

# Digital curation of public policy resources: discovery, access and management for policy and practice

Amanda Lawrence  
Research and Strategy Manager  
Australian Policy Online  
Swinburne University of Technology  
[alawrence@swin.edu.au](mailto:alawrence@swin.edu.au)  
 <http://orcid.org/0000-0003-2194-8178>

## **Abstract:**

*Public policy and practice relies on a wide range of resources, including traditional scholarly publications, and those produced directly by organisations, such as reports, discussion papers, briefings, reviews and data sets produced by government, academic centres, non-government organisations (NGOs), think tanks and companies. While heavily used, the collection and curation of digital publications (grey literature) is dispersed, inefficient and inadequate. This paper presents recent research on use, production and collection of policy publications and discusses the approach of Policy Online, a digital library that uses a variety of tools including crowd-sourcing content, linked data approaches, Digital Object Identifiers and more.*

## Introduction

The internet has profoundly changed the production, use and collection of research and information for public policy and practice, with publications produced by organisations playing an increasingly important role. Material produced and published directly by organisations, both in print and digitally, without recourse to the commercial or scholarly publishing industry is known as grey literature. Grey literature is notorious for being difficult to find, evaluate and catalogue. In the digital age a great deal more content is being produced and disseminated directly by organisations, so what does this mean for collecting services? What approaches and strategies are useful in collecting and providing access to digital content, particularly content needed for effective public policy and practice.

This presentation draws on the findings of the Grey Literature Strategies (GLS) research project, an Australian Research Council Linkage Project conducted in partnership with Swinburne University of Technology, Victoria University, the National Library of Australia, National and State Libraries Australasia, Australian Council for Educational Research and the Eidos Institute. Along with the author of this paper, other researchers involved in the project were Julian Thomas, John Houghton and Paul Weldon. The aim of the project was to investigate grey literature's role and importance in public policy and to find ways to enhance its value. The project considered the issues from the perspective of users, producers and collectors of grey literature and sought ways to enhance the value of grey literature through recommendations and guidelines. An overview of the research findings and recommendations was published as a discussion paper in 2014 (Lawrence et al.). This paper focuses in more detail on the findings of the collection survey and also looks at the approach taken by one particular digital library, Australian Policy Online (APO), which has been collecting publications produced by organisations since 2002. Over the last decade much has changed in how digital libraries need to approach their role and APO has had to continually adapt to new technologies and find a sustainable business model in a constantly changing information ecosystem.

## Organisation publishing

Organisations have been producing written materials for distribution publicly and through private networks for almost as long as the existence of the mass printing press. From the protestant revolution, through various political upheavals in Germany, France and Holland, religious and political pamphlets were a major force in the battle for influence and change (Briggs & Burke 2009). In many ways, the legacy of state libraries and even copyright law is inextricably bound to the adoption of new forms of publishing and dissemination by organisations.

Publications produced by organisations, also known as grey literature, circulate through a wide range of informal and ad hoc channels, in numerous formats and for many purposes. A key role has been to contribute to, advocate for, influence and inform social and political issues. In the digital era, grey literature has proliferated, and is a key part of the evidence-base produced and used for public policy and practice.

Grey literature is difficult to define but can be summed up as research and information resources produced and disseminated, publicly or privately, by organisations, outside of the commercial or scholarly publishing industry. The term is

simply a collective noun to bring together a wide range of publication types from reports and papers to evaluations, submissions and blog posts.

Despite its ubiquity, grey literature's role is often overlooked, as it does not fit easily into more recognised structures of commercial publishing and dissemination or scholarly communication and reward systems. As a result, it is also hard to find and collect by libraries and information services.

## **Research methods**

In order to understand how grey literature is used, produced and collected, a range of methods were used by the GLS project to gather research data, both qualitative and quantitative. They include online surveys and interviews with users, producers and collectors of grey literature, a study of subject databases and repositories, and a review of the current literature and policy context, particularly the broad areas of information, innovation and higher education policy in Australia. The analysis also featured methods drawn from various disciplines including that of economics, based on John Houghton's work on measuring the value of open access publishing and data repositories (see for example Houghton et al. 2009; Houghton, Steele & Sheehan 2006).

Three separate but related online surveys were conducted aimed at information users, collecting organisations and producing organisations. The Access and Use Survey (users) had 1012 responses with 943 from Australia (93%). The Producer Survey had 155 respondents with 144 from Australia (93%). The Collector Survey had 126 full responses with 115 from Australia (91%). The analysis provided here is based on Australian responses in order to provide a consistent picture.

For the survey of users, the government sector workers was the largest group, being nearly half of all respondents (44%), followed by those working in education (26%) and the NGO (non-government organisation) sector (23%). The number of respondents from the commercial sector was small, representing only 5% of respondents (N=49). There were also a small number of others not able to be allocated to any of the main four sectors.

The survey of producing organisations was directed at an organisation or a department level response. Of those that responded around a third were from education (35%) and NGOs (37%) with 21% from government and a small portion from the commercial sector (7%).

The survey of collecting organisations and services was also directed at an organization or department level response. The majority of respondents to the collector survey came from government (55%), followed by around a third from the education sector (29%), and 16% from NGOs. There were no respondents in the commercial sector. Figure 1 shows the breakdown of survey respondents by sector.

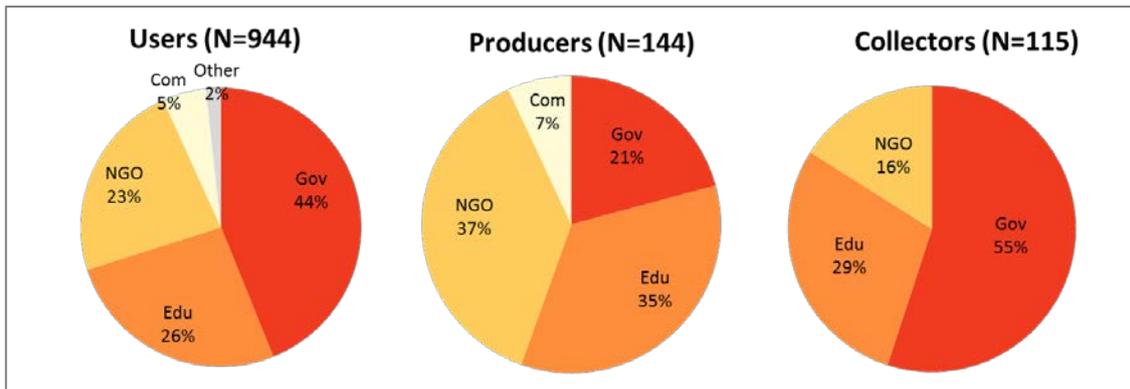


Figure 1. Survey respondents for the user, producer and collector surveys, by sector

### Public policy and practice relies on a wide range of resources

Digital technologies have allowed more people access to more information resources than at any time in the history of the world. Through online hosting, posting, linking, dissemination and sharing users can access massive amounts of information and many are indiscriminate in their choice of resources. As Carol Tenopir argues, readers are willing to read from non-journal peer-reviewed sources and these alternatives are increasingly available. “In the future, the combination of the unwillingness of readers to pay out of their own pockets for articles, the willingness to read from non-journal peer-reviewed sources, and the growing availability of alternatives provides a warning for publishers and librarians.” (Tenopir et al. 2010)

This has certainly been borne out by the GLS survey results showing that the most common resources, consulted regularly or occasionally by over 80% of surveyed policy information users, are reports, journal articles, discussion papers, and ‘briefings, guides and research reviews’. News reports and conference papers are used regularly by 79% of respondents, and two-thirds of policy workers use books and data regularly or occasionally. Working papers, submissions and evaluations are used by more than half of all policy workers regularly or occasionally.

The most important or very important resources used overall are reports (81%), journal articles (75%), discussion papers (69%), briefings, reviews and guides (66%) and data sets (61%). Other materials that are used and valued fairly highly are working papers, submissions and evaluations. The comparative data for use and importance is provided in Table 1, ordered by the material that is most used. The third column shows the importance for only those that use a type of material.

Table 1. Use and importance of materials for *Access and Use* survey respondents

<b>Materials used for work in policy and practice</b>	<b>Used %</b>	<b>Important/Very Important overall %</b>	<b>Important/Very important for those that use %</b>
Reports	86	81	94
Journal articles	85	75	88
Discussion papers	81	69	86
Briefings, guides	80	66	84
News reports	79	49	62
Conference papers	79	52	67
Books and eBooks	68	52	77
Data sets	67	61	91
Working papers	59	45	76
Submissions	58	44	77
Evaluations	51	45	89
Professional/trade mag	46	29	65
Audio/video material	42	14	35
Blogs	39	12	30
Social media	38	15	40
Technical documents	36	30	83
Theses	28	14	49
Commercial/market research	28	17	61
Archival material	26	14	54
Legal docs	20	15	75

Grey literature also makes up a large portion of the source material used for policy and practice work. When asked to estimate, information users report that grey literature makes up 60% or more of the material they consult for their work. For a quarter, grey literature constitutes more than 80% of their source material. Information users report that they value grey literature because:

- their work depends on grey literature;
- grey literature provides a broad view of the research environment and perspectives;
- grey literature is a unique source of information on topics, sources and issues not found elsewhere;
- grey literature is essential for public policy;
- academic journals do not cover the same issues;

- grey literature is widely available online free; and
- grey literature is often the most timely source of information.

An important aspect of this research project was estimating an economic value for grey literature use, production and collection. There are various ways to look at use value (that is, the value to users) including not only their cost to access or purchase but also the time spent accessing and using materials and the contingent valuation based on reported willingness to pay or willingness to accept. These figures were averaged across all user survey respondents and then scaled to the national level. As earlier indicated, it is estimated that in Australia around 3.8 million people might be grey literature users, and that therefore the use value of grey literature is around \$33 billion to \$43 billion per annum. (Further details on the economic valuations will be published in the coming year.)

Policy information seekers use multiple methods to search for and discover resources. As the number of potential access points and information providers has increased and roles and services converge, it has become more and more difficult to get a clear picture of exactly how policy workers find and access information. Dissemination, discovery and access have become increasingly complex in the digital environment. While traditional print and library channels continue to provide valuable services, many new digital curation and information services have emerged, and many producing organisations have also taken on some of these activities.

The Access and Use questionnaire asked respondents, “Is there any material you would use more often if it was easier to find or access?” This was an open question and almost half of the surveyed information users indicated they would use resources more often if they were easier to find or access. The most requested formats being journal articles, data and statistics, reports, and government material.

Users were asked: “What issues, if any, do you have accessing information and research, particularly grey literature, for your work?” Finding relevant resources including knowing what exists and where to look, and the amount of time required to sift and evaluate, are major issues for 45% of information users surveyed. Accessing resources — particularly the cost of journal articles and market research, and problems accessing government content — are problematic for 43% of information users surveyed. Poor production quality, the difficulty of evaluating credibility, the lack of collecting services and problems caused by digital content disappearing from online access (dead links) were also mentioned.

## **Policy publishing by organisations**

In the digital environment, the cost of producing and disseminating grey literature has reduced to the point where any person or organisation can write, publish and participate in the field of public policy discussion and debate. Online publication provides infinite flexibility in terms of format, content and style; these publications are cheap and easy to disseminate, and it is not surprising therefore that the production of digital grey literature has grown exponentially since the development of the internet. Based on data from the Australian Bureau of Statistics as well as other sources, it is estimated that at least 30,000 organisations in Australia currently produce grey literature for public policy and practice purposes.

Based on these survey results, it is clear that many organisations, across all sectors of society, produce their own publications, in order to translate and disseminate new

research or policy positions. More than 90% of producing organisations report that the most important reasons they produce material are to provide an evidence base and to inform public policy and practice. Other important aims are to translate knowledge for public use (84%), and to maximise public access to research and information (79%).

In terms of estimating the economic value of grey literature production, Australian respondents reported their organization or department spends a total of \$234 million per annum on projects that result in the production of grey literature, an average of \$3.3 million per annum per respondent. On this basis, total national grey literature related *project* spending could be around \$33 billion per annum. National R&D spending in Australia is \$28 billion per annum so this figure is commensurate with that amount.

## **Distribution and management of organisation publications**

Digital grey literature presents a far more complex picture of ad hoc distribution channels than traditional publishing supply chains. A producing organisation may send digital copies directly to readers as they do with print copies, they may send email newsletters with links to the online resource, and other organisations and information centres may do the same. Conferences and events are also a key place for dissemination of research results. Forward thinking producing organisations, and other intermediaries, also consider search engine optimisation (SEO), so that researchers seeking information on a given topic, or a specific title, may discover their resources via search engines that have indexed their content using robots and crawlers.

If the producing organisation has bought an ISBN, this may alert the National Library of Australia to its existence, and it may be able to secure a print copy through legal deposit requirements. However, the vast majority of organisation publications do not get ISBNs, and so there is no easy way for libraries or collecting services to know about their existence.

## **Link rot**

Without digital collections and stable links, link and reference rot to grey literature resources and other web content has become pervasive in online content and scholarly publications. The presence of grey literature on the internet today is no guarantee that it will be there tomorrow. Digital grey literature is often all too ephemeral due to organisational change, web site redesign and the removal of material for various reasons, leading to broken links, loss of access and frustrated users. Machinery of government changes following a general election are a well-known cause of dead links, with mass changes to department names and websites. Research projects come and go, as do their websites, and research centres and departments, while slower, have an increasing rate of turnover. NGOs and think tanks also come and go, and the increasing number of these means the churn, and resulting loss of online resources, is disturbingly high. "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."(Newbold & Grimshaw 2010)

A recent project investigating the extent of link rot has found one out of five science, technology and medicine (STM) articles suffer from reference rot, "meaning it is

impossible to revisit the web context that surrounds them some time after their publication”(Klein et al. 2014).

Since the early days of the internet, web archiving has been the most common approach adopted to try to preserve continuity of content access on the web. Web archiving is a good back up option for known content that has disappeared but does not help much with discovery of resources on the live web. Indeed, a number of studies of the use and potential of web archives have found that “the use cases for web archives are not well articulated, and have not engaged the research community in any significant way.” (Meyer, Thomas & Schoeder 2011, p. 4) While web archiving has many passionate exponents, and while it is true that it is the most cost effective way to mass archive vast amounts of the internet, it does not solve users need to discover, evaluate and manage content in the short and medium term. Recent research from New Zealand has found that only 7% of academics surveyed had used the New Zealand web archive and that “many respondents demonstrated confusion between an archived website, and archived items that are available via the web” (Riley & Crookston 2015). This confusion also exists in many of the discussions that occur on how to manage long-term access to research resources amongst information management professionals.

## **Digital curation of policy publications and resources**

While heavily used in policy and practice work, the collection and curation of digital publications (grey literature) is dispersed, inefficient and inadequate. This has created a black hole of resources that are being lost over time. The move to publish policy publications online has created a new paradigm for those whose task it is to support policy and practice work through effective resource provision and information management. However, so far there has been a lack of recognition of the need for new systems and the resources to support long-term access and management of valuable digital content.

The GLS survey of collecting services shows that journal articles and books are the most commonly collected type of material, with 90% of surveyed services collecting them, followed by reports (88%), conference papers (81%), audiovisual material (78%) and discussion papers (75%). Essays, policies and procedures, professional magazines, and archival material are collected by over two thirds of responding collectors.

Material collected the least amongst the respondents includes social media and blogs (12–15%), commercial research (23%), legal documents (27%), evaluations (30%), information sheets (33%), submissions (40%), and data sets (41%). As Figure 2 shows, there is some variation across government, education and NGO collecting organisations in what is collected.

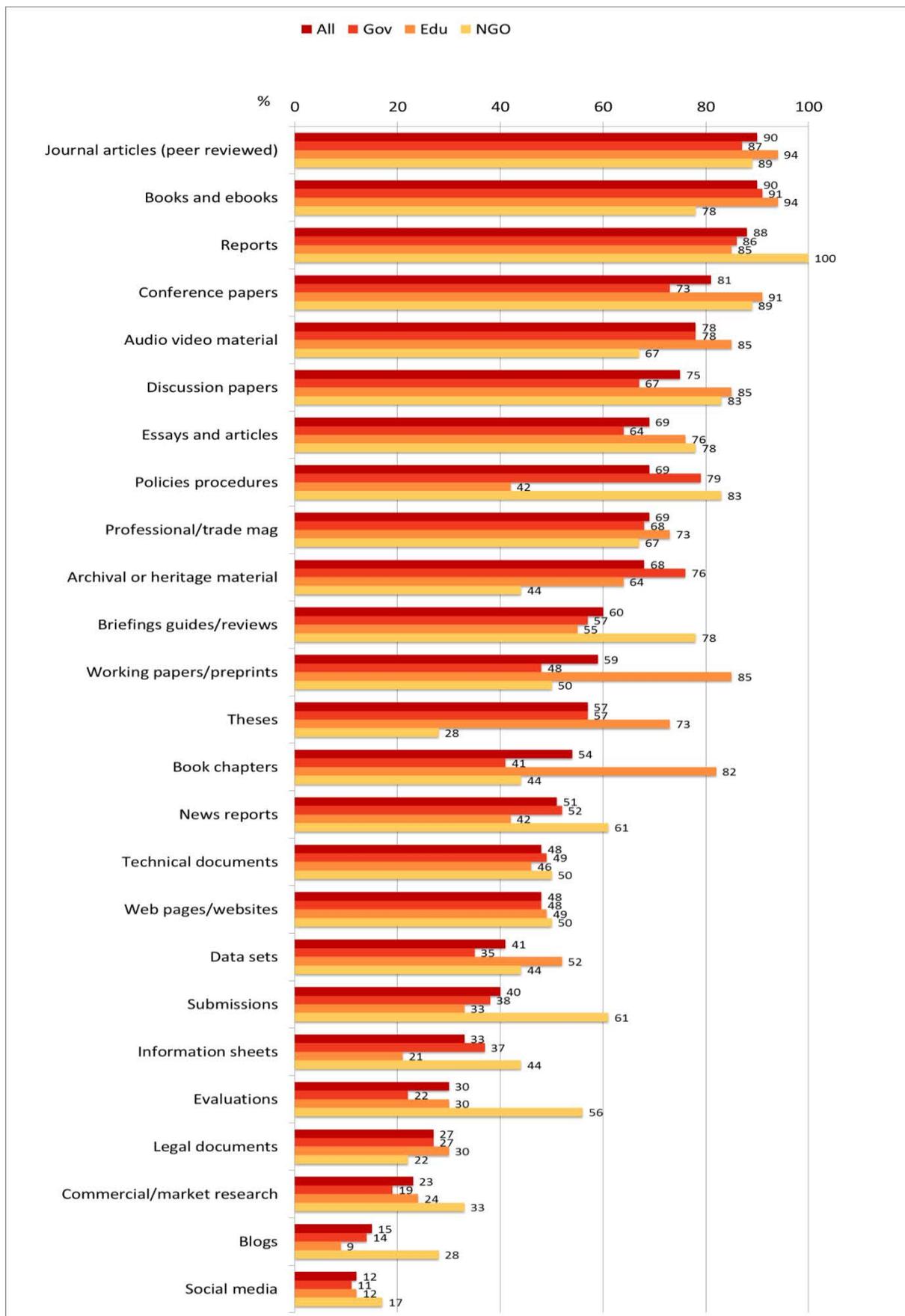


Figure 2. Materials collected by collecting organisations and services, by sector

The most important or very important materials for collecting services overall are also journal articles and books (78%), reports (68%), and conference papers (57%). These are followed by archival materials and professional magazines (54%), and, policies, standards and regulations (53%).

Only one-third of the surveyed collecting organisations consider briefings, data sets or technical reports important. Submissions are important for 27%, and only 21% consider evaluations important. On the other hand, where these items are collected they are considered very important. There are also clear differences across the sectors, with education institutions prioritising working papers, theses and book chapters, while NGOs value briefings and reviews, submissions and evaluations. This may provide a guide to how a distributed network of collecting could be developed to ensure all valuable materials are collected across a range of services.

Table 2. Importance of material collected – percent by sector

<b>Importance of materials collected</b>	<b>Gov %</b>	<b>Edu %</b>	<b>NGO %</b>	<b>All %</b>
Books and eBooks	81	91	44	78
Journal articles	73	91	72	78
Reports	65	64	89	68
Conference papers	46	76	61	57
Archival material	62	48	33	54
Professional/trade mag	54	61	39	54
Policies procedures	57	36	67	53
Discussion papers	43	55	67	50
Essays and articles	49	42	67	50
Audio video material	37	61	33	43
Theses	30	73	22	41
Book chapters	27	73	28	40
Working papers	29	67	33	40
Data sets	33	39	39	36
Briefings guides and reviews	33	33	44	35
News/media releases	32	33	50	35
Technical documents	38	27	33	34
Submissions	25	21	44	27
Web pages websites	21	27	39	25
Legal documents	21	30	17	23
Evaluations	19	12	44	21
Information sheets	17	9	28	17
Commercial/market research	6	15	17	11
Social media	5	9	11	7
Other	5	9	11	7
Blogs	3	6	6	4

## **Difficulty of collecting grey literature**

Librarians at the British Library argue that grey literature available for free online is more complicated to collect than traditional sources of publications, even those produced by organisations. “Whether in print or digital form, grey literature that is

worth considering for a collection is hard to monitor, identify, find, evaluate, catalogue, collect, store and preserve. This adds to both the cost and the expertise required to collect grey literature and, not unsurprisingly, the willingness of libraries or librarians to deal with it.” (Newbold & Grimshaw 2010)

Many resources that are considered important by users are not easy to find for collecting services (Table 3). Evaluations are considered the hardest item to find by responding collectors, with 69% of respondents rating them as not easy to collect. This is followed by data sets, archival material and submissions, which are hard to find for half of all surveyed collectors. Reports, which are the most important source of information for surveyed users, and are collected by 88% of surveyed collectors, are considered hard to find by 45% of collecting services. This indicates a major impediment to efficient collection and access for policy information and research.

Table 3. Materials considered easy or hard to collect by collecting services

<b>Materials</b>	<b>Easy/Very Easy %</b>	<b>Not easy/Not very easy %</b>
Evaluations	31	69
Data sets	49	51
Archival material	49	51
Submissions	50	50
Working papers and	55	45
Conference papers	55	45
Reports	55	45
Market research	57	44
Technical documents	62	38
Theses	63	37
Briefings, guides, reviews	66	34
Discussion papers	66	34
Audio/video material	72	28
Media releases	76	24
Information sheets	77	23
Essays and articles	77	23
Book chapters	78	22
Policies and regulations	80	20
Social media	82	18
Legal documents	83	17
Journal articles	83	17
Web pages/websites	86	14
Blogs	87	13
Professional mags/articles	87	13
News reports	87	13
Books and eBooks	91	10

## Digital curation approaches

“A collection should be a useful aggregation of resources based on user needs and demands...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.”(Newbold & Grimshaw 2010, p. 56) So, collecting involves two main roles: curating and storing resources, and enabling discovery and access.

When it comes to collecting, grey literature has long been a challenge for libraries and collecting services (Luzi 2000). Print grey literature often requires special collection teams and collecting policies. Whether print or digital, grey literature is time-consuming to find and catalogue; it is not distributed through established channels and it does not fall into standardised categories of document classification. Print grey literature is, however, relatively easy to store, requiring little more than some shelves and a computer to catalogue items. The idea of a ‘clearinghouse’ developed as a special collection of print documents curated to serve the interests of a particular group, often associated with specific policy domains.

As online content increased in the 1990s, many others outside libraries saw the need to collect and make accessible relevant resources for their interest groups and disciplines. Existing databases went online and new digital clearinghouses were developed along with electronic databases, digital libraries, portals, repositories, subject gateways, aggregators and archives. 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.

Digital publishing is changing the nature of the problem that libraries and collecting services must solve for information users. In a time where many existing services, such as government libraries, are being closed, and new databases and clearinghouses appear and disappear with similarly unstable funding, it must be considered how collecting and digital curation services can work together efficiently and cost-effectively to help policy workers find and access the information that they need.

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 availability of full-text indexed and preserved versions of government online grey literature in the Australian Government Web Archive and other grey literature archived in PANDORA and searchable via Trove, is an important step in this process and needs to be made better known to information users. However, the future lies in connecting online collections and sharing the load of cataloguing valuable online content, as even the largest institutions are unable to keep up with the massive amount of digital content being published.

## Policy Online

One online collection of social science grey literature trying to address some of these issues is Australian Policy Online ([apo.org.au](http://apo.org.au)). Established in 2002 by the Institute for Social Research at Swinburne University of Technology in Melbourne, the database now provides open access to over 22,000 titles from government, academic, NGO and professional research organisations, together with around 8000 policy-oriented journal articles.

As an open access digital library, Policy Online is actively dealing with the many issues presented by grey literature on a daily basis, as well as hoping to influence and change some of the more systemic problems that exist around production and collection.

### Collecting full text

While link rot is a general problem, in the world of public policy it has particularly devastating cycles and effects. Like many clearinghouses and online information services, Policy Online was developed as a bibliographic database, cataloguing and linking to external resources. Yet like many other digital collections relying on links to resources on the live web, link rot has been slowly devastating the value of the database and spreading its harm to other systems harvesting APO's metadata, such as Trove and World Cat. As a result, Policy Online has been changing over the last few years to become a full text digital library or repository for policy and practice grey literature. It is now estimated that 30 - 50% of APO's content is full text, and a small team of dedicated volunteers must be thanked for helping to fix links and upload text on older catalogue entries. This is a time consuming and complicated exercise for many reasons and is still very much a work in progress, but it is essential for APO's long-term effectiveness and functioning as a digital library.

One reason this has been a challenge is that Policy Online is built using the Drupal open source content management system rather than repository software. Drupal has many advantages over repository software: it is far more flexible in presentation and architecture, allowing for more customisation of the user interface and functionality while also having a range of modules that can be adapted to carry out the requirements of a repository. However, as it is not designed for this role, it requires considerable thought in terms of the architecture for full text hosting and document management compared to purpose-built software such as [DSpace](#), [ePrints](#) or [Islandora](#) (a combination of Fedora and Drupal). This has therefore required a considerable investment in time and effort by the APO team over the last few years, and while this is an ongoing journey, the architecture and functionality are now in place, along with greater expertise on the requirements for full-text file management.

A second issue in becoming a full-text collection is that copyright legislation in Australia means it is necessary to have permission to upload full-text resources. Creative Commons has been a great benefit in overcoming this; however, its use is still not widespread. Even governments that have declared their commitment to using Creative Commons often fail to implement it on their documents. One solution to this has therefore been to encourage user contributions in uploading content to the APO database. Contributors are encouraged to upload content to APO for free hosting, with listing moderated by APO curators, to ensure a level of quality control. As well as having their content being hosted, APO provides producing organisations

with various other benefits, such as Digital Object Identifiers (DOIs) for documents and data hosted in the repository, metrics such as downloads, shares and abstract page views, and pages for organisations and authors featuring the content hosted on the site

In order to combat link rot, the Digital Object Identifier (DOI), based on the Handle identifier system, was introduced to persistently identify journal articles. This uses DOI resolver software that keeps an up-to-date URL location for the DOI object, even when the article has changed web location. A similar approach is also being used for research data, which is increasingly regarded as an integral part of the scholarly record. Thanks to an Australian National Data Service (ANDS) Major Open Collections grant, APO is now able to mint DOIs for full-text hosted grey literature and data. This is important for many producers, as DOIs are seen not only as a way of managing resources, but also as a marker of symbolic capital in the world of scholarly publishing. This is as important for grey literature as it is for data in being acknowledged as a legitimate source of scholarly research. For more information on the policies and implementation issues of DOIs on APO, see the paper by Jaye Weatherburn in the VALA2016 Conference papers.

Another way this open contributor model works is for other collection services, organisations or researchers to develop their own collections using APO as a platform. This is occurring through partnerships with universities and organisations such as:

- the News and Media Research Centre at the University of Canberra, which is employing a staff member at the university to work on the Media and Communications collection;
- the Henry Halloran Trust at the University of Sydney, which is supporting an Urban Planning collection;
- the Australian and New Zealand School of Government, which supports the New Zealand Policy Collection and fortnightly newsletter; and most recently
- the University of South Australia, which is supporting a collection on Muslim and Non-Muslim understanding.

APO collections offer the benefits of a curated full-text collection without the expense of setting up and maintaining multiple individual repositories. The duplication of bibliographic databases is expensive and unsustainable with many databases closing after a few years – most recently the Homelessness Clearinghouse and the Australian Youth Clearinghouse. APO collections are designed to try to reduce costs and facilitate collaborative collecting.

## **Taxonomies and linked data**

Linked Data is an approach to publishing data. The use of this approach enhances the utility of information on the web by making references to persons, places, things, etc. more consistent and linkable, and therefore discoverable and able to be analysed within and across information systems. Key elements are the need for structured data, standardised classification systems or ontologies, and an understanding of the use cases, that is, the kinds of relationships and connections researchers and others need to undertake their work.

With support from an Australian Research Council Linkage Infrastructure and Equipment grant in 2013, APO has been investigating ways of implementing linked

data technologies. The process began with the implementation of internationally recognised interoperable taxonomies for topics and coverage terms. OCLC's Faceted Application of Subject Headings (FAST) taxonomy was chosen as a first step along the pathway of international interoperability. The FAST classification scheme is derived from the Library of Congress Subject Headings (LCSH) and makes the LCSH vocabulary available as a post-coordinate system designed for the web environment (The full vocabulary, which has over 1.7 million terms can be searched at <http://experimental.worldcat.org/fast/>). Terms tend to be single concepts that can be joined together in a faceted search system; for example, the terms internet and health can be applied by users to find any items with both terms (existing either as keywords or as taxonomy terms), without having had to apply these terms as a single concept by the cataloguer.

The advantage of this system is that it can be as simple or complex as required and if other collections use the same terms content can easily be gathered together. An example of this is the harvesting of policy-oriented journal articles from RMIT Publishing's Informit databases into APO. As Informit also uses the FAST taxonomy, all content on APO, both the grey literature reports and the Informit journal articles, is able to be discovered using the taxonomy system. Similarly, APO is now interoperable with WorldCat, OCLC's global library database, which harvests APO metadata using OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).

FAST has developed its own geographic terms for coverage, but is also linking to [Geonames](#), which provides the largest taxonomy of geographic terms and locations, allowing for links and visualisations using geospatial coordinates. A prototype visualisation is available at [apo.org.au/map](http://apo.org.au/map), with more sophisticated visualisations in development.

Implementing a new taxonomy system has involved a major reconfiguration of fields and navigation and extensive data cleaning, resulting in a significantly more flexible, and extensible database. This is an ongoing process, however, and using a large-scale classification system designed for a print-based library does not always mesh with the demands of an online collection classifying digital publications on the latest public policy issues. APO therefore still needs to have a tag field for local terms or terms that are not in FAST. (The omissions are often surprising – FAST does not have a term for public policy!) Staff at APO are working with the small FAST team at OCLC to contribute terms into the taxonomy and it would be great to see a global working group set up to work with the FAST.

Wikipedia is another intermediary source being used as a global reference point for terms in the APO database. Policy Online now has around 4000 producing organisations listed in the database and as an aid to evaluating grey literature sources the website includes embedded information from Wikipedia on listed organisations if it is available. For example, the Australian National Data Service page on APO lists various publications produced by this organisation, and embedded on the bottom right hand corner are details from Wikipedia about the organisation, helping orient readers and searchers to the context of the publisher (<http://apo.org.au/source/australian-national-data-service>). This allows APO to provide background information without additional resources. Linking subject terms to Wikipedia entries is another way to use this global resource and is something that

OCLC has been exploring recently. In this way, scope notes could be provided for terms used by APO, OCLC and other FAST users, as well as researchers.

As the size of the Policy Online collection has grown - it now has around 30,000 resource listings, 12,000 authors and 4000 organisations - it is now possible to see the database itself as a source of data for analysis as well as a repository of resources. Extending the metadata collected on organisations, including for example location, period of operations, personnel etc. and linking organisations to the topics on which they publish, is the next stage in providing data and visualisations of the interests and activities of policy organisations. Some examples of this kind of work and the visualisations generated can be seen on the Mapping Global Media Policy online database (*Mapping Global Media Policy* 2015; for an explanation of the research aims of this project see Raboy & Padovani 2010). Another interesting public policy tool that has inspired the next phase for APO's development as a data source is Bryan Jones and Frank Baumgartner's Policy Agendas Project (Baumgartner & Jones 2015). In October 2015, APO heard that it was successful in securing an Australian Research Council grant to continue its linked data work in 2016 and to develop these ideas of extending the utility of the database beyond being a digital library of documents and data to becoming a source of data in its own right.

## Conclusion

Open information infrastructures for scholarship, consisting of linked open data, open-source software tools, and a community committed to sustainability, are essential to meet the needs of scholars today. Finding and accessing public policy-related grey literature and data, understanding the relationships between organisations and the positions they advocate, mapping these over time and space, are essential for conducting research on increasingly globalised policy issues. Policy researchers spend large amounts of time searching, evaluating and collating resources. Systematic reviews, analysis of temporal and spatial elements, text mining of historical documents are all extremely difficult if not impossible without a large corpus of policy documents or data. There are currently few ways of monitoring research activity, trends and policy positions across lobby groups, NGOs, research networks and government agencies. The rise of data-driven science, and more recently social science, has changed the way research is conducted and disseminated. To date, traditional publishing and digital library collections have mainly focused on standard textual publications and bibliographic information. Access to timely research data and grey literature is becoming just as important particularly for the policy-related science and social science research dissemination and discovery and it is essential that libraries and information services work together to make this material more accessible and usable and to find ways to make digital collections sustainable.

## References

Baumgartner, F & Jones, B 2015, *Policy Agendas Project*, University of Texas, viewed 10 Nov 2015, <http://www.policyagendas.org/>.

Briggs, A & Burke, P 2009, *A social history of the media: from Gutenberg to the internet*, 3rd edn., Polity Press, Cambridge, UK.

Houghton, J, Rasmussen, B, Sheehan, P, Oppenheim, C, Morris, A, Creaser, C, Greenwood, H, Summers, M & Gourlay, A 2009, *Economic implications of alternative scholarly publishing models: exploring the costs and benefits*, JISC, London, viewed 13 September 2015, <http://apo.org.au/node/58411>.

Houghton, J, Steele, C & Sheehan, P 2006, *Research communication costs in Australia, emerging opportunities and benefits*, Department of Education Employment and Workplace Relations, Melbourne, viewed 10 Nov 2015, <http://apo.org.au/node/1811>.

Klein, M, Van de Sompel, H, Sanderson, R & Shankar, H 2014, 'Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot', *PlosOne*, vol. 9(12), [Doi.org/10.1371/journal.pone.0115253](https://doi.org/10.1371/journal.pone.0115253).

Lawrence, A, Houghton, J, Thomas, J & Weldon, P 2014, *Where is the evidence: realising the value of grey literature for public policy and practice*, Swinburne Institute for Social Research, viewed 10 Nov 2015, <http://apo.org.au/node/42299>.

Luzi, D 2000, 'Trends and evolution in the development of grey literature: a review', *The International Journal on Grey Literature*, vol. 1, no. 3, pp. 106-116.

*Mapping Global Media Policy*, 2015, International Association for Media and Communication Research (IAMCR), viewed 10 Nov 2015, <http://www.globalmediapolicy.net/>.

Meyer, ET, Thomas, A & Schoeder, R 2011, *Web archives: the future*, International Internet Preservation Consortium, <http://apo.org.au/node/58425>.

Newbold, E & Grimshaw, J 2010, 'Collection building with special regards to report literature', in J Schopf and DJ Farace (eds), *Grey literature in library and information management*, Gruyter, Berlin.

Raboy, M & Padovani, C 2010, 'Mapping global media policy: concepts, frameworks, methods', *Communication, Culture and Critique* vol. 3, pp.150–169. [doi:10.1111/j.1753-9137.2010.01064.x](https://doi.org/10.1111/j.1753-9137.2010.01064.x).

Riley, H & Crookston, M 2015, *Awareness and use of the New Zealand web archive: a survey of New Zealand academics*, National Library of New Zealand, viewed 10 Nov 2015, <http://apo.org.au/node/58430>.

Tenopir, C, Allard, S, Bates, B, J., LK, King, DW, Birch, B, Mays, R & Caldwell, C 2010, *Research publication characteristics and their relative values*, Publishing Research Consortium, UK, viewed 25 April 2015, <http://apo.org.au/node/55998>.