Australian Academic & Research Libraries, 2015
Vol. 46, No. 4, 229–248, http://dx.doi.org/10.1080/00048623.2015.1081712

Routledge
Taylor & Francis Group

Collecting the Evidence: Improving Access to Grey Literature and Data for Public Policy and Practice

Amanda Lawrence[a]*, Julian Thomas[a], John Houghton[b] and Paul Weldon[c]

[a]*Swinburne Institute for Social Research, Swinburne University of Technology, Hawthorn, Australia;* [b]*Victoria Institute of Strategic Economic Studies, Victoria University, Melbourne, Australia;* [c]*Australian Council for Educational Research, Camberwell, Australia*

The internet has profoundly changed how we produce, use and collect research and information for public policy and practice, with grey literature and data playing an increasingly important role. Reports, discussion papers, briefings and many other resources produced and published by organisations, without recourse to the commercial or scholarly publishing industry, are a key part of the evidence used for public policy and practice. Yet finding and accessing this material can be a time-consuming task made harder by poor production and management of resources and the lack of digital collecting services. Even knowing what is being collected and what collections exist is a difficult task. Based on research conducted as part of the Grey Literature Strategies ARC Linkage project, this article reports on the results of online surveys of users, producers and collectors of policy and research information with a particular focus on the results for collecting services. It discusses the state of collecting digital grey literature in Australia and the issues that need to be addressed to maximise the value of this public asset.

Keywords: digital curation; digital libraries; grey literature; public policy; access to knowledge; evidence-based policy

Introduction

Public policy work increasingly relies on a wide range of resources – some are traditional scholarly publications, but the majority are 'grey literature'. Reports, discussion papers, briefings, reviews and data-sets produced by government, academic centres, NGOs, think tanks and companies are heavily used and highly valued in policy and practice work, and form a key part of the evidence base. 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? To what extent is digital grey literature being collected and managed? What is the impact of not managing the digital resources that policy-makers and practitioners rely on for guidance and decision-making?

This paper draws on the findings of the Grey Literature Strategies 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. The researchers involved come from media and communications, economics, information management and education disciplines, a reflection of the interdisciplinary nature of the research problem. The aim of the project is to investigate

*Corresponding author. Email: alawrence@swin.edu.au

grey literature's role and importance in public policy and to find ways to enhance its value. The project considers the issues from the perspective of users, producers and collectors of grey literature. This article focuses on the role of collectors with data from a survey of collecting organisations helping to provide an understanding of what is currently being collected and how we might improve the efficiency and stability of digital collecting into the future.

Context

Grey literature is a term that often seems to obscure more than it illuminates and defining it is notoriously difficult with hard boundaries almost impossible to draw. A commonly cited definition is that grey literature is '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', defined at the Grey Literature Conference, Luxembourg, 1997 – expanded in New York, 2004.

To clarify this further, grey literature can be understood by focusing on three factors: the nature of the documents concerned, the types of producers and the means of dissemination. Grey literature features in many disciplines and professions as diverse as engineering, archaeology, and health and takes on particular characteristics depending on the context. In terms of public policy, grey literature document types include research and project reports, working papers, discussion papers, briefings and guides, data-sets, conference papers and various others, produced by organisations such as government departments and agencies, universities, think tanks, non-government organisations, corporations and professional bodies, and made available to the public directly either in print or digitally. Grey literature stands in contrast to content produced and sold by publishing companies which are generally part of a managed system of production and dissemination resulting in large-scale collection and preservation by companies and libraries. This is sometimes referred to as white literature (Feather & Sturges, 2003, p. 211) and other times as black literature, (PHCRIS, 2014) or in the scholarly community 'traditional publications' and includes books, journals, magazines, newspapers and certain conference proceedings. Table 1 provides an illustration of some document types of relevance to the public policy area. This should be considered only as a general guide to help clarify the concept rather than hard and fast rules of taxonomy as the boundaries very quickly blur and there are many exceptions to the rule.

Despite the lack of clarity around the term, grey literature still plays a useful role as a concept that draws together a range of documents types and resources that are

Table 1. General categories of white, grey and black literature in public policy.

Published (white) literature	Grey literature (grey) produced by organisations	Other (black)
Books	Reports	Emails and letters
Book chapters	Discussion papers	Personal notes
Journals	Briefings and guides	Minutes
Journal articles	Data-sets	Conversations
Magazines	Conference papers	Ephemera
Newspapers	Submissions	
	Evaluations	
	Working papers	
	Blogs	

otherwise often invisible within the larger discourses of publishing, scholarly communication and the open-access movement. In this way, the concept of grey literature brings together a range of disparate sectoral interests, publishing practices and information uses and provides a unique lens to explore how these operate within and inform the field of policy and practice.

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 isn't distributed through established channels and it doesn't 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. It could be based in an existing library, but was often no more than a room or part of an office building with some space marked out for pamphlets, reports, working papers, information sheets and other resources that were otherwise hard to find. Many organisations and libraries still have shelves of valuable print grey literature, with many older items in need of digitisation.

Print grey literature has also had the benefit of legal deposit legislation in many countries. In Australia, this requires producers to provide copies of their printed published material (anything made public) to the National Library and a state library in their jurisdiction. Copyright is also often not a major issue with print copies, as their use and distribution is limited and they are generally not copied by the collecting organisation. However, many educational libraries do still have to pay a fee to the Copyright Agency Limited if freely available, non-commercial publications are copied by patrons.

The internet has changed almost every aspect of the research publishing chain – the way research and information can be produced, disseminated, discovered, used, managed and collected. The huge amount of information and research published online provides unprecedented access to knowledge, from a wide range of sources, enabling a much greater level of understanding and participation in public interest issues. It also brings a number of challenges: searching, sifting, evaluating and accessing information and research are time-consuming and often frustrating tasks occupying a large portion of the day for those engaged in policy work.

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 policy discussion and debate. Online publication provides infinite flexibility in terms of format, content and style, they 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.

The Internet has also brought about a shift in the historic roles of publisher, researcher, distributor, collector and news service which are converging and transforming. The scale and rate of change has been hard to predict or adapt to, and many of our institutions, infrastructures, practices and legal frameworks are still stuck in print-based paradigms. These issues compromise some of the benefits of digital technologies and present a host of new challenges.

A key issue is the need for new ways of managing and curating digital resources of public importance. Recent investigations, such as the UK Finch review on improving access to research publications (Finch, 2012) and the US Blue Ribbon Task Force on

digital preservation (2010), share our concerns and recommend improvements in the way grey literature is managed and collected in order to maximise the benefits of publicly funded research.

> Clarification of the long-term value of emerging genres of digital scholarship, such as academic blogs and grey literature, is a high priority. Research and education institutions, professional societies, publishers, libraries, and scholars all have leading roles to play in creating sustainable preservation strategies for the materials that are valuable to them. (Blue Ribbon Task Force, 2010, p. 3)

It is in this context that we began our investigations to establish an understanding of the role of grey literature for public policy and practice.

Research methods

A range of methods are being used for this research, 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 research also features methods drawn from various disciplines including that of economics, based on Houghton's work on measuring the value of open-access publishing and data repositories (see e.g. Houghton & Beagrie, 2012; Houghton et al., 2009). The main sources of data which we are drawing on for this article are the online surveys. Forthcoming publications will include more details on other research results and the data will also be made public in due course.

Three separate but related surveys were conducted targeting individual users, collecting organisations and producing organisations. The surveys were all titled: information and research for policy and practice: a survey of access and use/producing organisations/collecting organisations and services. Grey literature as a term is not widely known or understood outside certain contexts (and even then the adoption is patchy) so it was important that a lack of familiarity with the terminology did not deter suitable respondents.

Another reason for a broad title was to situate the use of grey literature within the context of information seeking for public policy and professional practice. Various studies have been conducted looking at the ways in which research is accessed and used, particularly by those in the public sector and academia, and these commonly refer to a range of information resources including traditional scholarly outputs such as journal articles and books, grey literature content such as government or technical reports, the media and online commentary, and the role of colleagues and personal networks (see e.g. Cherney, 2012; Nutley, Walter, & Davies, 2007; Tenopir et al., 2010). We are therefore keen to be able to connect our research to these other studies where possible. This is not as easy as might be expected as many different terms are used for similar document types making comparisons across surveys somewhat difficult (e.g. reports may be referred to as technical reports, government reports, project reports, unpublished reports etc.).

Research population

Having reviewed a range of other surveys and instruments, the questionnaires were developed by the project team using interactive forms and web pages that could be

tested on representatives from partner organisations as well as others in education, government and NGOs. Each questionnaire asked respondents about themselves or their organisations to establish basic demographic information and other details important for various context and economic analyses (e.g. the number of staff employed by producing and collecting organisations or units). They were then asked to consider a fairly long list of resource types (approximately 20 items) and asked to indicate what materials they used, produced and collected for their work, how important these were for their work or their organisation, how they found and accessed the material they used. Following this a series of questions looked at how and why material was produced, methods of finding and accessing resources, what challenges were experienced with discovery and access, deadlinks (i.e. Internet links that no longer work), how they viewed various copyright issues, and to estimate the economic value of their use of grey literature – that is willingness to pay and willingness to accept contingent valuation questions. These were then able to be combined with other responses such as hours spent using grey literature, or number of staff hours spent collecting grey literature, to provide estimates of economic value at a national annual level. Where required, the term grey literature was used in some questions and was defined at the time of first use. Wherever possible questions were repeated in all three surveys, allowing for comparison across the different groups. In addition each survey had some questions uniquely relevant to its particular audience. All surveys are available online at the Grey Literature Strategies website (Lawrence, Houghton, Thomas, & Weldon, 2013).

Relatively little is known about how grey literature is used and produced in the field of policy and practice and given the complexity of the information environment, there is a need for an interdisciplinary intersectoral perspective (MacDonald, De Santo, Quigley, Soomai, & Wells, 2013). We therefore decided not to limit the research to a specific group, such as public sector workers or academic scholars, as is the case with many other studies on research impact. Our aim was instead to identify and describe the ways in which grey literature is used, produced and collected across the range of participants in policy and practice work. The population is therefore a convenience sample and it is difficult to know exactly what the overall population is for grey literature users, producers or even collectors would be.

Despite this, our research aims to estimate national costs and values, which rest on estimates of the relevant populations. These have been derived in the following ways. The estimate of the user population is based on an analysis of ABS labour force data by occupation and industry, and it is estimated that at least one-third of the labour force could be grey literature users. For example, there is a total of almost 4 million employed in the Information Media and Telecommunications, Financial and Insurance Services, Rental, Hiring and Real Estate Services, Professional, Scientific and Technical Services, Administrative and Support Services, Public Administration and Safety, and Education and Training industries. Similarly, by occupation there are around 4 million Managers and Professionals (Australian Bureau of Statistics [ABS], 2013a, 2013b). Hence, for the purposes of estimation we assume that one-third of the workforce or 3.8 million people use grey literature, to some extent, in their work.

For the policy-orientated grey literature producer population, we base our estimates on ABS data for non-profit organisations and businesses, listings of government departments and agencies, and other estimates of consultancies to arrive at an estimated producer population of at least 30,000 organisations, departments and companies.

The National Library's Australian Libraries Gateway reported a total of 5346 libraries and collecting organisations at the end of 2014 (excluding mobile, music and

toy libraries) (https://www.nla.gov.au/libraries). Added to this number are an estimated 60–70 digital libraries or databases not included in the Australian Libraries Gateway database, documented as part of this research. Hence, for the purposes of estimation we assume that there are a total of 5400 collecting organisations in Australia.

The questionnaire was set up and run by the Australian Council for Educational Research using Lime survey open-source software. The three online surveys ran for around three weeks in 2013 and were conducted using a convenience sample drawn from a targeted purposive audience of those engaged in public policy and professional practice work. Collecting and producing organisations were emailed directly based on addresses collated from publicly available sources. In addition all surveys relied on attracting participants via various dissemination methods and platforms where the target respondents were likely to be an audience. These included: Australian Policy Online (apo.org.au) newsletter and website; email newsletters produced by a range of organisations across various sectors and interests such as the Australian Council for Educational Research, PHCRIS, Eidos, National State Libraries Australasia, ALIA and many others; and social media, particularly Twitter and LinkedIn Groups on grey literature, public policy and other areas.

Respondents

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 coherent picture.

For the survey of users, government sector workers were the largest group, being nearly half of all respondents (44%), followed by those working in education (26%) and the NGO 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 intended to be answered at an organisation or a department level. 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 meant to be answered at an organisation or department level. 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 provides a summary of the respondents for each survey by sector.

The subject areas of interest for users, producers and collectors have been analysed based on ABS Socio-economic objective (SEO) classifications for research and development (Australian Bureau of Statistics [ABS], 2008). Respondents to all surveys had similar results with a third to a half being from Law, Politics and Community Services followed by Health, Economic development, Environment, Education and training and a small number from Cultural understanding and Defence, indicating the diversity of subject areas with a public policy dimension. A large portion of collectors were general collections and not able to be classified into a specific SEO.

Collectors generally came from either academic/research library (25%), special libraries (29%) or government libraries (21%) (collecting services based in the government sector may describe themselves as either a special library or a government library

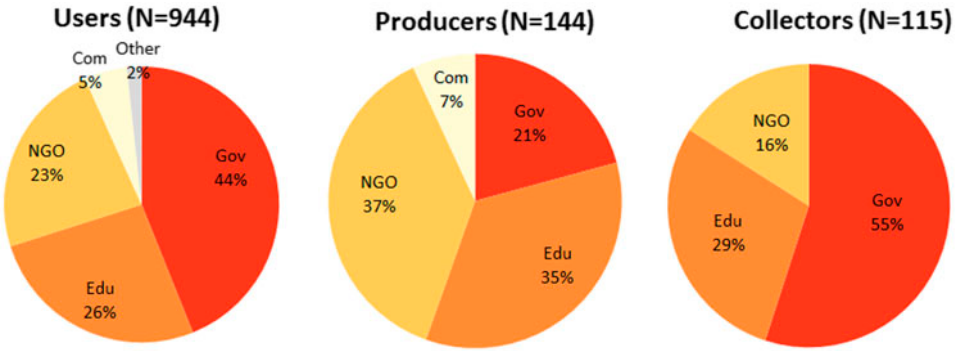

Figure 1. Survey respondents for the user, producer and collector surveys, by sector.

which explains the difference between these answers and the sector response). A further 20% came from either an institutional or a subject repository or collection. Collecting services were overwhelmingly small with 10 staff or less for 77% of respondents however there were also some very large collecting organisations responding.

Results and discussion

Access and use of research and information for policy and practice

While the focus of this paper is on the findings for the collection of grey literature, we will briefly summarise some of the key results on use and production in order to provide a context for the results from collecting organisations.

As already discussed, grey literature is difficult to define and it can be more useful to look at specific formats. The surveys therefore began not by asking respondents to indicate if they used grey literature, but rather to choose from a range of content types, both white and grey, and to indicate, 'all the types of materials you occasionally or regularly use in your work'. 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.

Respondents were then asked to indicate, from the list of resources that they used for their work, how important they consider each type. 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 are provided in Table 2 ordered by the material that is most used. The figures for importance are based on the whole sample and include those that don't use a type of material. If we only look at the importance of materials for those that use them (the right hand column), we get a slightly different picture. Reports and data-sets are important for over 90% of respondents, followed by evaluations (89%), journal articles (88%), discussion papers (86%), briefings, guides and reviews (84%) and technical reports (83%). Clearly then grey literature document types are a very important part of the information resources used by those working in policy and practice.

Table 2. Materials used by *Access and Use* survey respondents for work and their importance.

Materials used	Used (%)	Important/very important overall (%)	Important/very important for those that use (%)
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

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 for 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, we estimate 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–$43 billion per annum. (Further details on the economic valuations will be published in the coming year.)

Production of research and information for policy and practice

Grey literature is a key method used by surveyed organisations across all sectors of society to translate and disseminate new research or policy positions. The Producer questionnaire asked: ‘Why does your organisation produce its own research and

information materials (rather than publish with a journal or book publisher)?' – a question which defines grey literature rather than using the term. 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%).

As for the user survey, producers were asked to indicate from a list of materials what they produce and how important is it for their work? Conference papers are the most common type of document, produced by 82% of organisations, followed by discussion papers (77%), reports (76%), briefings/reviews (67%) and submissions (63%). This does not directly correlate with importance, as reports (93%), submissions (91%), discussion papers (89%), briefings/reviews (89%) and evaluations (89%) rate as the most important or very important material produced. Conference papers, which are produced the most, are only important for 69% of producers. It may be that conference papers are produced more as a requirement of participating in conferences rather than as an output in their own right. This is supported by the figures showing conferences are seen as an important dissemination method for 75% of producers.

In terms of estimating the economic value of grey literature production, as indicated we estimate that at least 30,000 organisations in Australia produce grey literature. Australian respondents reported their organisation 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 seems plausible.

Discovery and access to information and research

Policy information seekers follow many paths as they search for and discover resources. As the number of potential access points and information providers has multiplied and roles and services converge, it has become increasingly 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, a whole new range of digital curation and information services have emerged, and producing organisations have taken on some of these activities themselves.

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 – is 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 (deadlinks) were also mentioned.

Collecting information and research for policy and practice

Having established that grey literature is heavily used and highly valued as a source of research and information for those working in policy and practice, we will explore in detail the results for collecting services and organisations. As per the producer and user surveys, a list of content types was provided in the questionnaire for collecting organisations to select. From the results, we see that journal articles and books are the most commonly collected type of material with 90% of surveyed services collecting them (Table 3). This is 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. For example, briefings and guides are collected by 78% of NGOs but just over half of government and education sector collections. The opposite is true for working papers which are collected by 85% of education services whereas only around half of other sectors do so. NGOs are also working harder to collect submissions (61%) and evaluations (56%) compared to government collections (submissions 38%, evaluations 22%) and education services (submissions 33%, evaluations 30%), reflecting the importance of a range of formats for these groups and their audiences.

Table 3. Materials collected by collecting organisations and services – by sector.

Materials collected ($N = 114$)	Gov (%)	Edu (%)	NGO (%)	All (%)
Journal articles (peer reviewed)	87	94	89	90
Books and eBooks	91	94	78	90
Reports	86	85	100	88
Conference papers	73	91	89	81
Audio video material	78	85	67	78
Discussion papers	67	85	83	75
Essays and articles	64	76	78	69
Policies procedures	79	42	83	69
Professional/trade mag	68	73	67	69
Archival or heritage material	76	64	44	68
Briefings guides/reviews	57	55	78	60
Working papers/preprints	48	85	50	59
Theses	57	73	28	57
Book chapters	41	82	44	54
News reports	52	42	61	51
Technical documents	49	46	50	48
Web pages/websites	48	49	50	48
Data sets	35	52	44	41
Submissions	38	33	61	40
Information sheets	37	21	44	33
Evaluations	22	30	56	30
Legal documents	27	30	22	27
Commercial/market research	19	24	33	23
Blogs	14	9	28	15
Social media	11	12	17	12

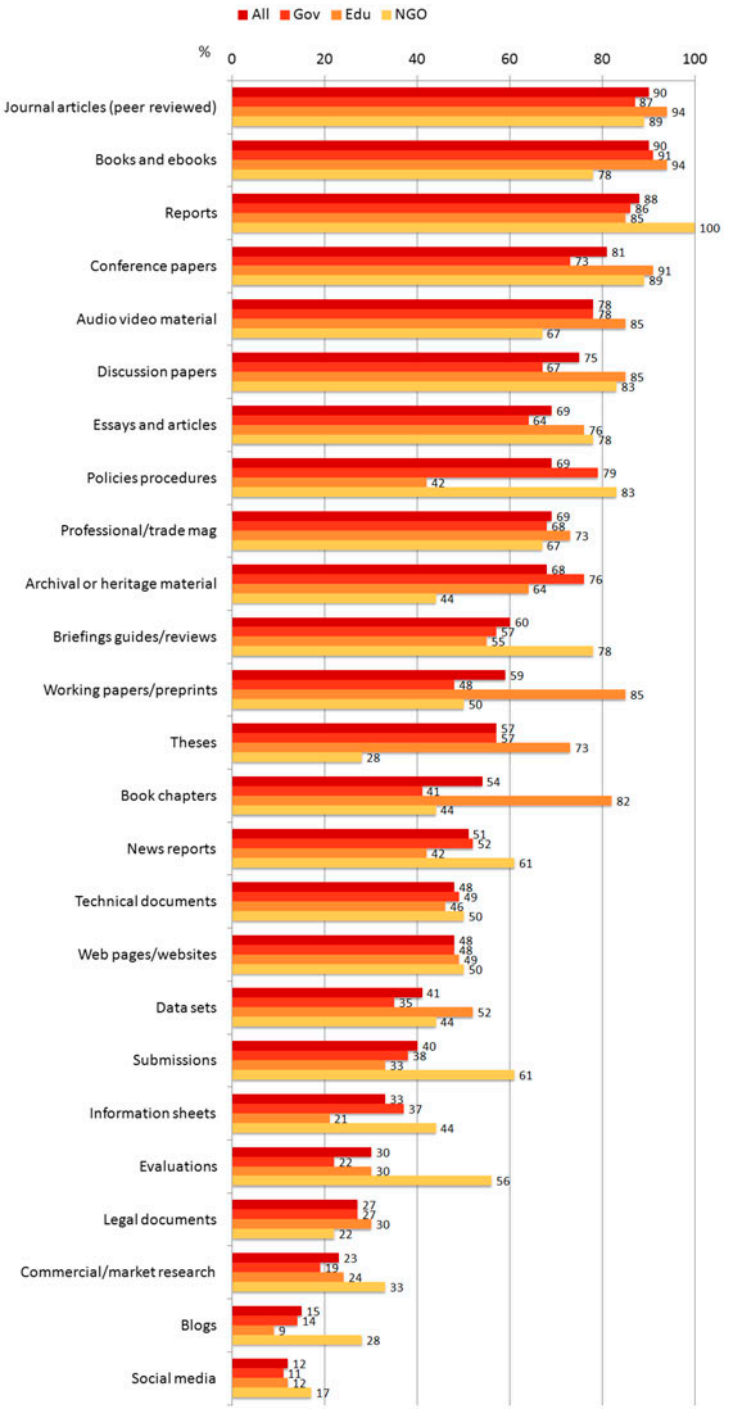

Figure 2. Materials collected by collecting organisations and services – by sector.

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 4).

A related issue of importance is how easily users and collectors are able to find content. The significance of this is that material that is important but hard to find for users should be a priority for collectors, while the degree of difficulty for collectors increases the time and cost involved, and therefore the likelihood of collecting. The surveys of users and collectors asked 'Of the materials that you use, how easy is it to find the information and research that you need for your work?' and the collector survey asked, 'How easy is it to acquire various types of items?'. Of the materials rated as the most important for users, a quarter had trouble finding reports (25%) and submissions (26%), a third said data was hard to find (31%) and a very large 44% reported that evaluations

Table 4. Importance of material collected – percent by sector.

Importance of materials collected	Gov (%)	Edu (%)	NGO (%)	All (%)
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

Table 5. Materials considered easy or hard to collect by collecting services.

Materials	Easy/very easy (%)	Not easy/not very easy (%)
Evaluations	31	69
Data sets	49	51
Archival material	49	51
Submissions	50	50
Working papers and preprints	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

are hard to find. Around 20% of policy workers responding find journal articles, professional magazines, discussion papers and briefings/guides hard to find.

Many resources that are considered important by users are not easy to find for collecting services (Table 5). 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.

Collections and collection size

Collecting organisations were asked 'What is the approximate overall size of your collection/database?' This was a very difficult to phrase for such a diverse range of collecting organisations and we were aware that it could be interpreted in many ways however it was essential to try to get some idea of the size of Australia's policy and practice collections. Responding collections are generally mid-range, two-thirds (63%) have more than 10,000 records and a fifth (20%) more than 100,000 records (Table 6 and Figure 3). Print collections are the largest, 17% of services have over 100,000 print items and around half (54%) have 10,000 print items or more. A third (29%) have less than 10,000 print items.

Table 6. Size of collections collecting organisations and services size.

Type of items in collection	Count	%
Records in your catalogue/database (N = 81)		
Don't know	7	9
Up to 999	10	12
1000–9999	13	16
10,000–99,999	35	43
100,000 or more	16	20
Print items held (N = 76)		
Don't know	12	16
None	7	9
Up to 999	5	7
1000–9999	10	13
10,000–99,999	28	37
100,000 or more	13	17
Digital documents held (N = 77)		
Don't know	15	19
None	8	10
Up to 999	14	18
1000–9999	21	27
10,000–99,999	15	19
100,000 or more	4	5
Audio/video items held (N = 72)		
Don't know	11	15
None	8	11
<1000	35	49
1000–9999	8	11
10,000–99,999	6	8
100,000 or more	3	4
Web pages websites archived (N = 69)		
Don't know	13	19
None	34	49
<1000	16	23
1000–9999	3	4
10,000–99,999	2	3
100,000 or more	1	1
Datasets held (N = 65)		
Don't know	13	20
None	30	46
<1000	16	25
1000–9999	6	9

In comparison only 5% of collections surveyed (four organisations) hold over 100,000 digital items and only a fifth (19%) have between 10,000 and up to 100,000 items. Just over a quarter (27%) have between 1000 and up to 10,000 digital items and a similar amount (28%) are very small, with less than 1000 digital documents. In other words 75% of collecting organisations responding have less than 10,000 digital items in their collections or don't know what they have. The collection of web pages and data-sets is considerably lower with almost no large collections.

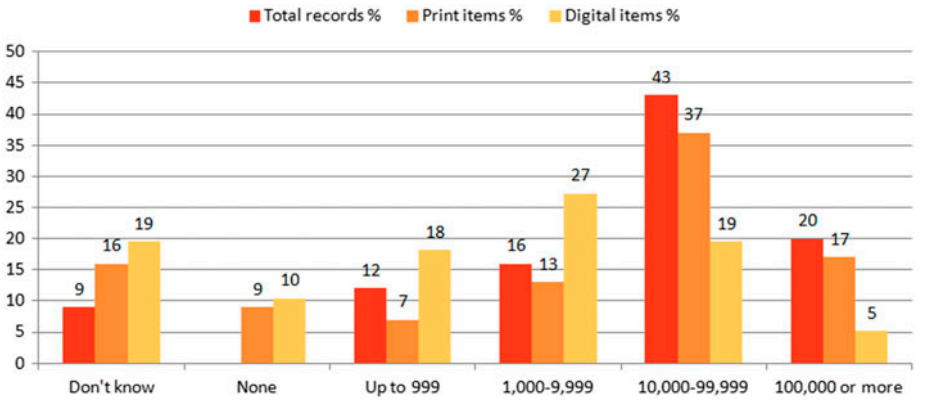

Figure 3. Collection size of responding organisations.

So although a great deal of digital content for policy and practice is made available online for free, (unlike print materials which are generally costly to acquire), it would seem from these figures that digital collecting at scale is not yet occurring in Australia. Another issue of concern is the inability of a sizable number of collecting organisations to audit and report on their collections with any accuracy. Nine per cent of collecting services were unable to say how many records they have, 16% how many print items and a fifth (19%) could not determine how many digital items they hold. Many others provided a figure but indicated in their comments that it was an estimate. It seems astonishing that in the age of databases and computers, many organisations are unable to readily count the digital content held in their collections. It may well be that many of those who did not respond to this question were actually unable to answer them. The reasons for this lie partly in the nature of legacy collections, but also in the lack of specific metadata that can assist with auditing collections and inadequate software and infrastructure. Without accurate figures, it is difficult to prepare or advocate for the kinds of investment required for effective information asset management.

Digital infrastructure not designed to store copies of digital content is one reason for the surprisingly slow adoption of digital collecting after 25 years of the Internet. While most people can easily download a report in common formats such as PDF or Microsoft Word and store it on a hard drive, collecting and providing access to full-text digital content requires specialist software. This is not complex or necessarily expensive (there are open source solutions), but it does require planning, set-up costs, and ongoing maintenance and management. There is clearly considerable activity in this space, with 27% of collection services reporting that they are developing digital infrastructure at the time of the survey, however this may also result in increased duplication rather than collaboration.

Rather than holding full text, many digital libraries and databases catalogue content, then link to where it is located online. This is generally a necessity given Australian copyright laws do not allow for the collection of digital content without permission of the copyright holder – an often time consuming and complex requirement. The effect of current Australian copyright legislation is that it limits the collection and preservation of all copyrighted material, including orphan works (where the copyright owner is unknown or no longer exists), 'unpublished' material and non-commercial, publicly

funded research and information produced in the public interest. This comes at a high price as content is lost from online access over time or is costly and time consuming to collect with the required permissions.

This has a number of flow-on effects. One is that it makes it particularly difficult to determine the amount of digital content that is actually collected as many collections have partial full text but what exactly this includes is often not even clear to the services themselves. It also causes major issues with 'link rot' or deadlinks – the loss of access to content previously available online caused by the removal or moving of content from its original URL. Link rot is a regular experience for most policy information users, with around half of user survey respondents reporting being unable to access an online resource because of a deadlink on a weekly basis or more often.

Given the prevalence of linking in digital collections, this is also a major issue for collectors (and has a flow on effect for aggregators such as Trove, World Cat and Google). When asked, 62% of collectors reported that they are concerned or very concerned about deadlinks affecting their digital collections. Collecting services end up with the problem embedded within their databases with over two-thirds (69%) of respondents reporting that they have deadlinks in their collections. Half estimate that this is up to 20% of their collection. A quarter does not monitor their links meaning there could also be extensive issues in these collections given the ubiquity of the problem.

Addressing deadlinks is also a costly exercise with 43% of collecting services reporting that they manually fixing link rot in their collections on a regular basis. A number of respondents commented that their approach to deadlinks involves *irregular* manual fixing, indicating a more ad hoc approach to this insidious problem. On average, collections spend 1.3 h a week fixing links and 50% of services are engaged in doing this on a weekly basis.

Other possible strategies to combat deadlinks are not used a great deal. Those in the education sector were the most likely to use identifiers such as handles or DOIs (54%) compared to an overall take up of only 30%. Dark archives and using web archiving services such as Pandora or the Internet Archive were used by just over a quarter of services (27%). A comment made in relation to this question reveals a combination of traditional methods and a touching hopefulness in others:

> We try to collect as much hard copy as possible. Will continue to do so until it becomes too expensive or we run out of room. We also trust that most government publications will be placed on respective websites. Collecting organisation, Government sector.

As indicated above, print collecting is also expensive. And unfortunately, expecting government departments and agencies to manage their publications is, at this point, highly unlikely. They may be placed on websites but they don't necessarily stay there for the long term as government departments change names, responsibilities and websites with changes in government. Calculated at the national level, we estimate that the combined cost of deadlinks in Australia, for users and collectors, is at least AUD$5 billion a year.

A major issue for collecting organisations and services across all sectors is financial sustainability. When asked where the funding came from for their work, two-thirds of collections reported that the most important source of support is their own organisation. This was followed by funding agreements and contracts, which are important for half (50%) of collections. Given that most respondents came from government, education and NGOs, it is fair to say that public funds are crucial source of income for collections. This is something that needs to be taken into consideration when both establishing

new services and when a decision is made to close a service, to ensure the public investment made in the collection is able to be transferred to another service rather than simply lost. Since conducting this survey some of the collection services that responded no longer exist: many government libraries at state and federal level have been closed, as have other specialist services. Our concern is that without a coherent approach to information management of public interest information and research, we may easily lose large amounts of publicly funded resources, both the publications themselves and the metadata collections which are also of great value and should be considered part of the public information asset registry.

Reform of information management practices for public policy and practice

> The disappearance of reports etc. that form grey literature from online is like an erasure of history. There has to be a mechanism of maintaining continuity of major corporate documents even if they no longer have status within an organisation. Libraries are important for this but they are becoming reduced in capacity and capability within government agencies. (Access and Use survey, government sector)

In short, finding and accessing policy information is a time-consuming task made harder by the lack of large-scale collection services able to help locate relevant, high-quality resources quickly and efficiently. Information users were asked to 'estimate what percentage of your overall working time you could save if grey literature were more readily accessible and systematically preserved? Please provide an estimated average percentage of working time per week'. Users estimated that on average they could save 16% of their work time if grey literature was more readily accessible. This is equivalent to six hours a week based on an average working week of 35 h, or 288 h a year per person – a significant amount of time. When calculated at the national level, the efficiency impact of grey literature being more readily accessible and systematically preserved would be worth around $17 billion per annum nationwide. A very small fraction of this amount could be spent improving production standards and establishing large-scale national and international collaborative digital collecting infrastructure.

In order to do this we need to make changes in a number of key areas: copyright reform, improve standards of production and management, greater collaboration amongst collecting services, and new mechanisms of regulation and measurement that will support innovation and efficiency gains. These recommendations are articulated in more detail in a recent discussion paper (Lawrence, Houghton, Thomas, & Weldon, 2014).

In terms of copyright, all three surveys included a series of statements on copyright and collection practices and respondents were asked to indicate how strongly they agreed with them. Over 80% of users, producers and collecting services (87%) agreed or strongly agreed with the statement that: 'Libraries and information services should be able to store copies of print and digital material, including grey literature, for long term access and use'. Two-thirds (67%) of collecting services felt strongly that not being able to store copies of information due to copyright law was a problem for their collection. Copyright law is fairly clear to the majority of collectors (64%) however a third (36%) of respondents indicated that they were neutral or unclear about their right to copy or store materials indicating a fair degree of confusion around digital collecting.

Some of the comments that were made in relation to this question indicate the problems that collecting services face:

> It would be good if the government could indicate whether all content is covered by CC licences, or only some. (Collection survey, Education sector respondent)

> I would like to see more freedom to store information and less onerous copyright restrictions. We pay a lot to license online materials, but many publishers do not include archival access and backfiles are available at the publisher's discretion. (Collection survey, NGO sector respondent)

> As pressure to discard print material increases it is essential that digital archiving be allowed. (Collection survey, NGO sector respondent)

> Copyright – including orphan works needs to be addressed. Also copyright where our agency is not the sole copyright holder. (Collection survey, Government sector respondent)

> I have obtained permission to keep copies of certain material on our server for internal staff use but this does not assist the wider public with ongoing access. (Collection survey, Government sector respondent)

Collecting services were asked what approaches would be the most useful for improving collection and access to grey literature (Table 7) with the most highly rated options being: standard bibliographic information on publications (85%), fair use copyright law (83%), agreed metadata standards (82%) and collaborative collecting (80%). Over three quarters (78%) of collecting services identified improved software and infrastructure, and sharing metadata as useful or very useful.

Applying these strategies would create considerable efficiencies for both information collectors and users. They would also facilitate the development of new specialist services for particular domains, with the potential to add value to larger aggregated services, such as Trove or Google Scholar.

Investment now in quality metadata and full-text collecting will have long-term benefits, especially in the light of emerging research techniques. There is the potential to create systems that enhance sifting, evaluating and reusing resources. Collection and information services need to consider what new aggregated or linked systems could be developed using linked data and international standards for interoperability. The time is right to investigate and invest in efficient, interoperable metadata across the whole policy information ecosystem in Australia.

Table 7. Most useful approaches for improving collection and access to grey literature.

Approaches for improving collection and access to grey literature	Useful/very useful (%)
Standard bibliographic info	85
Fair use copyright law	83
Agreed metadata standards	82
Collaborative collecting	80
Improved software or infrastructure	78
Shared metadata	78
Legal deposit for digital content	72
Interoperable systems	66
OAI harvesting	46
Other	30

Conclusion

The field of public policy and practice relies on information and research produced by a wide range of organisations including government departments and agencies, academic centres, non-government organisations and commercial consultants and companies, and professional publishers. These make up a complex set of resources and production practices that operate on many different levels of formality, production standards, distribution systems, reward and measurement systems, use value and long term management. In the face of the abundance of research and information products that are now being produced and disseminated online, collecting services are struggling to adapt. Collecting services need to come together to find more efficient ways of managing this abundance. And crucially the role of long-term collection needs to be recognised by governments and other institutions if we are to ensure valuable information and research resources and databases, often paid for by public funds, do not disappear from online public access.

Disclosure statement

Amanda Lawrence is the Research and Strategy Manager for Australian Policy Online and Julian Thomas is Chair of the Advisory Group. Australian Policy Online is a free open-access digital collection of grey literature based at the Swinburne University of Technology.

Funding

This research has been supported by an Australian Research Council Linkage [grant number LP120100309] in partnership with the National Library of Australia, the National and State Libraries Australasia, the Australian Council for Educational Research and the Eidos Institute.

Notes on contributors

Amanda Lawrence is a research and strategy manager for the Grey Literature Strategies ARC Linkage project (greylitstrategies.info) and Australian Policy Online (apo.org.au) an open-access research database. Amanda has an arts degree with Honours from the University of Melbourne, a Graduate Diploma in Library and Information management from RMIT University and is currently completing a PhD in Media and Communications at Swinburne University of Technology. Amanda has published a number of articles on publishing in Australia and Asia, online collecting, grey literature and public policy and is a member of the Global Internet Policy Observatory (GIPO) Advisory Group.

Julian Thomas is the director of the Swinburne Institute for Social Research, and Professor of Media and Communications at Swinburne University of Technology. He is the chair of the management committee of the website Australian Policy Online, and is a member of the Consumer Consultative Forum of the Australian Media and Communications Authority, the Telstra CEO's Consumer Forum and the ARC College of Experts. He is also a member of the ARC Centre of Excellence in Creative Industries and Innovation (CCI). Julian has published extensively on media and internet policy and his most recent book is *The informal media economy* (2015) published by Polity.

John Houghton is a professorial fellow at Victoria University's Victoria Institute of Strategic Economic Studies (VISES). He has published and spoken widely on information technology, industry and science and technology policy issues, and he has been a regular consultant to national and international agencies, including the Paris-based Organisation for Economic Cooperation and Development.

ORCID

Amanda Lawrence http://orcid.org/0000-0003-2194-8178

References

Australian Bureau of Statistics. (2008). *Australian and New Zealand Standard Research Classification (ANZSRC): Socio-economic objective* (1297.0). Author. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/1297.0Main%20Features62008?opendocument&tabname=Summary&prodno=1297.0&issue=2008&num=&view=

Australian Bureau of Statistics. (2013a). *Labour force, Australia, detailed, quarterly*. (Cat No 6291.0.55.003). Author. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/35006594EB1126B2CA257132000F9189?opendocument

Australian Bureau of Statistics. (2013b). *Labour force, Australia, Sep 2013*. Author. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/6202.0Main+Features1Sep202013?OpenDocument

Blue Ribbon Task Force. (2010). *Sustainable economics for a digital planet: Ensuring long-term access to digital information*. Author. Retrieved from http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Cherney, A. (2012). *What influences the utilisation of economics research? The perspectives of academic researchers*. Institute for Social Science Research. Retrieved from http://www.issr.uq.edu.au/filething/get/15141/Economists%20RU_Project%20Working%20Paper%20July%202012.pdf

Feather, J., & Sturges, R. P. (Eds.). (2003). *International encyclopedia of information and library science* (2nd ed.). London: Routledge.

Finch, J. (2012). *Accessibility, sustainability, excellence: How to expand access to research publications*. London: Working Group on Expanding Access to Published Research Findings, UK Government. Retrieved from http://apo.org.au/files/Resource/Finch-Group-report-FINAL-VERSION.pdf

Houghton, J., & Beagrie, C. (2012). *Economic impact evaluation of the Economic and Social Data Service*. The Centre for Strategic Economic Studies (CSES). Retrieved from http://www.esrc.ac.uk/_images/ESDS_Economic_Impact_Evaluation_tcm8-22229.pdf

Houghton, J., Rasmussen, B., Sheehan, P., Oppenheim, C., Morris, A., Creaser, C., … Gourlay, A. (2009). *Economic implications of alternative scholarly publishing models: Exploring the costs and benefits*. Retrieved September 13, 2013, from http://www.jisc.ac.uk/media/documents/publications/rpteconomicoapublishing.pdf

Lawrence, A., Houghton, J., Thomas, J., & Weldon, P. (2013). *Survey questionaires*. Retrieved from http://greylitstrategies.info/project-publications/survey-questionaires

Lawrence, A., Houghton, J., Thomas, J., & Weldon, P. (2014). *Where is the evidence: Realising the value of grey literature for public policy and practice*. Hawthorn: Swinburne Institute for Social Research. Retrieved from http://apo.org.au/node/42299

Luzi, D. (2000). Trends and evolution in the development of grey literature: A review. *International Journal on Grey Literature, 1*, 106–117.

MacDonald, B. H., De Santo, E. M., Quigley, K., Soomai, S. S., & Wells, P. G. (2013). Tracking the influence of grey literature in public policy contexts: The necessity and benefits of interdisciplinary research. *Grey Journal (TGJ), 9*, 61–68. Retrieved from http://ezproxy.lib.swin.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=88313512&site=ehost-live&scope=site

Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using evidence: How research can inform public services*. Bristol: Policy Press.

PHCRIS. (2014). *Introduction to accessing the grey literature*. Retrieved July 15, 2015, from http://www.phcris.org.au/guides/grey_literature.php

Tenopir, C., Suzie Allard, Bates, B., Levine, K. J., King, D. W., Birch, B., … Caldwell, C. (2010). *Research publication characteristics and their relative values*. Publishing Research Consortium. Retrieved from http://www.publishingresearch.net/PRCTenopiretalWord2010ResearchPublicationCharacteristics.docx