ELECTRONIC DOCUMENTS IN A PRINT WORLD: GREY LITERATURE AND THE INTERNET

Amanda Lawrence

Published in *Media International Australia*, 2012, (143), 122-131
http://www.uq.edu.au/mia/2012-issues#143


Abstract

Reports and documents from government and other organisations have existed for centuries, but in the post-war period their production increased significantly. Computers, databases, desktop publishing software and the internet have revolutionised the way documents can be produced and disseminated, allowing individuals, groups and organisations access to a whole new world of information. The result has been an explosion in online publishing that has transformed scholarly communication. Research reports – or grey literature as they are also known – are now an essential part of many disciplines, including science and technology, health, environmental science and many areas of public policy. While access to these reports has become easier in many respects, online publishing presents many challenges as well, particularly for collecting organisations faced with the task of adapting their systems. The management of grey literature raises many issues that are still not resolved today. This article provides some background to these ongoing challenges in Australia, the United States, the United Kingdom and Europe.

In the last 50 years, computers, databases, desktop publishing software and the internet have transformed the way we are able to produce and disseminate information and documents – electronically, online, with or without charges or copyright restrictions. As a publishing and distribution platform, the internet has been particularly effective in changing the way research reports and other content produced by governments and other organisations are able to be accessed and used. ‘Grey literature’ – material produced informally and circulating outside commercial channels – existed well before the internet, and has been subject to a range of policies and infrastructure across most Western countries to ensure it is identified, collected, catalogued and made accessible. This has never been easy.

In the era of print-only publications with small print runs and no easy means of dissemination or collection, grey literature was often defined by its elusive or fugitive quality. The advent of the internet has solved some old problems by making it easier to procure many reports, but also created new ones – like keeping track of the increased number of reports. The internet has not solved the problem of access to grey literature. As Smith (2009: 31) explains: ‘Despite its value as a source of information about what works, the grey literature is notoriously ill-organised, distributed and promoted. It was when it was print based, and continues to be so now that it is online.’

In this article, I provide a brief history of grey literature, its origins in print form and the policies and infrastructure that were put in place to manage it. I then look at the impact of computers and the internet, and the way these developments transformed the whole research communication enterprise, including the value of grey literature and the practices and infrastructure required to manage it, looking at examples of how access and collection have been dealt with in the United States, the United Kingdom, Europe and Australia. Finally, I analyse why the information policy settings and infrastructure applied to grey literature in the internet era have been inadequate for dealing with this important research resource.
Defining grey literature

The term ‘grey literature’ belongs to the world of the library and information profession. It is a relatively recent collective noun for ‘information produced on all levels of government, academia, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body’, as defined at the Grey Literature Conference, Luxembourg in 1997 and expanded in New York in 2004 (Farace, 1998; Farace and Frantzen, 2005). The term comes from the German phrase ‘Graue literatur’, and did not come into general use until it was adopted at the York Seminar in 1978 (Luzi, 2000: 106; Alberani and Castro, 2001: 236). Other terms used for this material include report literature, research reports, technical reports, government publications, fugitive literature, non-conventional literature, non-published literature, research outputs, small-circulation literature, unpublished literature, and self-publishing (Farace and Schöpfel, 2010). Desktop publishing and the internet added new terms to the list, including electronic publications, online publications, online resources, open access research, ePrints, digital documents, and so on. With this confusing number of terms, ‘grey literature’ has the advantage of being an agreed ‘controlled term’ that information professionals and researchers can use to discuss this distinct group of documents.

Grey literature is best understood by focusing on three factors: the nature of the documents concerned, the types of producers, and the means of dissemination. Using these factors, grey literature encompasses documents such as technical and project reports, working papers, discussion papers, technical manuals, information sheets, conference papers, theses, etc that are produced by government departments and agencies, universities, think tanks, non-government organisations, corporations and professional bodies, and usually lack systematic means of distribution or bibliographic control.

Grey literature is extremely important for many disciplines, including science, technology, health, engineering and social sciences, government and public policy arenas, and a range of professional and commercial practices. The aim of the producing bodies is to share key information with relevant parties on ‘what works’, in the form of technical specifications, project outcomes, changes to policy or legislation, and to do so quickly and easily without the delays and access restrictions of academic journals or book publishing. There is often little incentive or justification for these organisations or individuals to publish in academic journals, and no reason to charge for access to information that they may be mandated to share (Feather and Sturges, 2003: 210). The production and research quality may be extremely high because the reputation of the producing body is vested in the end-product, but they often lack the channels for extensive distribution and bibliographic control.

Grey literature history before the internet

Accounts of the origins of grey literature vary depending on how it is perceived. In a keynote address at the 2011 Grey Literature conference, Jens Vigen traced the history back to oral traditions of scientific knowledge exchange (Vigen, 2011). In 1813, the US government felt there was such a wealth of information being produced by Congress and its agencies that it warranted setting up the Federal Depository Library Program for national distribution and access to government documents (Putnam, 1990). Others point to the technical breakthroughs occurring in new industries and disciplines such as aeronautics, engineering and geology in the early twentieth century; these industries began producing and sharing their documentation with partners and other interested parties in the form of reports and manuals (Luzi, 2000; Augur 1998).

These reports were extremely important not just for sharing scientific information, but also for the development of information systems, public policy and public institutions. For example, the Australian government commissioned a report to look at how to organise scientific research and communication after World War I. Based on the recommendations of this report, Commonwealth Scientific and Industrial Research was established in 1926 – later to become the Commonwealth Scientific and Industrial Research Organisation (CSIRO). Often the decision to bypass conventional publishing systems was deliberate, based on security or commercial confidentiality, and the readership was intended to be restricted or limited. This is partly how grey literature got its reputation as fugitive, elusive and difficult to obtain (Augur, 1989: 1).
World War II saw an even greater urgency attached to the sharing of technical, scientific and policy information across departments, organisations and countries, particularly between the UK, US and its allies including Australia and Canada. Institutions like the US Office of Scientific Research and Development (OSRD), established in 1941 and later replaced by the Publications Board in 1945, were pivotal to the ongoing production, dissemination and control of technical information and reports, and also became the main source of bibliographic records for access and retrieval:”

Such institutions, numerous especially in countries such as the USA and UK, also had the task of divulging internally produced information and, at the same time, of spreading non-classified information to a broader user group to foster technology transfer. (Luzi, 2000: 107)

In the post-war era, government funding for research in the United States, and to a lesser extent in the United Kingdom and Australia, dramatically increased, and scientific, technical and policy reports, along with academic journal publishing, did likewise. In 1950, the US Congress instructed that a repository be established for technical information ‘from whatever sources, foreign and domestic, that may be available’, and to make ‘the results of technological research and development readily available to industry and business, and to the general public’ (AllGov, 2009). This would eventually develop into the National Technical Information Service (NTIS), which would become one of the largest collections of government scientific, technical and business information in the world.

The Weinberg report of 1963, titled Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information, provided an impressive analysis of the many issues involved in ensuring that the abundance of scientific and technical information being produced in the United States was managed and communicated effectively across government, agencies, industry and academia. It was one of the first documents to propose a wholesale information strategy, including valuing the role of report literature – which it described as the ‘crux of the current information crisis’ (Weinberg, 1963: 19). Weinberg reported that the United States was producing 100,000 informal government reports per year, in contrast to 450,000 journal articles – but unlike journal articles, the reports had no bibliographical control or comprehensive collecting system.

Information overload was a recurring theme of Weinberg’s report. One of his key recommendations was that new modes of information retrieval be explored along with the adoption of ‘information-handling technology’. At the same time, Weinberg warned that the promise and limitations of such technology should be well understood before they are adopted as ‘magical panaceas’ (1963: 21). Weinberg’s report encouraged the establishment of computer databases and bibliographic cataloguing systems for report literature – indeed, the NTIS moved to digital records the following year. Other important databases established around this time in the United States include the Education Research Information Centre (ERIC) for educational resources, which was also one of the first decentralised systems of collaborative collection and cataloguing (Luzi, 2000: 109). Weinberg’s report was considered so important that the inaugural International Conference on Grey Literature in 1993 had as its theme ‘Weinberg 2000’ in an effort to inspire the next generation to produce a similarly groundbreaking understanding of information policy.

In the United Kingdom, the British Library in the post-war period played a crucial role in collecting and managing grey literature – particularly the profusion of scientific and technical reports coming out of the United States fuelled by Cold War competition with the Soviet Union. The UK government’s desire to ensure this information spread to its own scientific community led to the establishment of a specialised collection of technical reports in the 1950s which soon expanded to include the social sciences. Not long afterwards, this became the British Library Document Supply Centre (BLDSC), and the British Library came to develop the largest collection of print grey literature in the world (Tillet and Newbold, 2006: 71). It now holds listings of over four million US reports along with those from the United Kingdom and Europe.

In both the United Kingdom and Australia, the Copyright Act 1912, revised in 1968, was a crucial tool for enabling the collection of a great deal of print material, as it enshrined ‘legal deposit’. Section 201, ‘Delivery of library material to the National Library’ requires that a copy of any work published in Australia must be deposited with the National Library of Australia (NLA) and generally the relevant state library soon after publication (NLA, 2010). According to the NLA’s guidelines: ‘A work is

Lawrence, A 2012, Electronic documents in a print world: grey literature and the internet, Media International Australia, 143
deemed published if reproduction of the work or edition has been made available, whether by sale or otherwise, to the public.’ (NLA 2010) Works that must be deposited include books, periodicals, newspapers, pamphlets, sheets of letter-press, sheets of music, maps, plans, charts or tables. The library interprets periodicals to include newsletters and annual reports (NLA, 2012). While not explicitly mentioning reports or grey literature, these have always been among the NLA’s collection in print form – particularly government documents. This broad harvest has resulted in a wealth of cultural and historical material of all kinds, and the NLA now boasts the greatest collection of Australian historical material, and also has an outstanding collection from Asia and the Pacific.

By using such a wide definition of published material, the NLA has not had to engage in debates and definitions of grey literature, and I have not been able to find any official documentation referring to this term in the NLA’s collection policies (see Pandora, 2005; NLA 2008). However, the NLA does have policies for the collection of print and online government documents from Australia, which it aims to collect comprehensively, and a policy for publications from international organisations that it has collected to a greater or lesser degree over the years (NLA Collections Policy, 2008). Thus, while it certainly collects grey literature, the NLA has not been active in the ongoing discussions of its nature and value, making an analysis of the NLA’s role in grey literature’s history more difficult.

The impact of computers and the internet on grey literature

From the 1970s, the use of databases and computerised catalogues for specialist collections and in libraries became widespread. These were often known as automated systems, as they replaced manual card catalogues. Grey literature was often at the forefront of new database systems, reflecting its importance in science and technology research and development, its quality of being informally published and distributed, and the fact that its collection often required special policies and practices.

One of the main motivations for building the internet was resource- and file-sharing between computers and groups of researchers and government. Produced and distributed informally for sharing amongst a select group of specialists, grey literature was a natural fit for dissemination via the new network. Email sharing of files became a major new channel for grey literature circulation amongst insiders. Before the internet, one of grey literature’s defining characteristics was that it was difficult to identify and retrieve, in contrast to journal articles that may have imposed a charge but were organised, indexed and locatable. The internet changed this whole equation by allowing individuals and organisations to send their reports and documents cheaply and easily around the world to peers, partners and other organisations.

The internet transformed the way documents could be produced, disseminated and retrieved. Documents could be published and accessible across the globe in an instant without the huge cost of print and distribution, and shared amongst large or small groups via a wide range of formal and informal networks. Bibliographic databases with full text holdings or links could be created for specialist and general audiences, providing access to a wealth of scholarship and technical know-how at very little cost. Governments and their agencies could provide access to policy changes and seek feedback via online documents and websites.

The impact of the internet was such that in 1997 a new definition of grey literature was created, which moved away from being hard to retrieve and focused instead on its form, circulation channels and producing organisation. The annual grey literature conferences that had begun in the early 1990s became pivotal for this redefining process, producing the formal definition outlined earlier that is now in common use.

Grey literature was at the forefront of what became the open access movement, forcing many involved in scholarly research to question the dominance of peer-reviewed journals, their value to the public purse, their efficacy for research dissemination and their claims to quality and accuracy (Thomas, 2005). Houghton et al. (2006) have shown that increasing open access to research articles has a potential benefit of $150 million per year in Australia alone.

The internet saw the rise of search engines able to crawl the web and whole documents for specific references or terms. It forced organisations to be more accountable, and generated new forms of online publishing including pre-prints, working papers, websites, multimedia and interactive formats, blogs, FAQs and more. All of this abundance needed to be collected and preserved for long-term access and
retrieval, and libraries and other players had to innovate quickly to keep up with the new information landscape, which required new forms of cooperation and infrastructure, new methods of evaluation and quality control, new systems for preservation and new metadata schemes to classify it all – not to mention more funding, authority and technical capacity to carry out such work. While the internet has created amazing new opportunities, it has not been the ‘magical panacea to information-handling woes’, as Weinberg wisely predicted.

**Catalogues, collections and the internet**

In Australia, the NLA introduced machine-readable catalogues in the early 1980s. The Australian Bibliographic Network replaced card catalogues for staff data entry and retrieval, although readers still had to use microfiche for another decade before they could do their own search of a library catalogue on computer. One of the first large-scale online catalogues was developed at Ohio State University in 1975, later to become the Online Computer Library Center (OCLC) and WorldCat – the world’s largest online collective library catalogue: ‘Using a dedicated terminal or telnet client, users could search a handful of pre-coordinate indexes and browse the resulting display in much the same way they had previously navigated the card catalog.’ (Wikipedia, 2012) It wasn’t until the 1990s that online public access catalogues (OPACs) became common across the library sector. The NLA’s OPAC went live in 1992 (Paton and McDonell, 1999).

With the massive increase in online content in the 1990s, many others outside libraries saw the need to try to collect and make accessible valuable resources for their interest groups and disciplines. Existing databases went online and new ‘clearing houses’ (specialist resource and information collections either print or online), databases, digital libraries, portals, repositories, subject gateways, aggregators and archives were established. Given grey literature’s unconventional nature, these have been developed by a wide range of bodies, including libraries, universities, government bodies, not-for-profits and commercial companies, as well as national, international and state-based initiatives, all trying to provide some kind of service for identifying, locating, collecting, cataloguing, disseminating or preserving online resources and publications. The internet has meant that: ‘Both customer needs and supplier routes for grey literature have changed significantly bringing new challenges.’ (Tillet and Newbold, 2006: 72) The business of creating order from such a highly unorganised system has been extremely challenging, not to mention the issues of competing interests, institutional legacies, outdated legislation and other complicating factors that have prevented the kind of streamlined access to content that would benefit all of society.

In the United States, for example, there has been controversy over the role of the NTIS and its commercial ventures. In the pre-internet era, the NTIS was set up with the expectation that it would recoup its operating costs, despite the fact that its main function was to collect scientific and technical reports, much of it produced by the US federal government and its agencies. In the years when access to information was expensive and difficult, the NTIS could justify its charges, but in the changed circumstances of the 1990s it began to lose money. In 1999, it was proposed that the service would go into partnership with a commercial company to provide a fee-based search function. The outcry over locating government-funded information behind a paywall was exacerbated by the now common assumption that the internet was about open access and free information.

The Secretary of Commerce, William M. Daley, announced plans to shut down the NTIS and move the archive and document-delivery service to the Library of Congress. Daley argued that it was no longer relevant for the NTIS to charge for government information when documents were being posted on government websites free for anyone to access:

> These changes in the information marketplace have made obsolete the need for NTIS to serve as a clearinghouse and, thus have in turn made it increasingly difficult for NTIS to maintain its operation on a self-sustaining basis, as established by Congress. (Quint and Hane, 1999)

The internet also put the NTIS into direct competition with the US Government Printing Office (GPO), and with new collaborations such as the Science.gov Alliance of scientists and organisations keen to make the most of the internet’s potential for sharing research. In the end, NTIS was not handed over to the Library of Congress, but its existence is continually under threat, and involved in territorial
disputes with similar agencies, as the United States tries to work out how to provide access to government information and research reports. The GPO and the Federal Depository Program still have their overlapping role of distributing government information. The Library of Congress remains another major player in the mix. In addition to these established institutions, new collections such as the Internet Archive and the New York Academy of Medicine’s Grey Literature Report have appeared among a plethora of US online portals and gateways. Duplication of effort, inefficiencies, confusion of purpose and competition among public entities are all problems that have beset the collection of online grey literature.

Following on from the York Seminar of 1978, European organisations became more active and organised in their approach to grey literature (Okoroma, 2011). The System for Information on Grey Literature in Europe (SIGLE) database was established in 1980, with a large proportion of its content provided by the British Library (Tillet and Newbold, 2006). SIGLE brought the Europeans to the forefront of collaborative cataloguing, and was also important for its broad collecting policies for grey literature – including social science as well as science and medical content. In some ways the United Kingdom and Europe were able to benefit from the slower pace of research funding and production compared with the United States, and from having smaller, distributed national collections that could remain autonomous but be integrated under a unified search system. Yet, like NTIS, the SIGLE database charged for access, and was not actually online, with access being provided by STI International or via CD-ROM. In 2005, the members agreed to end the SIGLE database as it was too hard to bring the system online, but within two years the archive was reclaimed by the French INIST-CNRS and with some technical work was made available via open access and online under the name OpenSIGLE. In 2011, the name was changed to OpenGrey (opengrey.eu), an online catalogue with bibliographic data on both print and online grey literature from a range of major institutions across Europe and the United Kingdom.

One online collection of social science grey literature with which I have been involved for a number of years is Australian Policy Online (apo.org.au). Established in 2002 by the Institute for Social Research at Swinburne University of Technology in Melbourne, the database provides free access to over 15,000 listings of government, academic, NGO and professional research reports and articles, with links to the full-text documents. Full-text copies of documents are not yet held in any systematic way due to copyright restrictions, but the issues of dead links, full-text search, duplicate cataloguing, and the need for identifiers and preservation strategies are prompting a review of how this could be achieved in the current copyright and technological environment. APO monitors around 400 sources of grey literature in Australia and some internationally across a wide range of social, health, economic and environmental subject areas. Importantly, it also provides an alert service via email newsletter, RSS feeds and Twitter.

The APO database uses the OAI protocol to feed content into the NLA’s Trove search engine, and is also easily searchable via Google, which helps to improve exposure for specific content. APO’s daily cataloguing and dissemination systems are designed to ensure the service and its content have the currency required for public policy debates. As a major US report into sustainable digital preservation and access states, issues around grey literature are by no means limited to libraries (Blue Ribbon Taskforce 2010: 3). This is the kind of function that a national, state or academic library has not traditionally provided, and the internet has created a clear and continuing role for services, databases and networks that complement major agencies by collecting and disseminating specialist content to particular audiences. The future lies in connecting up online collections and sharing the load of cataloguing valuable online content as even the largest institutions are struggling to keep up.

Ironically, in Australia and the United Kingdom national libraries have been constrained in fulfilling their charters in the internet age by copyright legislation. Although Australia’s legal deposit provisions are broad in scope, they do not require deposit of electronic publications. What this has meant in practice is that libraries must seek permission before copying an online publication or website into their archives and giving public access. McShane and Thomas (2010) have pointed out that ‘digital strategies have particular significance for major public libraries – the chief public information portals and cultural storehouses of liberal democracies’. Yet, as they argue, libraries have slipped out of the policy line of sight, as the policy rhetoric has narrowed in recent decades from information society,
through information economy, to the digital economy. The National Library has argued continuously for a change to the legislation over the last 20 years, yet – despite reviews of the Copyright Act in 1997, 2000, 2003, 2005 and 2008 – there has as yet been no change to legal deposit (Attorney General’s Department, 2011).

It is the very nature of the internet – which allows content to be accessed easily and reproduced – that poses a challenge for the NLA in seeking an extension of Legal Deposit. Commercial publishers and many other producers see this as a potential infringement of their rights. Change is possible, however. The UK legal deposit legislation was revised to incorporate electronic publications in 2003, although not without a considerable effort on the part of open access advocates and collecting agencies. In Australia, things have moved more slowly. A review of the Extension of Legal Deposit of ‘Library Material’ was initiated by the Attorney-General’s Department in 2007, but subsequently went dormant following the conclusion of the submission process in 2008. There was some improvement for the efficiency of NLA’s online collection practices when in 2010 the Australian Government Information Management Office announced whole-of-government arrangements for the NLA to ‘collect, preserve and make accessible in perpetuity Commonwealth Government online (digital) publications and websites’ (AGIMO, 2010). All unclassified government documents are now to be provided to the NLA by specified agencies without the need for copyright permission to be sought by the library (AGIMO, 2010).

Finally, in March 2012, a Review of the Extension of Legal Deposit as it applies to the National Library of Australia was announced, and a discussion paper was published (Attorney-General’s Department, 2012). It will be interesting to see whether this finally results in a victory for the NLA, and to ascertain the extent to which this will solve the collection and access problems for online publications, and grey literature in particular. The model proposed involves ‘deposit by demand’, based on the fact that everything on the internet is effectively ‘published’ – as in supplied to the public (Attorney-General’s Department, 2012). The NLA will still therefore have to face the issue of finding and evaluating grey literature and providing the resources to catalogue it. The model also proposes that content collected by the National Library will be available to the public within the existing copyright framework for libraries and archives, meaning that while the NLA can at least preserve a copy of an online work, it is not necessarily available for public access online. While there will be various administrative and strategic issues to deal with, Extension of Legal Deposit was generally supported in 2007, even by those concerned about its impact on commercial operations (Attorney-General’s Department, 2012: 8).

As well as individual documents, the idea that the internet itself, websites and web pages should be archived became a reality with the creation of the US-based Internet Archive (archive.org) in 1996 and Australia’s Pandora archive in the same year (pandora.nla.gov.au/about). The NLA established PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) with the aim of creating permanent records and an online archive of selected websites and online publications. From the beginning, it was clear that the amount of material being made available online would be beyond the capacity of the NLA alone, so Pandora was set up in collaboration with state and territory libraries, the National Gallery of Australia, the Australian War Memorial, the National Film and Sound Archive and the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS), with each institution choosing a selection of websites and resources to catalogue and archive based on their area of interest and expertise. Unlike some other national web archive strategies, PANDORA is not a snapshot of Australian websites at any given time, but a curated collection that includes publications such as grey literature.

From 2005, the NLA began conducting an annual snapshot of the .au domain; however, due to copyright restrictions, access to this archive is restricted. A review is currently underway of the technical capacity and the policy settings for the NLA’s web archiving, and it is anticipated that this will change significantly in the next few years.

**Conclusion**

Grey literature continues to play a vital role in the dissemination of research findings and government policy, a role that has greatly increased with the internet. Yet, despite 50 years of automated catalogues
and online collections, we are still a long way from resolving many of the issues that surround grey literature – online or offline. The internet has turned many aspects of grey literature management on its head, yet so many of the characteristics of report literature in the pre-internet era remain thorny issues today. Finding better ways to access, control, evaluate, collect and preserve grey literature is an important national and international issue in the twenty-first century, just as it was in the nineteenth and twentieth centuries, and possibly even earlier.

While much grey literature is now available as open access online, the intellectual battle against the dominance of subscription-only journals has only just begun. It has mainly been waged by a movement explicitly focused on changing the way peer-reviewed journals are published, and in their internal battle on which form of open access should predominate. There has been much less written about the quiet revolution that has occurred as grey literature has gone online over the last 20 years, although it has arguably had much greater impact in making research and technological innovation accessible to all, and may offer both lessons and synergies with those who seek open access academic journals. A concerted effort to achieve policy changes, infrastructure investment and collaborative approaches is needed to really unlock the potential, and reap the value, of grey literature in an online world.

References

Alberani, V. and De Castro, P. 2001, Grey Literature from the York Seminar (UK) of 1978 to the Year 2000, INSPEL, no. 35, pp. 236–47.


Amanda Lawrence is the Managing Editor of Australian Policy Online, Institute for Social Research, Swinburne University of Technology.