

Australian subject gateways - metadata as an agent of change

Debbie Campbell
Manager, Infrastructure Projects
National Library of Australia
dcampbel@nla.gov.au

Abstract:

Although the concept of a subject gateway has only risen to prominence in Australia during the last two years, several university library consortia and other education-based initiatives, including EdNA Online and the Australian Digital Theses Project, have already been successful in achieving collaborative approaches for the delivery of national access to electronic resources. The critical success factors realised, including the deployment of metadata and the community approaches to problem solving, mean that Australia is positioned to participate in new initiatives exploring service provision in a national infrastructure for integrated research and information access.

Introduction

The concept of a subject gateway has only risen to prominence in Australia during the last two years. Several university library consortia and other education-based initiatives, including EdNA Online and the Australian Digital Theses Project, have already been successful in achieving collaborative approaches for the delivery of national access to electronic or digitised resources.

A subject gateway may be defined as: " a Web-based mechanism for accessing a collection of high quality, evaluated resources identified to support research in a particular subject discipline where the resources are evaluated and described by information specialists in the field, such as science librarians." [SG]

The Australian higher education sector and its partners have established three new discipline-focused research entry points - to agriculture, chemistry, engineering and information technology. They are known as Agrigate, MetaChem, and AVEL (Australian Virtual Engineering Library). Their respective hosts are the University of Melbourne [AGRIGATE], the Australian Defence Force Academy for the University of New South Wales [METACHEM], and the University of Queensland [AVEL].

The first of these gateways, MetaChem, which was launched in January 1999, was approved on the basis of its goals attempting to change electronically based information service provision in Australia. This opportunity was provided by the use of metadata.

With EdNA Online [EdNA] and the Australian Digital Theses [ADT] Project, the gateways have the following features in common.

Common gateway features

1. The host institutions of each gateway have a mandate to be information providers, but they are not required to be information creators. All except EdNA Online (which is managed by a private company, Education.Au) are hosted by university libraries. But the Education Network of Australia has an even stronger link to the education community: it is funded by all of Australia's State and Commonwealth Departments of Education.
2. Each gateway has a distinctive logo and a reflective name, thereby positioning themselves for a significant Web presence. Distinctive branding also bears fruit as a common recognition technique.
3. Each gateway has selected and utilised a standard metadata schema for describing the resources incorporated into the gateways. The metadata is applied by librarians or educators with experienced knowledge of the disciplines.
4. The resources in each gateway have all been selected according to pre-determined criteria, published at each gateway site as part of a content coverage policy. A gatekeeper function, often a mix of computer and human intervention, ensures

adherence to the selection criteria. This function lends itself to a high level quality rating for the site as a whole, but in conjunction with the commitment by each gateway's partners to a valid, valuable service. While the gateways have targeted academic communities (due to the nature of their source funding and development), the currently free access means that they may be used by any member of the public with an interest in the discipline.

5. The coverage policies have made electronic resources, both 'born digital' and digitised, the highest priority for inclusion in the gateways; but they are all extensible to include books as well as bytes, and under-utilised or unknown resources such as databases, and descriptions of people.
6. The resources are supported by similar architectures, based on distributed creation and maintenance of their metadata, with a centralised facility for access. The metadata, if embedded in resources prior to their selection, may be enhanced or merely augmented but not ignored. The access facility, a navigation interface which provides for the metadata to be queried transparently or explicitly, is provided in at least one of three forms - keyword searching, index browsing, or structured pathways.
7. The navigation points are usually supported by the use of at least one thesaurus, to ensure reliability of resource description and subsequent search results. AVEL, for example, uses the Library of Congress Subject Headings (LCSH). Agrigate uses CABI and its Australian extensions, Agterms, as managed by the Kondinin Group in Western Australia. While a thesaurus enables subject terms to be selected for semantic consistency, the use of "street" terms for the creation of a dynamic thesaurus has also been considered.

Other features of commonality are discussed in a list developed by the National Library of Australia, which addresses business, functional and technical considerations [BESTPRAC]. While any single gateway won't be able to observe all of these best practices, adoption of a majority should lead to critical success.

Critical success factors

Critical success factors for the projects which produced the gateways, such as strong project management, are not discussed here. Rather, the longer-term guarantees of success, such as collaboration and the application of standards are described.

Collaboration

Each gateway has reviewed the successes and failures of its predecessors, so in a sense, their common features are also their critical success factors. However, some additional factors may be highlighted.

Firstly, and most significantly, is their collaboration across university campuses and other research institutions. Each gateway identified the need to pool efforts early in the gestation of the proposed service. While some of the factors driving this need were political, in general they were also pragmatic. Commitment to a common goal,

that of improving patron access to high quality Web content, also expedited the collaboration.

Secondly, the collaboration has been reflected during the development phases of each gateway. The establishment of discussion lists for problem solving proved effective, and their continued existence ensures faster information exchange and issue resolution.

Thirdly, collaboration has resulted in strong content coverage by each gateway. The combination of single institutions (a content consortium) brought specialisation to each gateway's coverage which would have been difficult for any one individual institution to cover robustly.

Collaboration could be extended in the future because the gateways have the potential to allow other individual institutions to contribute - single libraries do not need to maintain their own 'Internet subject guides' to various disciplines. The potential for a reduction in the duplication of these types of services is enormous.

The application of standards

The use of standards has also been important. The first standard considered was metadata. While each gateway development team conducted an independent assessment of available metadata schemas, the Dublin Core was selected as being a suitable baseline in all cases. The gateways also chose elements from the Australian Government Locator Service [AGLS] and EdNA schemas (already based on the Dublin Core), and conducted some experiments with the A-Core, a draft management schema. [A-Core]

Where the Dublin Core is not the schema of first choice for description, it may still be used as a transfer syntax for older applications. For example, the Library of Congress has made mappings available which permit the conversion of USMARC to Dublin Core and vice versa. The National Library has participated in OCLC's Cooperative Online Resource Catalog Project to experiment with this type of application [CORC]. Harmonisation is another possible technique, whereby an older existing schema or a proprietary schema may be standardised by replacing its elements with the Dublin Core and necessary local extensions.

The Australian Digital Theses Project is also exploring ways of converting the Dublin Core metadata created via the thesis submission process to USMARC, which can be integrated into cataloguing workflow processes. The MARC records are provided to Kinetica and local catalogues.

Thesauri, a second standards application, were not implemented until after distributed metadata creation demonstrated their need. Each gateway has selected one, although metadata can usually support more than one. It should be noted that most thesauri are not linked in an electronic sense to the points of metadata creation or patron navigation, although some projects are attempting their integration in these ways.

Cross-gateway searching

There is still discussion over a third standard, for the support of cross-gateway searching. The trade-offs between the use of Z39.50, data export or import, and proxies which query repositories simultaneously are being tested. The scope for this is yet to be determined. While initiatives such as the national ZedWeb Project have tested use of Z39.50 across multiple servers, the concept required the patron to have fore knowledge of each service in order to choose between them [ZEDWEB].

Extensions to services such as cross-browsing require the development of other standards, for example in the area of collection description metadata. This has been a focus of activity at the UK Office for Library and Information Networking [UKOLN] in 1998-99. As a complement to this, the National Library of Australia is exploring the possibility of exploiting information already made available in the areas of subject strengths on the Australian Libraries Gateway which includes previously maintained Conspectus details, to provide links to the subject gateways.

Cross-searching and cross-browsing could be facilitated by simple links between existing pieces of information, rather than building whole new structures which individual institutions may feel unable to commit to. Two services already exemplify this: the Register of Australian Archives and Manuscripts [RAAM], which provides situational links to information in the Directory of Archives in Australia, and the Bright Spars archive which links to the same directory, and to related information in the National Archives of Australia [BSPARCS].

The ubiquitous standard which underpins all of the gateways is the use of the hypertext transmission protocol in Web browsers for which the gateways interfaces were designed. Without this uniform technology, the gateway projects would not have been successful in their early stages.

Benchmarking

To measure true service delivery success, it is necessary to benchmark subject gateways against other information services. The services need to be carefully selected, because their long-term goals may be different. Two types of information services are commercial search engines and library catalogues, specifically Online Public Access Catalogues which are provided for access to the collections of legal deposit institutions such as the National and State Libraries, and University Catalogues. Several of their features are examined here.

All three types of services differ in how they initially identify resources, but they also borrow techniques from each other to achieve this, for example, by relying on notification from interested providers. Both gateways and search engines encourage this in a serendipitous sense, while library catalogues still benefit from long relationships with their business partners, publishers.

The end result of this identification process also varies - search engines do not discriminate between resources. Library catalogues, while more selective, still aim to be comprehensive. There is a reluctance, however, to encompass electronic resources.

This is in part due to a reasonable, albeit unquantified, analysis that the investment required for the creation of bibliographic description was not worthwhile for transitory resources. New views have changed this - more Web sites are created by trusted providers, and metadata schemas are available which can be implemented by automated generation of descriptive data.

The quality and quantity of metadata applied to the resources in each service varies considerably. Catalogues have always permitted rich bibliographic description in MARC. By contrast, search engines have shown little interest in the potential of metadata, with a few known exceptions. Early use of descriptive metadata elements has been attempted by Alta Vista. Some manual (human) analysis is provided by Yahoo subject categories.

Subject gateways have been very successful at exploring the usefulness of new metadata schemas, and using them as a vehicle for facilitating change in the provision of Web resources for research purposes. It is possible to create rich descriptions of resources with these schemas, but they go beyond identification and location information. Rights management, quality ratings, and conditions of access are all being explored under the gateways' auspices through the application of metadata.

The addition of these capabilities via the exploitation of metadata may even change the service name 'subject gateway' to that of 'portal'. It is too early to use the 'portal' concept as a benchmarking service, as their scope is not yet clearly understood. The origin of the term occurred when commercialised search engine services were re-engineered to be tuned to individual requirements. In the library community, they may be defined as an amalgamation of services to the patron where the amalgamation is achieved through seamless integration of existing services by using binding agents such as customisation and authentication services, search protocols such as Z39.50, loan protocols such as ISO10161, and e-commerce. The result is a personalised service which allows the individual to access the rich content of both print-based and electronic systems.

There are no so-named portal services in Australia, but two international library initiatives are being monitored: My.Gateway and My.Library. My.Gateway is an initiative of the University of Washington Library [UW]. The service equates itself to bookmarking functionality in Web browsers, but it also offers some push technology by providing alerts about new resources. My.Library is a project of North Carolina State University [NCSU]. It provides a single interface to various types of information - news about the library and its services, as well as personal links to library resources. The scope of these initiatives suggests that we potentially already have many portals in Australia - it is merely the use of the terminology which is different.

The use of metadata for managing the quality of Australian portals (or subject gateways of the future) will need to remain. The application of (broken) link checkers which may be triggered by metadata, for example, has increased confidence in subject gateway resources. Like catalogues, the gateways aim to remain trusted information sources by facilitating discovery through the use of enhancing metadata such as subject keywords.

But subject gateways have deliberately limited the number of resources they describe. They link to other related services wherever possible. They also focus on content which is accessible in real time. An important feature they share with search engines is their ease of access to content without mediation. Carl Grant of Ex-Libris made the point recently at the national Library Technicians Conference that although information providers might not want their patrons to use the Web as their exclusive access point for research purposes, patrons are making this happen [GRANT].

To make sure it happens in the best way possible for both patrons and providers, libraries need to facilitate more collaboration. Search engines are unlikely to do this unless there is a commercial imperative. Catalogues and subject gateways, on the other hand, already have an infrastructure at their disposal to share their resources - human resources, financial resources and content.

Extended resource sharing

Human resources

Sharing of human resources will occur through the permanent institutionalisation of subject gateways. Libraries and other organisations need to ensure their long-term commitment through continued funding and integration of new initiatives with existing services.

Financial resources

The redeemed value of gateway services is yet to be quantified. While there is some comfort in knowing that there is routine identification and assessment of Web resources in particular disciplines, more funding could strengthen and enhance the scope of these efforts. Alternative sources of funding, such as sponsorship or new fee-based services where the gateways manage a digital archive on behalf of key stakeholders, have not yet been explored in Australia comprehensively. There are no clear business models in Australia or overseas for the sustainability of subject gateway services.

Content

The collaborative theme of gateway providers spread across several universities and other research institutions has allowed each provider to derive benefit by sharing the responsibility for identifying and describing content. They are also able to reuse older content.

Sharing of content may also be effected by establishing partnerships with national gateways in other countries. Shared content may include different levels of metadata: resource description, subject classification, or collection description. Through their involvement in the international subject gateway mesh initiative, known as IMesh, Australia gateway providers have established partnerships with related UK and US gateways. For example, Agrigate has commenced discussions with a United States equivalent - AgNIC - to explore data exchange. Agrigate is also featured in the ISAAC Scout report [SIGNPOST]. MetaChem has explored the use of the subject

categories deployed by BUBL, which is itself based on the Dewey Decimal Classification scheme [BUBL]. AVEL has formed a partnership with its UK equivalent, the Edinburgh Engineering Virtual Library [EEVL].

Collaborating beyond Australia

The issues described are not unique to Australia, although our environment shapes the processes we go through when attempting to find solutions. There is already extensive potential to participate in like-minded United States and United Kingdom-based initiatives which are attempting to find solutions in these areas. The test beds are the subject gateways.

A few of the initiatives are worth discussing in more detail: the United States' ISAAC Network, the United Kingdom's interoperability focus, and a project which spans both continents: the IMesh Toolkit.

The "Information Seeker's Avenue to Authoritative Content" Network

The ISAAC Network has its auspices in the Internet Scout Project. SCOUT is funded by the National Science Foundation, at the University of Wisconsin-Madison, and provides a service to the U.S. higher education community by identifying research resources. SCOUT offers to show searchers the way to the best resources on the Internet. Librarians and educators filter hundreds of announcements looking for the online resources most valuable to the education community, although they are not restricted to other searchers. The alerting service is similar in style to that provided by EdNA's *'What's New'* service. EdNA is in fact a partner in the ISAAC Network [ISAAC].

The IMesh Toolkit

The IMesh Toolkit project began on 1st September 1999, as a joint initiative of UKOLN and the United States' National Science Foundation. It will build on existing UK subject gateway software to develop a configurable, reusable and extensible toolkit for gateway providers. In addition, the project will sponsor a research strand to consider related issues such as the operation of a distributed international subject gateway development, which may focus on metadata sharing and reuse. [TOOLKIT].

The formal deliverables from the project include: an architecture, application programming interfaces, software distribution, an integrated development environment for metadata, a metadata registry with complementary tools, documentation, and recommendations for addressing issues faced by both patrons and service providers. Some of these outcomes will be delivered through identifying existing research efforts, others will be developed. Cooperation will be facilitated with the international subject gateway community via the IMesh mailing list. [IMESH]

Interoperability Focus

The interoperability focus is an initiative of the United Kingdom Office for Library and Information Networking [UKOLN]. Its goal is "for exploring, publicising and mobilising the benefits and practice of effective interoperability across diverse information sectors, including libraries and the cultural heritage and archival communities. A key aspect of this work is the identification and exploitation of synergies with existing UKOLN and external projects, with a view to maximising returns on the ongoing work of projects such as MODELS and other initiatives." [IF]

Interoperability is considered to be possible in five strands. Firstly, in the technical strand, where systems are permitted to communicate by using software protocols such as Z39.50, ISO10160-61 for InterLibrary Loan, and eXtensible Mark-up Language (XML). The second strand is semantic, where content needs to be mapped or harmonised in order to be joined into a centralised service. Standards such as thesauri can assist these processes.

Thirdly, organisations, their staff and their patrons need to negotiate and agree to reach interoperable status. Fourthly, inter-community relationships permit interoperability to occur. An Australian example is the collaboration between the National and State Libraries, and between cultural heritage institutions. And lastly, international interoperability is of key significance, both for common problem resolution and sharing of content.

Australia is participating in a number of these initiatives. Significant development work has been started by UKOLN to encompass collection description metadata, part of the middleware layer which supports interoperability. The National Library is exploring the potential of this schema by assessing it against the collection strength data held in the Australian Libraries Gateway and updated by its owners - more than 5,000 libraries around Australia [ALG].

Conclusion

In 2000 and beyond, Australia would like to capitalise on collaboration to research and develop new services involving the broadest possible group of information service providers. The initiatives mentioned, and others identified as a result of the work as it proceeds, will help to progress Australia towards a stronger national infrastructure for an integrated research libraries and information network. Metadata will play a small but vital part in this transition.

The potential to integrate services is very strong. The subject gateways have established a new trend in collaboration, which can only revitalise library service provision in the future. The only constant in the changing environment of provision of bytes has been the use of metadata itself. Everything else is subject to change.

Bibliographical References

[ADT] Available: www.library.unsw.edu.au/thesis/thesis.html. 22 September 1999.

[AGRIGATE] Available: www.agrigate.edu.au/. 22 September 1999.

[ALG] Available: www.nla.gov.au/libraries/. 22 September 1999.

[AVEL] Available: <http://avel.library.uq.edu.au/>. 22 September 1999.

[BESTPRAC] Available: www.nla.gov.au/meta/sg/criteria.html. 22 September 1999.

[BSPARCS] Available: www.asap.unimelb.edu.au/bsparcs/bsparcshome.htm. 22 September 1999.

[BUBL] Available: <http://bubl.ac.uk/link/sci.html>. 22 September 1999.

[EdNA] Available: www.edna.edu.au/. 22 September 1999.

[EEVL] Available: www.eevl.ac.uk/. 22 September 1999.

[GRANT] *A Look at the horizon - Is it a sunset or a sunrise ?*, Carl Grant, President, Ex Libris (USA) Inc., at Projections / Reflections, 10th ALIA National Library Technicians Conference, 8-10 September 1999

[IF] Available: www.ukoln.ac.uk/interop-focus/about/. 22 September 1999.

[IMESH] Available: www.imesh.org/. 22 September 1999.

[ISAAC] Available: <http://scout.cs.wisc.edu/research/index.html>. 22 September 1999.

[METACHEM] Available: <http://metachem.ch.adfa.edu.au/>. 22 September 1999.

[NCSU] Available: <http://my.lib.ncsu.edu/>. 22 September 1999.

[RAAM] Available: www.nla.gov.au/raam/. 22 September 1999.

[SG] Available: www.nla.gov.au/initiatives/sg/. 22 September 1999.

[SIGNPOST] Available: <http://scout7.cs.wisc.edu/pages/00010715.html>. 22 September 1999.

[TOOLKIT] Available: www.imesh.org/toolkit/. 22 September 1999.

[UKOLN] Available: www.ukoln.ac.uk/metadata/cld/. 22 September 1999.

[UW] Available: www.lib.washington.edu/resource/help/MyGateway.html. 22 September 1999.

[ZEDWEB] Available: www.dstc.edu.au/Research/Projects/zedweb/. 22 September 1999.