

A bridge over grey waters: collecting and preserving public policy research

Amanda Lawrence

Abstract

This paper looks at the role of Australian Policy Online (APO) as a collection of 'grey' public policy research and considers how this fits within the broader context of open access publishing and policy development. The amount of grey literature now being produced presents challenges for APO as well as other collecting agencies and society in general. This paper looks at who is operating in this space in Australia, the issues for collecting and preserving grey literature, and how bibliographic exchange and interoperability can help us to manage the flood of new material and reduce the duplicated effort that currently exists. While APO has been a foundational player in this field, it has reached a stage where a new solution is required - a national online collaborative collection of full text digital resources created with a vision for the benefits that the semantic web will bring.

Australian Policy Online (apo.org.au) is a web-based, open access collection of Australian public policy related research and information. The core service involves finding, cataloguing and disseminating information on the latest reports, articles, analyses, working papers and statistical documents published 'informally' by research institutes, government and non-governmental organizations, and other agencies. This material, known as 'grey' literature, is not controlled by commercial publishers (Greynet International), meaning it is only collected or catalogued by libraries, publishers or other services in an ad hoc manner. While APO is also an information network and a community space for those working on public policy issues, this paper focuses particularly on APO's role in collecting grey research and considers how this fits within the broader context of open access publishing and policy development. The amount of grey literature now being produced presents challenges for APO as well as other collecting agencies and society in general. This paper looks at who is operating in this space in Australia, the issues for collecting and preserving grey literature, and how bibliographic exchange and interoperability can help us to manage the flood of new material and reduce the duplicated effort that currently exists. APO has been a foundational player in this field, however it has reached a stage where a new solution is required - a national online collaborative collection of full text digital resources created with a vision for the benefits that the semantic web will bring.

APO's strength is that it catalogues and disseminates scattered resources and provides a central archive for further research. It thereby solves a long-standing problem in public policy – ensuring that those who make and implement policy have access to the latest research findings produced by academic institutions and other organisations; and conversely, that researchers and writers are able to disseminate their findings to policy makers in a timely and effective manner.

APO is one of the major ways of bridging the academic/public policy divide, allowing research to flow between academics and practitioners. Around 200 centres and organisations from around the country form the core of APO's content source including: multi-institutional research networks; university research centres; think tanks from all sides of politics; NGOs, charities, associations and professional bodies; government

agencies; and government departments and inquiries from Federal, State and Local government.

Public policy issues encompass a great breadth of subject matter and the site covers a wide range of disciplines, as evidenced by its ten major areas (see the screen shot in Figure 1.): creative economy, economics, education, environmental and planning, health, Indigenous issues, international issues, justice, politics and social policy. So although public policy is a subject in itself, APO does not focus on any one particular discipline. It could be described as predominately social sciences but with considerable cross over into areas of technology, health, law and natural sciences.

Edited by the Institute for Social Research, Swinburne University of Technology

Research Commentary Websites Guides Audio Video Events Jobs Submissions A - Z Courses

APO AustralianPolicyOnline
ORG.AU "THE BEST ONLINE SOURCE OF AUSTRALIAN POLICY INFORMATION"

Advertise on APO Email Subscription RSS Login TEXT: LARGER | SMALLER

Search APO GO

Home [Creative Economy](#) [Economics](#) [Education](#) [Environment & Planning](#) [Health](#) [Indigenous](#) [International](#) [Justice](#) [Politics](#) [Social Policy](#)

NEW RESEARCH

Enabling Australia: Inquiry into the Migration Treatment of Disability
Joint Standing Committee on Migration
30 June 2010 | Australia's current health requirement for migration to this country is prejudiced against people with a disability and their families, according to this report.

Building a modern committee system
House Standing Committee on Procedure
30 June 2010 | This report recommends a set of complementary measures designed to strengthen the system of committees in the House of Representatives.

IPTV and internet video delivery models: video content services over IP in Australia
Australian Communications and Media Authority
30 June 2010 | Australians are increasingly able to view video content on their own terms, including when, where and how they want it, according to this report.

State of our public hospitals June 2010
Department of Health and Ageing
30 June 2010 | The seventh in an annual series, this report on public hospital performance in 2008-09 includes information on private hospital performance and has chapters on hospital use by Aboriginal and Torres Strait Islander people and on hospital maternity services.

The future of cloud computing
Lee Rainie, Janna Quitney Anderson | Pew Internet and American Life Project
28 June 2010 | Technology experts and stakeholders say they expect internet users will 'live mostly in the cloud' in 2020 and not on the desktop, working through

Compulsory income management: a flawed answer to a complex issue
22 June 2010 | From 1 July 2010 major changes are due to be made to the way income support payments are paid to recipients in the Northern Territory, and potentially across Australia.

COMMENTARY

Campaigning to Christians
30 June 2010 | The elevation of Julia Gillard means the Australian Christian Lobby now needs to engage afresh with a new Prime Minister, writes John Warhurst in Eureka Street

Anxious overture
29 June 2010 | Kevin Rudd's downfall revealed the limits of Labor's ability to introduce reforms opposed by powerful interests, writes Geoffrey Barker in Inside Story

The way ahead is less complicated than it seems
28 June 2010 | Kevin Rudd departed the prime ministership this week

MOST VIEWED: THIS WEEK

- Who will benefit from the 1 July 2010 tax cuts?
- Campaigning to Christians
- Our future world: an analysis of global trends, shocks and scenarios
- Review of Discovery laws and practice to improve access to justice
- Senior Innovation & Strategy Officer

QUICKLINKS

- Citizenship
- Creative industries
- Disability
- Employment
- Housing
- Mental health
- Schools
- Climate change
- Defence
- Education policy
- Health policy
- Immigration & refugees
- Poverty
- Social problems

SUBSCRIBE TO THE APO WEEKLY BRIEFING

JOBS

Senior Innovation & Strategy Officer
Innovation & Strategy Unit, Victorian Department of Justice
Melbourne

Senior Urban Planner: Consulting Opportunity Urbis
Brisbane

Urban Planner: Consulting Opportunity Urbis
Brisbane

EVENTS

WORKSHOP Developing Monitoring and Evaluation Frameworks
13 May 2010 - 6 Aug 2010
Australian capital cities & Cairns

WORKSHOP Using purposeful program theory and logic models for evidence-based policy and practice
28 Jun 2010 - 27 Jul 2010
Melbourne, Adelaide, Brisbane, Darwin, Sydney

CONFERENCE

Figure 1. Home page of Australian Policy Online with the site's ten major subject areas forming the navigation.

Audience and funding

According to an online reader survey conducted in 2007, about a third of APO's audience is within the education sector, at least another third come from local, state or federal government sectors, and the final slice of pie is a mix of NGOs, charities, media, professionals and librarians. By definition, public policy is of concern to the public, as well as specialist researchers, which means that a broad range of people engage with APO for a wide variety of reasons in a way that is different to proprietary research databases or specialist subject archives. As a result the service is particularly appreciated

by those unable to access alternative sources of scholarly material such as journalists, independent researchers, charities and NGOs.

According to Google Analytics figures, APO's readership is growing by at least 30% annually and the site now attracts over 65,000 visits each month, and over half a million for the year, viewing nearly 1.5 million pages. Over 14,000 people subscribe to the *Weekly Briefing* email newsletter and many access the site via RSS feeds, Twitter or the National Library's Australian Research Online harvester. In 2009 APO was ranked in the top 10 for politics websites in Australia by Hitwise (Hitwise 2009).

The site receives financial support from the Institute for Social Research, Swinburne University and the ARC Centre of Excellence for Creative Industries and Innovation (CCI) and has received a number of Australian Research Council (ARC) grants for infrastructure and equipment funding.

Background and project development

Australian Policy Online was established in 2002 as a project of the Institute for Social Research, Swinburne University of Technology, based on a site in the US called Moving On (since closed). The original idea was to gather research reports from academic centres and institutes based on free membership of APO and provide a weekly email newsletter alert. Each week 10 – 15 reports were posted along with 1 or 2 original commentary pieces, some events, calls and the rare job.

With the increased use of the internet as a publishing vehicle for all kinds of organisations including government, it became clear that limiting APO's content to only member organisations was too restrictive. In 2006 the collection policy was expanded to include any policy related report or document including those from government, NGOs and associations where appropriate. With this extended scope APO increased its value, becoming a central resource that allowed key players producing policy research to share information with each other. It also massively expanded the amount of information the editors needed to filter and to share with the APO audience; the site now posts up to 50 items each week with around 35 being research publications.

Over time it became clear that APO needed to move onto a new platform to better cope with the rapidly increasing size of the database and to maximise the potential of new software and functionality. Thanks to an ARC grant, the database was migrated from a proprietary content management system to the open source system Drupal in 2009. As part of the upgrade the site had a visual and architectural redesign, a new subject classification structure was created, new metadata fields were added, the type of content collected was extended to include audio, video and web resources, plus new sections for course and book listings. Various other functions and features were added including faceted search, tailored email, most viewed results and a bookmarking system.

A catalogue or a repository?

In upgrading the site, it was intended that APO would have the capacity to upload and store full text PDFs of documents. This was desirable for two reasons. Firstly to avoid the ongoing and inevitable problem of links going dead as a result of documents being moved and URLs changing. Secondly to allow for full text searching of resources on APO. While this is now technically possible and some publications are uploaded two issues have arisen in carrying out this aim. One is copyright and the obligation to secure permission to store a copy of a copyrighted work. The second is that preserving

documents involves a much greater degree of time, effort and cost. More importantly, it involves a whole new approach to technical and standards issues than APO has dealt with previously. It means becoming a repository rather than just a catalogue, something that has not been part of the site's conception.

The impact of trying to store full text holdings on the site has brought to the fore the need to understand the larger issue of grey literature collecting and preservation within the broader research and publishing environment in Australia. This is not a bad thing, for as Bryson (2004) points out:

To respond effectively to changes in their environments, public and non-profit organizations (and communities) must understand the external and internal contexts within which they find themselves so that they can develop effective strategies to link these two contexts in such a way that public value is created.

Scholarly publishing and the open access movement

A key question facing us today is, are there new opportunities and new models for scholarly communication that could enhance the dissemination of research findings and, thereby, maximise the economic and social returns to public investment in R&D? (Houghton, Steele & Sheehan 2006)

The answer to Houghton's question is both yes and no. The backbone of academic scholarly publishing and communication was, and still is, the traditional system of peer-reviewed scholarly journals along with peer-reviewed monographs (Harley et al. 2010). These are predominantly assessed by academics but published and distributed by publishers and aggregators. To read these journals, even in an online environment, requires access to subscription only databases paid for by university funds and accessed via the academic library. As Harley et. al. (2010) attest, "These traditions, which rely heavily on various forms of peer review, may override the perceived "opportunities" afforded by new technologies, including those falling into the Web 2.0 category'.

In contrast to the persistence of traditional models of publishing is the ongoing exploration of the new opportunities and models referred to by Houghton – primarily with the open access movement. In Budapest 2002 it was declared that 'many different initiatives have shown that open access is economically feasible, that it gives readers extraordinary power to find and make use of relevant literature, and that it gives authors and their works vast and measurable new visibility, readership, and impact' (Budapest Open Access Initiative 2002).

This has been no overnight revolution, and there is still no clear answer to what models work and no solution to the financial and institutional barriers to these new models. A recent JISC report has identified three main options for scholars to consider when publishing research as open access: open access journals, open access repositories and open access repositories with overlay services – all of which only account for formally published journal articles and conference papers (Swan 2010).

For many involved in the open access movement in higher education, the publications of interest are peer-reviewed journal articles or their pre-print versions and peer-reviewed conference papers. The management and preservation of material beyond these parameters is often not considered part of the investigation.

While most university based researchers are able to access both subscription and open access research, many in government, NGOs, the community sector, media, professions, and the general public have had limited access to scholarship. As a result, in the public policy arena (which in itself cuts across many social science and some science and medical disciplines) there has developed a strong culture of informal publishing that looks set to continue and in fact, increase massively.

APO's rapid collection of grey literature is particularly important in public policy as new research can be influential if it is timely and accessible (i.e. not held up by the slowness of journal publishing or the need to write for an academic audience) and reports are often produced explicitly to provide evidence for policy decisions.

Also in public policy there are many producers of research outside of academia including government and non-government organisations, associations and think tanks. Grey literature is often the best source of up-to-date research in specific areas and it is generally written in a more accessible style than journal articles, providing a clear, concise introduction to difficult or complex topics. In health, law, environmental studies, economics, creative industries to name a few, grey literature is a major part of the research material published and used. As Tyndal (2008) points out in the health sciences,

While searching the published (black) literature is a given, there is a growing recognition that grey literature should be included to fully reflect the existing evidential base... Literature reviews need to include the most significant research available. And that might mean conference papers, reports, legislation or working papers - all grey.

As a result of simplified tools for desk top and web publishing, the explosion in content management systems, and the power and popularity of internet search engines, it is easier than ever for individuals and organisations to produce and publish their own research online. The amount of quality research material and other relevant and legitimate material being published outside commercial publishing channels is now completely unmanageable without a major, coordinated and collaborative change of approach, plus considerable technical input.

Collecting grey literature

Despite its importance and increased prevalence there is little consensus on how grey literature should be published, collected, disseminated, evaluated or preserved in Australia (or elsewhere). In a recent call for papers for *Cataloguing & Classification Quarterly* on the issue of grey literature the editors state that:

Despite the advantages offered by new information technologies and, in particular by the ability of search engines to retrieve documents on the web, grey literature still lacks bibliographic control. ...unlike journals and books, the absence of commercial stakes has contributed to a situation where the way in which the different types of grey literature are referenced still depends more on choices made by the bodies that produce, collect, or distribute these documents than on any national or international standard (A Library Writers Blog 2010).

In Australia there is still no systematic collection or preservation of grey documents, or any agreement on how this should be done. In terms of government documents, the National Library of Australia collects some government publications but is not obliged to do so in any exhaustive way. The National Archives collects all government documents but not until departments agree to finally hand them over, which may be years after items

are published. This is set to change to some extent with the recent adoption by the Australian Government of the recommendations from the Gov 2.0 taskforce (Australian Government 2010).

Due to a major injection of funding, most Australian universities have instituted repositories in the last 5 years to collect and preserve their own staff's publications (whether open access or not) (ARROW 2008). However the majority of university repositories limit their policies to collecting only scholarly journal publications and conference papers (exceptions include Swinburne University of Technology and University of Technology, Sydney).

As for the not-for-profit or 'third sector', it seems that no one is thoroughly collecting or preserving the output of NGOs, charities, service agencies, associations, corporate researchers, professional sectors or independent scholars. Yet these organisations are commissioning and publishing substantial research material with direct bearing on public policy issues.

Australian Policy Online is one of the largest collections of social science grey literature in Australia and yet it is not comprehensive, does not preserve material and restricts its selection to public policy related material. Of course APO is not the only subject-based gateway to online research, there are in fact over 55 according to a review conducted recently of online collections. Like the objects they collect, there are many terms that are currently in use to describe an online database of resources – clearing house, digital library, resource centre, subject gateway etc. and there may be subtle differences in the aims and functions of each one. Many have been created to resolve a perceived need for "a collection of resources" on a particular topic and thus a new database and website is born. They range from small to large – from the Victorian Women's Disability Clearinghouse to Health Insite for public health guides – but all are subject specific. Another type of collection is the university and organisation repository but here the collection policy is based on author association rather than subject, and, as mentioned earlier, most do not collect grey literature.

The problem with so many clearing houses and subject collections is that currently they are all cataloguing online content which means there is a great deal of duplicated cataloguing of grey literature for subject clearing houses, libraries and specialist collections – mostly at tax payers' expense. There is little sharing of metadata records for materials freely available online and even though many websites are in fact digital libraries, most aren't currently available to browse within the public or academic library systems.

In terms of preservation various problems exist including:

- No agreed frameworks or metadata standards for cataloguing publications and other related items.
- No mandate or agreement to preserve documents and no systematic archival procedure.
- No agreement on the allocation of persistent identifiers and therefore no capacity to link documents and research data or associated content.
- No clearly articulated publishing standards for organisations wishing to publish research informally i.e. using permanent URLs, Dublin core metadata etc.
- No place for individuals or organisations to upload their own documents in an Australian repository.

As the recent report from the Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010 states:

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.

International systems

Internationally there are a number of interesting examples of how open access research and grey literature are being managed. Like Australia there are many clearing houses and specialist subject gateways however there are also a number of large scale collecting operations that provide interesting models for consideration.

Social Science Research Network (www.ssrn.com) is an international repository where researchers are able to upload their documents. The system is run as a business by charging annual subscriptions to subject specific email newsletters in various disciplines. It also includes abstracts from open access journals. This kind of mass collaborative model is set to take over as the kind of search system researchers want to use. However additional services such as the SSRN's email newsletters could be provided by database partners if a more communal model was adopted.

A more general example is Scribd (www.scribd.com), which describes itself as 'the largest social publishing and reading site in the world.' Scribd allows anyone to post anything and turns all documents into web-based pages, allowing tagging and sharing. Harvard University Press uses Scribd to post its catalogue for readers to look at on Scribd – so it is both a publishing tool for non-publishers and a marketing tool for publishers.

These kinds of systems represent a major change in how individuals and organisations are able to disseminate their own material – they don't even need their own website, nor do they require the editing services of a site such as APO. It therefore appears necessary for APO to seriously question the extent to which it should be cataloguing all its own material and instead work towards creating a larger, open repository where grey literature could be uploaded and catalogued, with layers of services added on including metadata exchange and sharing, filtering and dissemination.

Filtering and evaluation

In the information age, this ability to filter effectively has moved from an essential of survival to one of the primary determinants of success (Dawson 2003).

As the quantity of quality research being made available increases along with the volume of grey literature, so does the need to filter it. It is not enough for APO to find, catalogue and disseminate research, the service must find ways to filter information so that readers are not bombarded with ever increasing amounts of undifferentiated information. Even the traditional systems of peer-review are no longer adequate given the amount of material available and various attempts at post publication peer review are now being made. Houghton, Steele & Sheehan (2006) suggest that from this problem springs a number of opportunities:

there are potential futures for publishers to become value adding service providers overlaying open access content (e.g. peer review services, bibliometrics and webometrics for research evaluation, etc.). In turn, these might enhance research evaluation and lead to better focused R&D expenditures.




Filtering systems range from simple automated systems to massive international collaborations by skilled experts. In terms of the former, filtering systems can be created automatically based on user interaction with a site's content. For example by displaying how many people viewed a particular page or downloaded a report, how many other publications cited it, how many recommended it or found it useful.

Yet while these numbers have a place they are limited within the research domain. Recommended for what? Useful in what way? The rating system on the video site 'Ted talks' (ted.com) doesn't just ask users to say if they liked or disliked a video, or to click a thumbs up or down symbol, it asks users to consider whether the item was jaw-dropping, persuasive, funny, longwinded, unconvincing etc. Perhaps therefore we should be looking at a rating system that might provide options such as: ground breaking, well-written, good synthesis, what works etc. In this way the site might incorporate a web 2.0 type of peer-review for grey literature.

An interesting example of post-publication peer review is Faculty of 1000 which currently has two sites, one for Biology and one for Medicine. The screen shot below shows the rating system that can be applied by over 2000 biologists from around the world, in addition to short paragraphs indicating a publication's research value. The Faculty of 1000 is an interesting model for applying post-publication peer review to grey literature.

For each paper they select, the Faculty member:

- Writes a brief Comment** explaining the main findings and why they find the paper interesting. These comments are : intelligible to a general biology audience. Faculty members often mention related seminal research articles within the relevant papers.
- Assigns a Rating**

 Recommended F1000 Factor 3.1	"Recommended": Recommended reading for a section (i.e. of specialist interest) (F1000 Biology rating of 3)
 Must Read F1000 Factor 6.2	"Must read": Must read for more than one subject/section (i.e. of general interest) (F1000 Biology rating of 6),
 Exceptional F1000 Factor 10.7	"Exceptional": A landmark paper representing the top 1% of publications (F1000 Biology rating of 9).

The ratings assigned to each paper is then used to calculate its F1000 factor (see below), which also takes into account the same paper.
- Classifies it** into any of seven paper types
 - New finding
 - Technical advance
 - Interesting hypothesis
 - Important confirmation
 - Controversial findings
 - Refutation
 - Novel Drug Target

Figure 2. Faculty of 1000 Biology evaluation system <http://f1000biology.com/about/system>

Tyndal (2008) points out that while there is some analysis and acknowledgement of grey literature's importance in the health sciences, there is major uncertainty about what evaluation criteria should be used as there is no clear hierarchy to evaluate grey literature, with different disciplines focussing on different resources. In her view the standard evaluation criteria for websites should be used: authority, accuracy, objectivity, coverage and date – known as AA OCD. This could be used as the basis for discussing collection policies for grey literature – a very difficult issue that is beyond the scope of this paper.

Standards and interoperability

Trends in the web environment are emphasizing interoperability among multiple standards for metadata used and recycled in the book industry, rights management, and library and information sectors (Program for Cooperative Cataloguing Mission Statement 2010).

The library sector has long known the value of bibliographic exchange and has been working on this for the last 30 years (Hider & Harvey 2008). In World Cat, using the Online Computer Library Centre (OCLC) Connexion software, cataloguers can input URLs for web-based materials and Connexion will automatically generate brief bibliographic records for them which human catalogues then edit and add to (OCLC 2010).

Another type of automation is the extraction of keywords for subject classification such as the Open Calais system (www.opencalais.com). Academic reference management and bookmarking software such as Mendeley (www.mendeley.com), Connotea (www.connotea.org), Icyte (www.icyte.com) and Zotero (www.zotero.org) allow readers to post metadata details with the click of a button – but the effectiveness of such systems depends to some extent on all producers of research documents using metadata in the first place. This may require the promotion of Dublin core to informal publishers of research in Australia as well as the installation of the capturing technology in online collection infrastructure.

Mendeley appears to be one of the rising stars in this area. It allows readers to create their 'own personal bibliography online' by indexing and organising PDF uploads. It also looks up PubMed, CrossRef, DOIs and other related document details automatically. The Web Importer allows you to quickly and easily import papers from resources such as Google Scholar, ACM, IEEE and many more at the click of a button.

Protocols such as RSS and the Open Archives Initiative (OAI) allow APO to be harvested by various systems, including the National Library of Australia and the OCLC's World Catalogue via OAIster. This is a good start but there also needs to be collaboration on how information enters the system in the first place. With the combination of technology and metadata standards, there seems to be little justification for research, including grey literature research, to be catalogued multiple times when it is so easy to share metadata if agreed standards are used.

Standards for grey literature publishing and collecting have been agreed upon and implemented to some extent in the sciences (Committee 2007) but there has been less take up and implementation within the social sciences.

Interoperability is the capability that allows different computer systems to share information across a network... By adopting a common set of best practices, controlled vocabularies, and interoperable system architecture, institutions can

increase their visibility and provide opportunities for new connections with others to serve the shared needs of constituent communities (CDP Metadata Working Group 2006).

With so many standards and schemas now in existence it is hard to know where to start to identify what might be the best options for grey literature. According to Hider there are two solutions to the problem of standards proliferation - translate or unify; 'The main attempt at unification has been Dublin Core...' (2008 p. 214). Dublin Core stands out as a schema because it aims to be used across information domains rather than be more specialised. However it was designed to describe web pages and is therefore limited in its ability to contextualise information. Despite this no other metadata schema has been adopted as widely. One of Dublin core's great advantages is that it works with the open archives initiative protocol for metadata harvesting (OAI PMH) which has become a popular vehicle for sharing metadata across collections in the library community.

However, according to the Common European Research Information Format project, the current use of Dublin Core or even MARC means that 'the metadata is machine readable but not machine-understandable. The end-result is that the input, update and retrieval processes are human-intensive. With increasing availability of repositories of grey literature this will not scale.' The Europeans are therefore creating their own system - called CERIF. It is not clear at this stage which system if any will become a universal standard so this situation needs continued monitoring.

Warwick Cathro at the National Library of Australia (2009) identifies some of the key standards for data interoperability as:

- MARC (Machine-Readable Cataloguing) – used by library catalogues
- MODS (Metadata Object Description Schema) – developed by Library of Congress to provide a link between MARC and Dublin Core.
- METS (Metadata Encoding and Transmission Standard) Structured for encoding descriptive, administrative, and structural metadata. Designed by the Library of Congress especially for metadata about electronic publications in digital libraries. Includes extra administrative metadata.
- PREMIS (Preservation Metadata – Implementation Strategies) - A data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials.

In the case of software inter-operability, relevant standards include the NISO Circulation Interchange Protocol (NCIP), the Information Retrieval Protocols (Z39.50, SRU), and the Open Archives Initiative Protocol for Metadata Harvesting (OAI).' (Cathro, 2009). Another important standard for maximising the potential of the long awaited semantic web is Resource Description Framework (RDF). This a standard framework for structuring metadata in XML and all metadata schemas expressed in XML should conform to the RDF. The advantage of RDF is that it allows relationships between web documents to be identified by a computer. It is a core element of the semantic web which many are hailing as the next big leap for the internet (O'Reilly 2009).

Finally there is a need for Uniform Resource Identifiers (URI) and the two systems currently operating to provide these: DOI and Handles. These protocols enable a distributed computer system to store identifiers, known as handles, of arbitrary resources and resolve those handles into the information necessary to locate, access, contact, authenticate, or otherwise make use of the resources – in particular to always be able to

locate and link them to other material. DOI's are proprietary while Handles are open source. DOI's have been adopted by publishers while Handles are being used by University repositories.

Linked Data

Interoperability and standards are at the core of the semantic web, which is about making sure that the relationships between information are machine readable and therefore able to be connected and visualised. This raises the issue of how publications and data can be linked. In biology researchers have been working on how to connect publications and data for some time and Philip Bourne (2005) presents a compelling case for what might be possible:

...what if the data in an online paper became more alive? Some databases let you download data into spreadsheets or other client-side applications that render and analyze data. Papers could be treated this way, too. The technology is there to create these ubiquitous clients that are independent of operating systems and hardware and that are downloadable on demand. New levels of comprehension might be achievable.

A great deal of public policy and social science research is quantitative and the opportunity to analyse the data as well its interpretation would be a great boon for evidence-based policy. A major opportunity presented by the semantic web is the potential to connect research publications with the data held in collecting agencies such as the Australian National Data Service and the Australian Social Science Data Archive (ASSDA). But to do this there needs to be a commitment to preserving grey literature and implementing standards using handles or DOI on all grey literature publications and have relatable identifiers on datasets.

Conclusion

As a catalogue of public policy grey literature, Australian Policy Online needs to understand the context in which its services are used and the way in which the type of content it holds is perceived, created, managed, preserved and disseminated. This analysis has revealed that in fact the service that APO provides is a unique and valuable one but that there is a greater issue that APO could be assisting to resolve – that of collaboratively preserving, cataloguing and sharing metadata about Australian grey literature. At the same time there are many efficiencies that APO could gain from implementing automated data extraction now where possible. Cataloguing and preservation need to be separated from services such as specialist subject access, evaluation, filtering systems, value added services, and community of practice stewardship. APO aims to be a driving force in helping to solve this national research infrastructure problem while continuing to provide specialist services for its constituent community of public policy researchers and practitioners. We welcome any comments or thoughts from others interested in creating such a resource.

References

- A Library Writers Blog 2010. *CFP: Special issue: Cataloging Grey Literature (Cataloging & Classification Quarterly)*. Available at:
<http://librarywriting.blogspot.com/2010/01/cfp-special-issue-cataloginggrey.html> (accessed 17 February 2010).
- Australian Research Repositories Online to the World (ARROW). Available at:
<http://arrow.edu.au/> (accessed 10 April 2010).

- Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010, *Sustainable economics for a digital planet*, OCLC. Available at: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf (accessed 4 June 2010).
- Bourne, P. 2005. Will a Biological Database Be Different from a Biological Journal? *PLoS Computational Biology*, Vol 1, Issue 3. Available at: [2010.http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1193993/pdf/pcbi.0010034.pdf](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1193993/pdf/pcbi.0010034.pdf) (accessed 24 May).
- Bryson, J. M. 2004. *Strategic planning for public and nonprofit organizations: A guide to strengthening and sustaining organizational achievement*, 3rd edition., John Wiley & Sons, Hoboken.
- Committee, GLIS 2007. *Guidelines for the production of scientific and technical reports: how to write and distribute grey literature*. Available at: <http://www.glisc.info/> (accessed 15 March 2010).
- Budapest Open Access Initiative, 2002. Open Society Initiative. Available at: <http://www.soros.org/openaccess/read.shtml> (accessed 15 May 2010).
- Cathro, W. 2009. Collaboration Strategies for Digital Collections: The Australian Experience, International Conference on Libraries, Seoul, Korea, 25-26 May 2009. Available at: <http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1433/1738> (accessed 22 April 2010).
- CDP Metadata Working Group 2006. *Dublin core metadata best practices*, Collaborative Digitization Program, University of Denver. Available at: <http://www.bcr.org/dps/cdp/best/dublin-core-bp.pdf> (accessed 13 June 2010).
- Dawson, R. 2003. *Living networks*, Prentice Hall, Upper Saddle River, New Jersey.
- Faculty of 1000. Available at: <http://f1000.com> (accessed 10 April 2010).
- Australian Government, 2010, *Government Response to the Report of the Government 2.0 Taskforce*, Canberra. Available at: <http://www.finance.gov.au/publications/govresponse20report/index.html> (accessed 30 June 2010).
- Greynet International, nd. Available at: <http://www.greynet.org/greynethome.html> (accessed 15 May 2010).
- Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S. and King, C. J. 2010. *Assessing the future landscape of scholarly communication: an exploration of faculty values and needs in seven disciplines*, Center for Studies in Higher Education, University of California, Berkeley. Available at: http://escholarship.org/uc/cshe_fsc (accessed 10 June 2010).
- Hearst, M. A. 2006. *Design recommendations for hierarchical faceted search interfaces*. School of Information, UC Berkeley. Available at: <http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf> (accessed 15 March 2009).
- Hider, P. and Harvey, R. 2008. *Organising knowledge in a global society*, Centre for Information Studies, Charles Sturt University, Wagga Wagga.
- Hitwise 2009. *Hitwise Top 10 Online Performance Award 2009*, 2009. Available at: <http://www.hitwise.com/awards/popup.html?market=au&sDomain=www.apo.org.au&iDate=200904&iCatnum=295&Cal=&semi=2> (viewed 15 June 2010).
- Houghton, J. Steele, C and Sheehan, P. 2006. *Research communication costs in Australia, emerging opportunities and benefits*, Centre for Strategic Economic Studies, Victoria University, Melbourne. Available at: <http://www.cfses.com/documents/wp24.pdf> (accessed 27 April 2008).
- OAIster, nd. Available at: <http://www.oclc.org/oaister/> (accessed 15 June 2010).

- OCLC 2010. Connexion, Ohio. Available at: <http://www.oclc.org/connexion/> (accessed 16 May 2010).
- OECD 2010. *The economic and social role of Internet intermediaries*. Available at: <http://www.oecd.org/dataoecd/49/4/44949023.pdf> (accessed 16 April 2010).
- O'Reilly, T. & Battelle, J. 2009. *Web Squared: Web 2.0 Five Years On*. Available at: http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper.pdf (accessed 1 September 2009).
- Program for Cooperative Cataloging Mission Statement*, 2010. Library of Congress. Available at: <http://www.loc.gov/catdir/pcc/stratdir-2010.pdf> (accessed 13 June 2010).
- Research and the scholarly communications process – principles*. 2006. Research Information Network, London. Available at: <http://www.rin.ac.uk/system/files/attachments/Research-scholarlycommunications-principles.pdf> (accessed 29 March 2010).
- Shirky, C. 2009. *Here comes everybody: the power of organizing without organizations*, Allen Lane, London.
- Swan, A. 2010. *Modelling scholarly communication options: costs and benefits for universities*, JISC, London. Available at: <http://ierepository.jisc.ac.uk/442/> (accessed 13 May 2010).
- Tyndal, J. 2008. *How low can you go? Towards a hierarchy of grey literature.*, Dreaming08: Australian Library and Information Association Biennial Conference, 2-5 September 2008, Alice Springs. Available at: <http://hdl.handle.net/2328/3326> (accessed 18 May 2010).

About the Author

Amanda Lawrence is managing editor of Australian Policy Online (apo.org.au) and Creative Economy (creative.org.au) at the Institute for Social Research, Swinburne University of Technology. She also manages the website of the ARC Centre of Excellence for Creative Industries and Innovation (CCI). Amanda created the information architecture and oversaw the upgrades for APO and CCI websites in 2009. She recently completed a Graduate Diploma of Library and Information Management at RMIT. Amanda has a Bachelor of Arts (Hons) from the University of Melbourne and was Manager of the Literature Program at the University's Asialink Centre from 1996 – 2005. She has also worked as a bookseller and researcher. Email: alawrence@swin.edu.au