Chapter 4
Collection building with special Regards to Report Literature

Elizabeth Newbold and Jennie Grimshaw
The British Library, UK

4.1 Collection Building

In this chapter we will look at the meaning of collection building in the digital age, concentrating on some of the drivers and issues that can be applied to any type of library, but with an emphasis on our experiences in the United Kingdom. Specific to this chapter, we will look at report literature in the sciences and social sciences. This can include, but is not limited to: research, practice, evaluation and development reports distributed by government, international or intergovernmental organisations; policy, regulatory and guidance materials from central or local government and agencies; reports or technical papers produced by research institutes, think tanks or consultants; and material from voluntary and community sector organisations and charities. Firstly, let us consider what we mean by collection building and collection development?

Collection building is a process usually conducted over a period of time that shapes a collection of resources into a cohesive, balanced and useful set of material for a given user community. Collection building encompasses not only material owned by a library in both physical and electronic formats but the also fee-based electronic resources to which it subscribes and free Internet sites to which it links. Collection building takes place not only in all academic, research and special libraries but also among providers of abstracting and indexing databases and information gateways. It includes the organisation and presentation as well as the acquisition of material.

4.1.1 Why build collections?

Mostly librarians have in the past built collections to have materials available for patrons ‘just in case’ they are needed. However, is this still a relevant and useful activity in an increasingly digital and connected environment where information is but a click away on any networked computer? Should librarians still be building
collections and what is the purpose of this activity? A collection should be a useful aggregation of resources based on user needs and demands. Easier access to material online has placed users at the centre of collection building activities and libraries are adopting more ‘just in time’ approaches to collection development. Collection building is not just about what a library or information unit holds in either physical or digital locations but also about how it enables resource discovery and facilitates access to material. The role of collection development and collection building is to organise and index the wealth of information available in print and electronic formats so that users can home in on what they need quickly and accurately and to ensure that access is maintained over time. So, although collections are now hybrids including both print and electronic material, building them requires traditional activities and skills:

Selection

Selection has always been fundamental to collection building. Selection brings together material from different sources and locations and in a range of formats to provide a coherent resource for users. While researchers have domain expertise in their subjects, the role of the librarian is to have domain expertise in the information related to those subjects, where to find it, what quality criteria to apply and how to source it. The role of selection goes beyond that which is found by a simple internet search; it identifies a range of sources of relevant material in print and electronic formats, monitors their ongoing availability and scans the horizon for new sources.

Organisation

One of the most important aspects of collection development is the organisation of information so that it can be quickly, comprehensively and accurately retrieved by users. Report literature is found in multiple formats and is produced by a wide range of organisations. It may be owned by the library or information unit, have been deposited in an institutional or subject repository, be included in a fee-based database, or be published free on the Internet. Traditional library skills of resource description and indexing are needed to facilitate its efficient discovery by users.

Long term access

One of the reasons for building collections is to ensure long term access to material. Grey literature in hard copy has always been difficult to find and acquire because it is not covered by mainstream abstracting and indexing services, is produced in limited print runs, and is not available through the book trade. With the advent of web publishing grey literature is in some ways easier to discover through simple internet searches, but long term access to the material is much less
certain. Much of the material is now only available as a web page or as a document attached to a web page and has no print equivalent. Web pages are transient in nature; some estimates put the life span of a web page at just 75 days (Lawrence, 2001). Links to material on web sites are often broken when these are redesigned and uniform resource locators (URLs) are changed. Whether the content has disappeared or only moved is to some extent immaterial; librarians building collections will need to solve the problem of ensuring long-term access to it.

Continuity of access - switch from print to online only publication

Many reports and especially official publications texts are published in series. In some instances, especially in the case of government publications, the series may have been in existence for decades. Libraries hold long runs of the print publications, which can suddenly cease when the documents migrate to the web and the hard copy version is no longer produced¹. For the convenience of researchers seeking information over time, the successor electronic versions need to be collected and made accessible by the same institutions. Ideally the whole time series should be made available at one location, and the print and electronic versions linked in library catalogues.

4.1.2 What is the role of grey literature in a collection?

As we know there is considerable debate over the definition of grey literature. The representation of grey literature in library collections varies considerably: in some specialist and technical libraries the majority of the collection may consist of grey material, while in other institutions it may be a small percentage of the total holdings. Librarians have traditionally been wary of grey literature, due to the difficulties involved in identifying, acquiring, cataloguing and shelving it. One of the most common words that comes up in conversation with librarians about grey literature is “difficult”. Grey literature is an often-overlooked resource and does not always figure in the collection development policies or selection guidelines of libraries (Lehman and Webster, 2005).

The fact that grey literature has often been overlooked and is difficult to deal with does not diminish its worth as it offers many benefits to users. Grey literature, especially report literature, provides access to high caliber research often not published elsewhere. Documents produced and prepared for one organisation or purpose may have relevance or resonance for other audiences. It is

¹ For example UK Civil Service Statistics have been produced since 1950 and are available online at http://www.civilservice.gov.uk from 1970. They are no longer produced in print and the data is now produced and provided by the Office for National Statistics at https://www.statistics.gov.uk The run of the series is now split between print only editions, dual print and electronic versions and at least two different web locations
often free or low cost and can provide a faster route to publication than that offered by academic monographs or peer-reviewed scholarly journals. Information in research reports is often more detailed than in journal articles, where page limits may have been imposed. Moreover they cover failures as well as successes.

There is a gradual movement among researchers to recognize the value of grey literature and it is seemingly becoming an increasingly important and useful resource, particularly for multi-disciplinary research and in areas such as systematic reviews in evidence based health and social policy, in the dissemination of best practice in social care, and in the relatively new field of systematic reviews in environmental sciences. Systematic reviewers in all fields not only need access to research results published in report literature, but also themselves publish their work as reports. They are thus both users and producers of grey literature. The growing importance of evidence based practice in applied sciences such as healthcare and in social policy is creating greater awareness among researchers of the relevance and usefulness of grey material. This is in turn generating a demand for its inclusion in library collections.

4.2 Approaches to collection building

As grey literature is in essence not commercially published it has to be obtained in most cases directly from the producers. In the ‘traditional’ print environment the two main mechanisms for acquiring material were either purchase on subscription or as individual items, or by donation. Both of these approaches are still relevant today but given the migration of grey material to web-only publication libraries are faced with a slightly more complicated range of options for collecting it:

- reliance on internet access – instead of libraries spending time creating, organising and providing access to a physical collection, the role of the information professional is to assist researchers in locating material on the web.
- downloading and archiving – including the harvesting of metadata, downloading of individual documents from the internet, and the archiving of whole web sites
- use of institutional and discipline repositories
- use of commercial and non-commercial services

It is highly likely that most organisation and information professionals adding grey literature to their collections will need to consider all approaches as at the time of writing, material is available in many formats and has not migrated fully to the digital world.

---

Further information on systematic review in environmental sciences is available from the Collaboration for Environmental Evidence http://www.environmentalevidence.org/
4.2.1 Reliance on the internet

Many researchers argue that all the materials that they need are available free on the Internet and accessible via search engines (Snyder, 2008), and there is a growing body of research on user behaviour that illustrates the use of Google and other search engines as the premier tools for information retrieval. Libraries need to make a decision on whether to continue to seek to own material or to take a linking approach to building a collection of electronic materials. If a linking approach is chosen, the librarian or information professional needs to consider the stability and longevity of the source. They then need to consider how to provide the link to the resource, for example via a portal, through home-produced subject guides or by links from records in a traditional library catalogue. In this scenario the role of the library has completely changed. Instead of creating, organizing and providing access to a physical collection of locally held material, the information professional’s role is to assist researchers in locating the materials they require via a range of free and priced electronic resources and to provide links to material held elsewhere.

This route has been followed by a number of libraries to a greater or lesser extent, with some information professionals providing sophisticated linkages to content whilst others are assuming more of an enabling role, training end users in electronic information retrieval techniques rather than building collections of material. In principle, the availability of such a wealth of information online should make research much quicker and easier, and negate the need to own material. However this approach may be short sighted for a number of reasons. There is ample evidence that the presence of grey literature on the internet is ephemeral due to organisational change, web site redesign and the removal of material considered to be out of date, leading to broken links, loss of access and very frustrated users.

Problems of loss of access

When organisations change name, merge or are abolished material on their websites is often not migrated to the site of the new body. The old site may be maintained for a limited period, but not indefinitely. The skill for the information professional is to anticipate when this may happen and react accordingly. Whilst this may sound like crystal ball gazing, such proactive monitoring and awareness of the changing information landscape is becoming an essential skill for collection building and ensuring continuing access to material. Whilst some changes to web sites and loss of material may happen without any prior warning, others are predictable, for example the general overhaul of public communication which follows a general election when a different political party comes into power.

Other examples of organisational reforms leading to radical web site redesign include changes in the machinery of government, when departments split and merge. Departmental histories in the UK are extremely convoluted and it is not
unusual for departments to split, merge, change name or disappear within the lifetime of a government and not just due to changes following a general election. With each change the website is redesigned and it is not unusual to find that some information has been deleted during the transition. Other information will have moved to a new web address. In these situations there is usually some warning but it is essential for the librarian to react swiftly and it requires effort on the part of the library to ensure that key documents are tracked down and links reestablished. National website archiving programmes may help. In the UK the UK Government Web Archive and Web Continuity Project developed by The National Archives seeks to ensure that links to government websites continue to work and that pieces of online information cited remains accessible in perpetuity. Initiatives such as this ensure that material remains available and enables librarians to adopt a linking approach to collection development with more confidence.

Information may also be deleted simply because it is out of date or no longer reflects current policy. It would appear to an external audience that some organisations have prioritized website housekeeping without regard for the value of the information on the site and take the approach of simply deleting older information. The length of time a webpage or website has existed is not an adequate indicator of value and this approach causes immense difficulties to librarians as they attempt to maintain links to older information. Automated tools that check website links are a useful aid, but librarians need to develop in depth intelligence on the practices of key organisations in order to anticipate changes and take appropriate action. In other cases material remains online but is moved and the URL is changed, so that links to pages and documents bookmarked by users, cited in academic research, or retrieved by search engines are broken, again requiring the expenditure of scarce resources to track down the new location.

4.2.2 Downloading and Archiving

Having considered reliance on the internet, what are the options if the host site is not considered stable enough for long-term maintenance of access? We shall now consider the feasibility of downloading or harvesting individual documents or archiving whole websites. An attractive option is the harvesting or manual downloading of individual documents freely available on the internet.

Individual documents

Libraries wishing to download and keep individual documents in their subject area published free on the Internet have two options. They may choose to seek formal, written permission to download and archive from the web site publisher and/or the

3 Further information on the UK Government Web Archive and Web Continuity project is available at http://nationalarchives.gov.uk/webarchive/
rights holder, or they may opt for an informal verbal assurance that there is no objection. Experience at the British Library (BL) suggests that the latter approach would be more fruitful than the former. An action research experiment took place within Social Sciences Collections and Research at the British Library in April-July 2008. Staff contacted a sample of UK organisations in the field by telephone or email and initiated a dialogue aimed at persuading the body to voluntarily deposit their electronic publications for long term preservation and use within the reading rooms. This piece of action research identified a number of barriers:

- Finding an appropriate contact, especially one that had the authority to agree to electronic deposit, proved challenging. Organisational structures and staff changes added to problems here, with the final agreement often needed from senior management representatives.
- Organisations were very wary of e-deposit and were unclear about who had power of decision to authorise it. They were unclear about what they were “letting themselves in for” and were fearful of unforeseen consequences.
- Legal barriers proved a deterrent, especially the burden placed on publishers by the Library’s requirement that they get permission to deposit from third party rights holders, i.e. authors of individual reports. This involved them in a great deal of work. On the other hand, many reacted with puzzlement and asked why the BL did not simply download a copy of whatever free web documents it required, referring to the example of the activities of the Internet Archive as a precedent. In other words, they were quite happy for us to gather the material, provided that they did not have to take the risk of signing a formal licence or use scarce resources contacting authors of individual reports to get their permission to deposit.
- Lack of interest within the organization was a significant problem, when staff was busy and the question of deposit of electronic publications did not figure on their list of important things to do. In the case of small charities, staff, often volunteers, are focused on serving their users and do not have spare time or energy to engage with electronic deposit issues.
- Technical barriers also manifested themselves, as not all organisations had staff with sufficient knowledge or competence to manage the transfer if a “push” rather than “pull” approach was used.

The downloading of individual documents for retention in library collections is something of a legal minefield even though many organisations offer the option of downloading documents from their websites free of charge at least for private research and study. Information professionals need to be aware of the intellectual property rights associated with that content, which are often complicated and time consuming to unravel. In some instances it is perhaps easier to fall back on to the familiarity of print copies of documents, where they are available, to avoid the added challenges associated with grey literature living wild on the web.
Web site archiving and harvesting

Another approach to the securing of long term access to grey literature published on the web is through web site archiving programmes. In the UK the British Library has been selectively archiving sites since 2005 initially as part of the UK Web Archiving Consortium (UKWAC)\(^4\) and since 2008 as an individual contributor to the UK Web Archive. Tight legal constraints prevent the archiving of web sites without permission from the rights holders. In order to comply with the letter of UK copyright law, written permission to archive needs to be obtained from the web site publishers and other rights holders, including authors of individual documents and even contributors to discussion forums. Sites which the BL and its partners have gathered are made publicly accessible via the UK Web Archive web site\(^5\), permission therefore has to be sought to both gather the material and make it publicly available. Seeking these permissions has proved time consuming and problematic and the success rate is currently very low at about 25% of requests made being granted. Common concerns raised by organisations from which permission to archive is sought are similar to those indentified in the experimental programme for requesting authorisation to download individual documents and include:

- Uncertainty about who in the organisation has authority to give permission, and lack of time and motivation to address the issue.
- The burden laid on the site publisher of gaining permission for the BL to archive material on their web site where the copyright is held by a third party. Such material includes documents like research reports where the copyright is retained by the author, contributions to discussion forums and blogs, and images licensed from picture agencies. It is often impractical for website managers to look through large sites for third party copyright material and then to negotiate with rights holders.
- Concerns about privacy issues, when personal information about individuals is included on web sites.

Thus the barriers to collection building in the digital age using materials available free on the internet are perhaps as much political or legal as technical. We have used examples from our experiences in the UK, but programmes in other countries

---

\(^4\) The consortium was formed in 2004 by six institutions working in partnership: the British Library, the National Libraries of Scotland and Wales, the Wellcome Trust, the National Archives and the Joint Information Services Committee (JISC) to select and archive web sites. It is now evolving into a forum for policy development and sharing of technical expertise and best practice, with an additional advisory function in support of any institutions interested in developing a web archiving programme. In early 2008 the BL launched its new Web Curator Tool, (WCT) software developed in-house for archiving web sites. Access to the WCT, and to a hosting service for archived web sites was made available on a subscription basis, initially to the original UKWAC partner.

\(^5\) UK Web Archive is available at http://www.webarchive.org.uk
face similar if slightly different problems\textsuperscript{6}. Even so, web site archiving provides a useful tool and option for collection building. There is a growing use of web archiving tools to build collections of web resources for different user communities. It is hoped that, as web archiving activities continue to grow, the problem of the requirement to get formal permission to archive from rights holders will be resolved through new legislation, at least as far as the British Library is concerned.

4.2.3 Use of Institutional Repositories and Discipline Repositories

Collection building may also be undertaken internally as a mechanism for storing and preserving the research outputs of organisations. Many research organisations are setting up institutional repositories and an increasing, amount of data is being deposited in institutional and discipline focused repositories.

Much of the activity has been in higher education institutions or research intensive organisations and most of the data collected is in the form of academic journal articles or theses, but it can include reports, conference presentations, working papers, statistical datasets, visual media, and sound recordings. Institutional repositories offer a hub for collecting and preserving the intellectual output of an institution in digital form. They are usually, but not always, open access, and their general aim is to make publicly funded research available, to showcase the research outputs of university departments, and to preserve them for the lifetime of the repository. Content can be retrieved via a number of routes including links from the Library catalogue, search engines such as Google, the repository’s own search interface and specialist services such as Intute Repository Search (IRS)\textsuperscript{7}. It should be noted however, that the quality of the search results is dependent on the quality of the metadata in the original repositories (Kerr, 2008).

There is evidence that digital repositories with high quality structured metadata which is accessible to search engines are successful in making content accessible to users outside of the host institution. Research at Oregon State University showed that only 25% of users of the digital repository were local. The remainder, were users external to the university from other parts of the USA, Canada, India, Europe and Asia including the Middle East. Twice as many users from Europe and Asia access Oregon University’s digital library as from the USA. These external users access the digital library via search engines or the Integrated Library System (Reese, 2008).


\textsuperscript{7} The IRS project is a continuation and enhancement of work carried out under the ePrints UK Project, which aimed to harvest metadata from institutional and subject based eprints archives using the Open Archive Initiative Protocol for Metadata Harvesting. The IRS project builds on this initial work and in the Summer of 2008 was searching across 321,038 working papers, journal articles, reports, conference papers and other scholarly digital objects deposited into 89 UK eprints repositories.
When populating an institutional repository, managers and authors need to comply with intellectual property rights. This will involve identifying the rights holder, and if this is an entity other than the author, e.g. a commercial publisher, gaining their permission to place a copy of the work in the repository. Unfortunately this situation is complicated because authors either do not know who owns the copyright in their published works or are unaware of the limitations that agreements they have signed with commercial publishers put on their re-use. For example, Nieminien (2008) reports that authors at Bradford University were surprised to learn that, while publishers may be willing to permit self-archiving in open access repositories, they are not keen to allow the final published PDF versions of works to be placed in institutional repositories. There is a clear need for further training and awareness building in this area, for authors to be encouraged to retain copyright in their scholarly output (Oppenheim, 2008) and for librarians and information professionals to understand how this material can be re-used. As it is, accessibility outside of the host institution makes institutional repositories an increasingly useful tool for improving and widening access to grey literature.

4.2.4 Commercial and Non-Commercial Services

Commercial services

So far we have looked at options for acquiring or accessing grey literature which is available free of charge, although there are always costs associated with the selection and maintenance of access to the material. We shall now briefly consider some of the commercial services available.

One of the most common forms of dissemination of report literature used to be microform. Although this is still a distribution channel for some producers, it is becoming less commonly available and the most prevalent form of dissemination, and preferred choice for access, is now online internet based services.

Some commercial services are emerging which aggregate and provide fee-based access to full text grey literature. For example, PsycEXTRA™ produced by the American Psychological Association, is a companion to the scholarly PsycINFO. Most of the content was written for professionals and distributed outside of peer-reviewed journals. It includes full text of research reports, newsletters, policy statements, factsheets, annual reports and consumer brochures. In 2008 CABI made the decision to add full text to the standard subscription to the CAB Abstracts database, which includes conference proceedings and reports from government and international organisations. It aims to be a permanent sustainable repository and grow by 10,000 documents per year (CABI, 2008).

---

8 PsycEXTRA information available at http://www.apa.org/psycextra
9 PsycINFO information available at http://www.apa.org/psycinfo
10 Information on CAB Abstracts available at http://www.cabi.org
this instance grey literature is discoverable alongside mainstream journal literature to satisfy growing user expectations of immediate access full text documents. This approach is similar to PsycEXTRA in that they both aim to be a repository for full text materials, but they differ in that one is concerned solely with grey documents and the other provides grey material alongside conventional literature. These two similar but different approaches perhaps reflect differences in how researchers in various disciplines approach information seeking and use of grey literature. However it has to be said that this is still relatively unusual and most indexing services simply link to full text documents on the web, sharing the ever present problems of links which break and content which has been removed. Increasingly we are seeing abstracting and indexing services providing access to full text documents but they still predominantly focus on peer-reviewed journal articles and not grey literature.

More common are bibliographic databases that index report literature. Two long running and well known such services are the National Technical Information Services (NTIS)\textsuperscript{11} database which indexes reports on research sponsored by the United States and selected foreign governments and the Education Resources Information Center (ERIC)\textsuperscript{12} database which covers both journal and report literature in the field. Both these services provide bibliographic records and are backed up by an ordering mechanism for documents. Services such as these are constantly evolving and they too are starting to offer a growing collection of full text content.

The obvious constraint on use of commercial services for building collections will be the size of acquisitions budgets but they can be a cost effective and reliable option. It is perhaps more common to assess them using criteria laid down by an electronic resources collection strategy, but for databases containing full text it is important to also consider them as part of the overall collection of primary research materials. A database such as PsycEXTRA provides a valuable resource in psychology as it offers discovery and delivery of full text through a single stable platform. One of the main barriers to wider use of grey literature is the difficulties researchers can face in getting hold of the document once a reference has been found. Commercial services which aggregate hard-to-find materials in full text have a valuable role to play in enabling swift access. While there is cost associated with subscription to commercial resources, when this is compared to the life cycle costs associated with selecting, acquiring, processing and storing individual documents and the time consuming and resource intensive activities associated with maintaining links to free internet versions, they start to become very attractive options.

\textsuperscript{11} NTIS National Technical Information Service http://www.ntis.gov
\textsuperscript{12} ERIC Education Resources Information Services http://www.eric.ed.gov/
Non commercial services

In addition to subscription based resources, there are also a number of freely available services that offer single points of access to grey literature in a given field through portals and integrated search interfaces. These have a twofold relevance in collection building: a) they can act as a selection tool; and b) they can provide a relatively stable option for end user access. In this instance someone else has done the hard work of collection building for you and has dealt with the problems of selection, metadata creation and links maintenance, unless you are involved in building this type of facility yourself! For the majority of librarians the only task involved is locating and identifying these resources, assessing their relevance and then providing links to them for researchers to find and use. These resources tend to be developed by experts in the field and by information professionals in specialist units, leading to the creation of high quality subject specific collections. Rather than using tools to automatically collect material from the web in the way that some gateways do, the material is hand selected by experts, following defined criteria. Such services therefore provide access to authoritative information. An example of such a gateway is Science.gov\textsuperscript{13}. The advantages of resources such as Science.gov are that they bring together material from very diverse organisations and provide search mechanisms to retrieve documents and information held in a distributed network of repositories.

4.3 Conclusion

How collections are built and what they contain will continue to evolve and develop and collecting practices will adapt to cope with changes in how information is published and disseminated. Ultimately, though, the approach to collection building in relation to grey literature chosen by individual information professionals will be dependent on the goals of the organisation. It will balance the staff and monetary resources available, the relative availability and accessibility of grey material in the subject, the aims of the organisation and the needs of the users. The aim has to be to provide coherent and integrated access to content regardless of where the content is produced and held.

We have outlined some of the current options and activities for collection building with reference to technical and government report literature in the sciences and social sciences, considering some of the benefits and problems associated with the different approaches used. In conclusion, when building collections including grey literature:

- Understand your users and have a clear vision of why you are building the collection in order to provide the right resources in the right format for your users.

\textsuperscript{13} Science.gov available at: http://www.science.gov/
• Understand the current transition from ownership of locally stored resources to the provision of access to resources, with the ‘collection’ offering a relevant mix of print and electronic owned, subscribed to and linked to content.

• Be pragmatic about the acquisition of material and pick approaches that suit your users and organisational needs. In reality all collection building will comprise a mixture of approaches; there is no ‘one size fits all’ when working with grey literature.

• Produce collection development policies, statements and selection guides to make explicit the rationale for the inclusion of content and to explain the exclusion of certain resources. Include criteria used for decisions on whether to own, subscribe to, or link to resources.

• Understand the legal environment that you are working in and what restrictions there may be on access, storage and re-use of both print and electronic content. Working with grey literature available free on the Internet is more complicated than operating in a purely print or purely commercial environment.

Grey literature has always been difficult to collect and manage and will continue to be so. However, the challenges are not unique to grey material and similar issues are being faced by librarians dealing with licensing access to mainstream commercially produced online resources. Perhaps the differences between grey and non-grey material are less clear than they once were.

Collection building in a grey digital world opens up opportunities for librarians and information professionals to develop new skills, provides fertile soil for growing partnerships and collaborations, and most importantly opens the door for the creation of rich and varied collections of resources and development of enhanced access to content for users and researchers. There is still a role for libraries in the building of collections and the inclusion of grey literature in them. What is changing is the traditional notion of what a library collection looks like and the mechanisms for creating it.

References


Niemiinen, S.  Bradford Scholars: an online repository for the University of Bradford, ALISS Quarterly, vol.3, no. 4 2008, p. 20-23

