadata, formats, workflows, system architecture, etc. As the papers presented at the ETD conferences show, divergence is the rule, not interoperability or standards.

“[...] that are protected by intellectual property rights [...]”; generally, dissertations are the result of years of individual research and, legally speaking, original creation of the mind, that is, they are an intellectual (literary) work put into a readable format and, as such, protected by copyright laws. Yet, copyright is not the only legal regime that applies to dissertations (Schöpfel and Lipinski, 2012). In some countries, they are also regulated by administrative laws (as a document required for a state diploma). Also, in the context of open data and open science, there is a growing debate on the limits of the individual author’s intellectual property, in particular, when ETDs are the result of publicly funded research (Harper, 2011, Hawkins *et al.*, 2013). Other limitations of the protection by intellectual property rights arise when the underlying research has been conducted by and with industrial R&D funds (creating rights via corporate innovation and interests) and when the

ETDs contain data sets with different and genuine legal regimes and challenges (*sui generis* database rights, privacy, etc.).

“[...] of sufficient quality to be collected and preserved by library holdings or institutional repositories [...]”; in a general way, the process of doctoral studies guarantees a minimum level of quality, via the supervisory arrangements and individual follow-up, the formal, institutional and/or legal requirements for the dissertation and the final, oral examination by a committee or jury, before or after completion of the submission (*viva voce*, “defense”) (Juznic, 2010). On the other hand, supervisory arrangements, requirements and examinations can be very different between universities and countries, even inside the same university; also, the quality of a PhD dissertation depends largely on the excellence and reputation of the institution, on the quality of the supervision and on the candidate’s research (and writing) skills. The best process cannot completely prevent scientific misconduct; fabrication of data, falsification of research results and plagiarism are real problems calling for institutional awareness and measures. For example, an institutional repository should implement a specific ETD workflow that makes it impossible to deposit faked, non-validated or otherwise fraudulent documents (Ferrerias-Fernandez *et al.*, 2015; Noge and Duskova, 2013). As a final observation, dissertations are present in all academic libraries. However, their status is somehow different from journals or books insofar, as collections of print or digital dissertations often have institutional or mandatory character – that is, their acquisition policy is generally not selective but exhaustive (“all dissertations from a given institution, country, discipline, field, . . .”) – and the quality assurance is not a matter for the information professional (acquisition librarian, etc.), but is expected from the institution that delivers the doctoral degree and disseminates the dissertation.

“[...] but not controlled by commercial publishers i.e. where publishing is not the primary activity of the producing body”; at first sight, the situation is simple and self-evident – doctoral degrees being delivered by academic institutions, theses and dissertations are considered (and evaluated) as part of their scientific output – in other words, the “producing bodies” are mostly universities which may “externalize” the dissemination of “their” theses and dissertations to academic networks or agencies, such as in India or France. However, one part of the ETDs is (also) available via corporate vendors and academic publishers, which may, in some cases, also control the ETD metadata.

In summary, ETDs are generally compliant with the criteria of the Prague definition of grey literature. In other words, yes, ETDs are (still) part of grey literature, even if single cases may be different; for example, when a dissertation is classified or when it is disseminated through commercial channels. A paradox development even reinforces the greyness of ETDs. The creation of institutional repositories and ETD workflows does not make all items more accessible and available, and a significant part of ETDs remains embargoed and/or limited to on-campus access (Owen *et al.*, 2009; Schöpfel and Prost, 2014). Non-English ETDs, especially those written in vernacular languages, are part of the problem. As long as this unsatisfying and inefficient situation persists, as long as their findability and accessibility remain limited, ETDs are still grey literature – not because of their lack of peer review or “uncertain quality” (a false argument, as shown above), but because their identification, collection and use continue to be a challenge.

No, they are not (no longer) grey literature (cons)

As mentioned above, grey literature was originally a term for describing *documents* characterized by the *difficulty* of the *librarian* to get them, a term that emerged from the library environment and acquisition context. But with time, “greyness” went beyond library doors and became a widespread term in other contexts, such as universities, laboratories

and research institutes. While acquisition staff did not intend to assess the quality of information resources with this label, outside the library, grey literature became increasingly associated with unsure or deficient quality (Jeffery and Asserson, 2011, 2014; Motta *et al.*, 2016).

One of the key consequences of applying the label *grey* to large parts of academic literature is that academic communities have focussed their research on white literature, while grey literature has been pushed to the margins. The steadily declining impact of dissertations from 1980 onwards (Larivière *et al.*, 2008) may be seen as a result of this labelling. Moreover, this link between grey literature and (lack of) quality may have motivated scientists and institutions to prefer white literature as a vector for research output to grey items, which, in turn, further contributed to decreasing quality, removing them from the research ecosystem, ignoring their impact, etc.

Insisting on applying the label of *grey* to many different types of documents proves that the print world's legacy still dominates various concepts in the digital age, as well as keeps these concepts from being defined in a different way. We have entered a new era with old rules and approaches from the print age, whereas the new age has its own requirements. The digital age has challenged old and obsolete definitions; from the concept of learning (Sharpe *et al.*, 2010) and literacy (Tyner, 2014) to plagiarism (Evering and Moorman, 2012). Hence, should we not rethink grey literature in the digital age instead of continuing to use the term as if nothing has changed?

The world of information management has changed, and so have the dissertations and their formats, platforms and dissemination vectors. Does the term grey literature really cope with this new situation? Let us start with a flashback to the definition of “grey literature”. As noted above, the dominant concept in the early definitions is difficult-to-get. We can interpret two sides of difficult-to-get. The first point is that scholars cannot access a certain resource, usually because of physical distances. The second point is that there was no institution to collect and manage all the resources so that acquisition staff could subscribe to some services to access resources.

New technologies have revolutionized (scholarly) publishing and built an infrastructure for information sharing in a more effective way (Borgman, 2010) and “new formats and contents are challenging research communities and the information industry [. . .] Academic publishing has definitively left the Gutenberg era” (Schöpfel *et al.*, 2014a, p. 612). Electronic publishing was one of the consequences of this revolution, so ETDs are now produced, published, distributed and retrieved digitally. As one of the vital promises of the digital age is access to digital resources anywhere anytime, digital resources, such as ETDs (at least in theory if not in practice), are, not difficult-to-get anymore.

Also, following Learned Publishing's editor-in-chief Pippa Smart (2015), the main issues of grey literature in actual academic publishing are inaccurate citing, lack of archiving for posterity and, sometimes, quality. However, regarding ETDs, these are secondary problems, because the situation has significantly improved for 15 years now, through standards, infrastructures and institutional control.

Today, integrated information systems, such as institutional repositories, ETD databases, digital libraries and current research information systems (Schöpfel *et al.*, 2014b), collect ETDs of one or more institutions in a single database and facilitate access to ETDs. Even if institutional repositories do not necessarily make all items more accessible (Schöpfel and Prost, 2014), they have the potential to do so. Several databases, portals and other discovery tools have been developed to manage ETDs at different levels (institutional, local, national, regional and global). For example, several different countries have developed their own national ETD databases or gateway, such as *EThOS* in the UK, *Theses Canada* in

Canada, *Digital Australian Theses* via *Trove* in Australia, *Theses.fr* in France, *BDLTD* in Brazil, *IranDoc ETDs* in Iran, *National Thesis Center* in Turkey, *NDLTD* in Taiwan, *ETD Portal* in South Africa, *CALIS ETD* in China, *Digital Dissertation Library of the Russian State Library* in Russia, *doiSerbiaPhD* in Serbia, *eLABa ETD* in Lithuania, *Shodhganga* in India, *TDX* in Spain and *NARCIS* in The Netherlands. In addition, other regional (such as *DiVA* for Scandinavian institutions and *DART* for European countries) and global services (*NDLTD*, *WorldCat*, *OAIster* and *Cybertesis*) contribute to increase in the findability and accessibility of ETDs.

Another way to increase the findability of ETDs is the minting of unique identifiers, in particular, the digital object identifier (DOI), which helps producers and users of information resources to organize and locate intellectual objects in the digital environment (Chandrakar, 2006). Some national ETD programmes are investigating how DOIs could be allocated to their records (Schöpfel *et al.*, 2014b), such as EThOS in the UK (Gould, 2016). On the other hand, for more than 10 years now, DOI registration agencies have been attempting to include new items, for example, *CrossRef* has been investigating “the addition of new types of scholarly content – theses and dissertations, patents, working papers, technical reports and a whole range of grey literature” (Pentz, 2004, p. 185). DOI helps to organize and retrieve ETDs in an effective way, as well as it is useful in citation indices.

Another element of the definition of grey literature is that these documents are “not controlled by commercial publishers”. Apparently, this concept was also challenged by the digital age. As noted above, academic institutions are the original producers of ETDs. However, there are differences between producer and publisher, technically (Li *et al.*, 2016). While producers generate content and information, publishers publish information and content through various types (books, articles, databases, etc.). Therefore, publishing ETDs may be controlled by an independent publisher, not necessarily by the academic institution itself. For example, some corporate companies (such as ProQuest) currently control ETD publishing for some institutions around the world through the *PQDT* database, and there are some other commercial vendors working at the national level; for example, in Italy and Spain, some companies collect and sell dissertations (Rasuli *et al.*, 2015).

As noted above, while the primary definitions of grey literature did not mention or tackle the quality of the documents, scientists and librarians associate grey with low quality. For example, in the review studies, researchers are hesitant about including grey literature in their analysis as they are not sure about the quality of such literature (Adams *et al.*, 2016). Yet, ETDs “contain the results of at least three years of scientific work, accomplished within a laboratory, a research team or an institute, school or company” (Schöpfel *et al.*, 2014a, p. 616); also “it is reasonable to assume that high-quality work is published outside the white literature by individuals who are not under pressure to publish in academic journals” (Adams *et al.*, 2016).

Moreover, as ETDs are reviewed by academic committees, we can assume that they have enough acceptable quality to impact future research, especially at the PhD level (Larivière *et al.*, 2008). As a matter of fact, researchers are regularly citing ETDs in their academic writings. If we consider citation as a credit which reflects the quality of a document, then we can say that ETDs have academic quality even if there may be differences in citing ETDs and the number of citations, depending on disciplines and publication types and also on the data sources and indexing tools (Meho and Yang, 2007). Also, many white literature items, such as journal articles and books, are derived from ETDs, which indicates their quality.

In summary, some parts of the grey literature definition (for example, “that are protected by intellectual property rights” or “in print and electronic formats”) are generic and are applicable to all types of information resources, such as grey, white or black literature. But the core concept of the definition of grey literature (e.g. “difficult-to-get”) is challenged by the new technological advancements of the digital age. This is especially true regarding ETDs that are collected, organized, distributed and retrieved more systematically than other “grey” items. With the transition from print to digital formats, dissertations and theses started to move out of the field of grey literature, and it may not be really appropriate to continue calling ETDs grey literature as before, as if nothing has changed during the past 20 years.

For several decades now, the ecosystem of formal academic communication is dominated by white literature, while other colours (i.e. grey and black) play a more marginal role. Labelling documents with valuable data and information as grey can decrease their potential impact on future research. Therefore, the concept of grey literature should be used with caution in the field of ETDs. In fact, this label may be or become a problem for the impact and promotion of ETDs. As information culture is a critical success factor in developing ETDs databases (Rasuli *et al.*, 2016), changing the way of seeing and labelling ETDs – that is, stop calling them grey – can help improve their accessibility. Removing the label may result in more visibility and impact of ETDs and perhaps convince scientometric databases, such as Web of Science and Scopus, to develop an index for ETDs.

Conclusion

In the past (Gutenberg era), PhD dissertations were generally considered as grey literature, mainly because of their non-commercial production and dissemination, because of their interest in academic special collections and because of specific problems with identification and acquisition. For some 20 years, new technology – in particular, the Web – has completely modified the environment, infrastructures, tools and formats of these documents in such a way that it seems legitimate to ask if it (still) makes sense to include ETDs in the field of grey literature.

Our debate presented some strong arguments that one can still apply the concept of grey literature to ETDs, for the same reasons as before. One main argument is that they are still collected by academic libraries because of their quality and interest for research and development, and that this acquisition is organized outside the usual (commercial) book channels. Arguments on the other side question the interest of the “grey label” not only because of the significant improvement of digital ETD infrastructures and discovery tools but also because this label, with its negative connotation regarding quality, may be a risk for the status and impact of ETDs.

Grey? Not grey? Confronting the argument’s pros and cons, there is in fact one shared assessment or conviction, that is, the “greyness” is an artificial problem more than a “natural” feature. While for other documents (e.g. reports, working papers or communications), greyness can be considered a “normal attribute” largely accepted by authors, institutions and librarians as a usual property of this specific type of information resource, the same label represents a challenge for ETDs and a barrier to remove. Perhaps, this is the main difference between ETDs and other categories of grey literature.

So, we can conclude that here, in the specific environment of ETDs, greyness as defined above should be considered as a transitional phenomenon, as one step or stage in the transition to open science. Grey literature may be a helpful concept to describe and analyze

some specific problems related to the publishing and intermediation of ETDs. Above all, it is an expression or indicator of these problems. Resolving the crucial problems with ETDs will reduce and remove their greyness.

To improve the dissemination and impact of ETDs, institutions and authorities must tackle the “grey issues”. In the context of open science, this approach can be described with the FAIR (findability, accessibility, interoperability, reusability) principles developed for research data management (Wilkinson *et al.*, 2016). In other words, any ETD policy should set four goals, with *ad minima* the following objectives:

- *Findability*: It is the assignment of a globally unique and persistent identifier (DOI by default), description with rich metadata and registration or indexing of these metadata in a searchable resource.
- *Accessibility*: It is the retrieval by identifier by using a standardized communications protocol, which is open, free and universally implementable and which allows for an authentication and authorization procedure, where necessary.
- *Interoperability*: It is the use of a formal, accessible, shared and broadly applicable language for knowledge representation of metadata and ETDs.
- *Reusability*: It is a rich description with a plurality of accurate and relevant attributes, which are released with a clear and accessible usage license, are associated with detailed provenance and meet domain-relevant community standards.

As the papers and discussions during the ETD conferences show, the transition to ETD infrastructures compliant with the requirements of open science is well engaged. As mentioned above, there is often no need for new facilities and infrastructures. Moreover, in this specific field of ETDs, there are largely accepted standards and good practices which (have the potential to) improve their findability and interoperability. Today, the problem with ETDs lies upstream, in local (academic) contexts and also in national legislation and jurisprudence that facilitate decisions in favour of embargoes and restricted access and which reduce their accessibility and reusability, for instance, for text and data mining. “To put it in a simple way, pipes exist, but there is a lack of both fuel and pressure for ETDs and open access” (Schöpfel *et al.*, 2015b). In other words and to conclude our debate, if by 2020, ETDs should be completely integrated in the emerging open science infrastructures, as open as possible (and just as closed as necessary), easily retrievable and accessible and largely reusable by content mining tools, greyness would no longer be a problem.

References

- Adams, R.J., Smart, P. and Huff, A.S. (2016), “Shades of grey: guidelines for working with the grey literature in systematic reviews for management and organizational studies”, *International Journal of Management Reviews*, Vol. 19 No. 4, pp. 432-454.
- Artus, H.M. (2003), “Old in new bottles? Developments in electronic information and communication: structural change and functional inertia”, paper presented at the 5th International Conference on Grey Literature (GL5), 4-5 December, Amsterdam.
- Auger, C.P. (1989), *Information Sources in Grey Literature*, Bowker-Saur, London.
- Borgman, C.L. (2010), *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*, MIT Press, Cambridge, MA.
- Chandrakar, R. (2006), “Digital object identifier system: an overview”, *The Electronic Library*, Vol. 24 No. 4, pp. 445-452.

-
- Chillag, J.P. (1993), "From Weimar to Maastricht and beyond: half a century with grey literature", paper presented at the 1st International Conference on Grey Literature (GL1), 13-15 December, Amsterdam.
- Evering, L.C. and Moorman, G. (2012), "Rethinking plagiarism in the digital age", *Journal of Adolescent & Adult Literacy*, Vol. 56 No. 1, pp. 35-44.
- Ferreras-Fernandez, T., Garcia-Penalvo, F.J. and Merlo-Vega, J.A. (2015), "Open access repositories as channel of publication scientific grey literature", paper presented at Technological Ecosystems for Enhancing Multiculturality (TEEM'15), 7-9 October, Porto.
- Gould, S. (2016), "UK theses and the british library EThOS service: from supply on demand to repository linking", *Interlending & Document Supply*, Vol. 44 No. 1, pp. 7-13.
- Harper, G. (2011), "ETDs, open access and intellectual property issues", paper presented at USETDA 2011: The Magic of ETDs. . . Where Creative Minds Meet, 18-20 May, Orlando.
- Hawkins, A.R., Kimball, M.A. and Ives, M. (2013), "Mandatory open access publishing for electronic theses and dissertations: Ethics and enthusiasm", *The Journal of Academic Librarianship*, Vol. 39 No. 1, pp. 32-60.
- Jeffery, K.G. and Asserson, A. (2011), "GL transparency: through a glass, clearly", *The Grey Journal*, Vol. 7 No. 2, pp. 99-104.
- Jeffery, K.G. and Asserson, A. (2014), "Data intensive science: shades of grey", paper presented at the 12th International Conference on Current Research Information Systems (CRIS 2014), 13-15 May, Rome.
- Juznic, P. (2010), "Grey literature produced and published by universities: a case for ETDs", in Farace, D.J. and Schöpfel, J. (Eds), *Grey Literature in Library and Information Studies*, De Gruyter Saur, Munich, pp. 39-51.
- Larivière, V. (2012), "On the shoulders of students? The contribution of PhD students to the advancement of knowledge", *Scientometrics*, Vol. 90 No. 2, pp. 463-481.
- Larivière, V., Zuccala, A. and Archambault, E. (2008), "The declining scientific impact of theses: implications for electronic thesis and dissertation repositories and graduate studies", *Scientometrics*, Vol. 74 No. 1, pp. 109-121.
- Li, H.C., Ajoux, P.V., Demailly, L., Guz, L.A., Song, Y.J., Roy, S. and Kumar, S. (2016), "Subscription groups in publish-subscribe system", United States Patent No. US 9344395 B2.
- Lippincott, J.K. and Lynch, C. (2008), "Survey of US ETD programs: progress and state of acceptance", paper presented at the 11th International Symposium on Electronic Theses and Dissertations (ETD 2008), 4-7 June, Aberdeen.
- McMillan, G. (1999), "Perspectives on electronic theses and dissertations", paper presented at the 4th International Conference on Grey Literature (GL4), 4-5 October, Washington, DC.
- Meho, L.I. and Yang, K. (2007), "Impact of data sources on citation counts and rankings of LIS faculty: web of science versus Scopus and Google Scholar", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 13, pp. 2105-2125.
- Motta, G., Puccinelli, R., Reggiani, L. and Saccone, M. (2016), "Extracting value from grey literature: processes and technologies for aggregating and analyzing the hidden 'big data' treasure of organizations", *The Grey Journal*, Vol. 12 No. 1, pp. 15-25.
- Noge, J. and Duskova, M. (2013), "Central registry of theses and dissertation and the anti-plagiarism system as a comprehensive solution at national level", paper presented at the Seminar on Providing Access to Grey Literature, 23 October, Prague.
- Owen, T.M., Hackman, T. and Harrod, T. (2009), "ETDs in lock-down: trends, analyses and faculty perspectives on ETD embargoes", paper presented at the 12th International Symposium on Electronic Theses and Dissertations (ETD 2009), 10-13 June, Pittsburgh.
- Pejšová, P., Mynarz, J. and Simandlová, T. (2011), "A linked data vocabulary of the types of grey literature: version 1.0", paper presented at the 13th International Conference on Grey Literature (GL13), 5-6 December, Washington, DC.

- Pentz, E. (2004), "Recent developments at CrossRef", *Interlending & Document Supply*, Vol. 32 No. 3, pp. 183-185.
- 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. [Patternmatch.
- Rasuli, B., Alipour-Hafezi, M. and Solaimani, S. (2015), "Understanding electronic theses and dissertations through a business model perspective: the case of Irandoc ETDs", paper presented at the 18th International Symposium on Electronic Theses and Dissertations (ETD 2015), 4-6 November, New Delhi.
- Rasuli, B., Alipour-Hafezi, M. and Solaimani, S. (2016), "The identification of critical success factors in the development of national ETDs programs", paper presented at the 19th International Symposium on Electronic Theses and Dissertations (ETD 2016), 11-13 July, Lille.
- Reeves, S., Hagen, J. and Jewell, C. (2006), "Unlocking scholarly access: ETDs, institutional repositories and creators: highlights of ETD 2006, the 9th International Symposium on Electronic Theses and Dissertations", *Library Hi Tech News*, Vol. 23 No. 7, pp. 12-15.
- Schöpfel, J. (2010), "Towards a Prague definition of grey literature", paper presented at the 12th International Conference on Grey Literature (GL12), 6-7 December, Prague.
- Schöpfel, J. (2013), "Adding value to electronic theses and dissertations in institutional repositories", *D-Lib Magazine*, Vol. 19 Nos 3/4.
- Schöpfel, J., Chaudiron, S., Jacquemin, B., Prost, H., Severo, M. and Thiault, F. (2014a), "Open access to research data in electronic theses and dissertations: an overview", *Library Hi Tech*, Vol. 32 No. 4, pp. 612-627.
- Schöpfel, J. and Farace, D.J. (2010), "Grey literature", in Bates, M.J. and Maack, M.N. (Eds), *Encyclopedia of Library and Information Sciences*, 3rd ed., CRC Press, London, pp. 2029-2039.
- Schöpfel, J. and Lipinski, T.A. (2012), "Legal aspects of grey literature", *The Grey Journal*, Vol. 8 No. 3, pp. 137-153.
- Schöpfel, J. and Prost, H. (2014), "Back to grey: disclosure and concealment of electronic theses and dissertations", *The Grey Audit: A Field Assessment in Grey Literature (Fifteenth International Conference on Grey Literature)*, Bratislava, 2-3 December.
- Schöpfel, J., Juznic, P., Prost, H., Malleret, C., Cesarek, A. and Koler-Povh, T. (2015a), "Dissertations and data (keynote address)", paper presented at the 17th International Conference on Grey Literature (GL17), 1-2 December, Amsterdam.
- Schöpfel, J., Prost, H., Piotrowski, M., Hilf, E.R., Severiens, T. and Grabbe, P. (2015b), "A french-german survey of electronic theses and dissertations: access and restrictions", *D-Lib Magazine*, Vol. 21 Nos 3/4.
- Schöpfel, J., Zendulkova, D. and Fatemi, O. (2014b), "Electronic theses and dissertations in CRIS", *Procedia Computer Science*, Vol. 33, pp. 110-117.
- Sharpe, R., Beetham, H. and De Freitas, S. (2010), *Rethinking Learning for a Digital Age: How Learners Are Shaping Their Own Experiences*, Routledge, New York, NY.
- Smart, P. (2015), "Twenty-five shades of grey", *Learned Publishing*, Vol. 28 No. 3, pp. 163-165.
- Tyner, K. (2014), *Literacy in a Digital World: Teaching and Learning in the Age of Information*, Routledge, New York, NY.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J. (2016), "The FAIR guiding principles for scientific data management and stewardship", *Scientific Data*, Vol. 3, p. 160018.
- Wood, D.N. and Smith, A.W. (1993), "SIGLE: a model for international co-operation", *Interlending & Document Supply*, Vol. 21 No. 1, pp. 18-22.

About the authors

Joachim Schopfel is a Senior Lecturer in information and document sciences at the University of Lille and Researcher at the GERiiCO laboratory. He was also the Director of the French national reproduction centre from 2012 to 2017 for PhD theses. He is interested in scientific information, academic publishing, open repositories, grey literature and usage statistics. He is a member of the Grey Literature Network Service (GreyNet) and one of the directors of the Networked Digital Library of Theses and Dissertations (NDLTD). Joachim Schopfel is the corresponding author and can be contacted at: joachim.schopfel@univ-lille3.fr

Behrooz Rasuli is a PhD Candidate in library and information science at IranDoc and Research Assistant at the Information Studies Laboratory (InfoLab) at the University of Tehran. Currently, he is working on his dissertation on designing a business model for providing digital information resources (the case of ETDs). He is interested in scientometrics, business models, ETDs, digital libraries and copyright. His homepage can be accessed through irandoc.ac.ir/rasuli/rasuli.html.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com