The Internet Resources Project: an Exercise in Collaboration

Nathalie Schulz
Librarian, Technical Services
Griffith University Library
N.Schulz@mailbox.gu.edu.au

Sarah Fredline
Library Systems Manager
Queensland University of Technology Library
s.fredline@qut.edu.au

With the assistance of Ann Huthwaite, Brian Reis, Alice Steiner, and Cheryl Stevens.

Abstract:
This paper describes a collaborative effort to select and describe free Internet resources by Griffith University and Queensland University of Technology (QUT) Libraries in Brisbane, Australia. After considering the options available, the Libraries decided to trial selection and description of free Internet resources by Reference staff using a web input form, with some records upgraded to full cataloguing.

InfoQuest

In 1998 Griffith University Library and Queensland University of Technology Library began a collaborative project to provide bibliographic access to Internet resources by creating a subject gateway called InfoQuest (Griffith University Library and Queensland University of Technology Library, 2001). InfoQuest used ROADS (ROADS Project Team, 2001) software, and was offered to clients as a standalone web-accessible database. As the Dublin Core metadata standard was in a state of flux at the time, a locally designed record structure was used, with the intention that it would be easily mappable to other schemas such as Dublin Core in the future. InfoQuest was a hybrid subject gateway, and included records for some print items, as well as Internet resources. Records for both subscription items such as databases, and freely available materials were included. It had a very broad subject scope, embracing all subjects studied at either university, but it was not intended to be comprehensive. Responsibility for resource selection and description was shared out between the two institutions, and Reference staff entered details directly into the database using a web form. InfoQuest itself was seen as primarily a Reference Services project, with little involvement from cataloguing staff in Technical Services. By mid 2000 InfoQuest contained about 4000 metadata records.

InfoQuest was used heavily, however examination of search logs revealed that non-librarian users didn't really know what they could expect to find in it. Despite explanatory text detailing the nature of the database, users searched for very specific items like the "mating habits of fairy penguins in Antarctica", and became frustrated when they received no hits. There was also confusion amongst library staff as to the overlap between InfoQuest and the catalogue.

InfoQuest was formally reviewed by an external consultant in June 2000. Joint meetings with relevant staff from both institutions were also held, and InfoQuest was recognised as a valuable learning experience which had established a favourable climate of cooperation. The decision was made to explore other options, and as a result a working party was formed and charged with examining possible models and metadata standards for describing free Internet resources. The working party consisted of four members, two from each institution, with equal representation between the Systems area and Technical Services.

Options for Describing Free Internet Resources

Since the emergence of the Internet as a major medium of communication, libraries worldwide have grappled with the issue of whether or not to select and describe free Internet resources for their clients. Some, such as Baruth, make the case against cataloguing of Internet resources, arguing that it is too labour intensive and the resources are too numerous (Baruth, 2000). Gorman is one author who has suggested that the bibliographic control of electronic resources be based on their "value", with only a very small proportion receiving full cataloguing, some receiving Dublin Core records and the majority retrieved by search engines (Gorman, 1999).

The Working Group identified four models for describing free Internet resources in a collaborative environment, two catalogue based and two external repository based. The InfoQuest model of a shared stand-alone database available to clients was not considered.

Catalogue as Repository

One of the primary advantages of using the catalogue as a repository for information relating to free Internet resources is that the number of information sources for clients is reduced and resources are integrated with resources in other formats, and those electronic resources that have already been catalogued. In addition, there are no additional software costs and the sophisticated searching interfaces of the library system, enhanced by use of standards, facilitate retrieval. The existing expertise in resource description of Technical Services staff can be used. A disadvantage of using the catalogue is that it is not marketable in the same way as a standalone service such as InfoQuest.

Model 1 – Full records

In this model all electronic resources selected for description are catalogued fully into the respective library catalogues of the two institutions. Reference staff identify resources, and advise Technical Services staff who perform the description. Selection activity is coordinated, but other potential for collaboration is limited to sharing of full catalogue records. A disadvantage of this model is the possibly serious workflow implication for Technical Services staff.

Model 2 – Full and brief records

In this model Reference staff create brief records for resources, and mark resources of lasting value as requiring enhancement to full cataloguing. The brief records are created either by use of a web form or by direct input into the library system by Reference staff. The use of a web form and shared selection allow great potential for collaboration. There are fewer workflow implications and cost considerations with this model, but brief records may not be as retrievable as full records.

Repositories Outside the Catalogue

Model 3 – Shared repository

In this model a new database is developed as a repository for brief records. Reference staff from both institutions create brief records using a web interface. This repository is not searchable by clients, but is rather a source of records for inclusion in other products. The potential for collaboration is high, including shared development of the repository and direct sharing of brief records. This model allows content to be reviewed easily because records exist as a discrete subset, and allows alternative subject hierarchies to the catalogue. In this model, some items are given full cataloguing in addition to residing in the repository. A disadvantage of this model is the paucity of suitable software at an acceptable cost.

Model 4 - CORC

CORC (Cooperative Online Resource Catalog)(OCLC, 2001) is an international database of cooperatively created brief and full records for web-based electronic resources. The University of New South Wales is one Australian library that has participated in CORC (UNSW Library, 2001). Records can be exported in MARC or Dublin Core format, and CORC can also be used to create pathfinders. In this model CORC is used as the mechanism for creation and storage of brief records, and as a source of records for inclusion in the catalogues. By definition CORC represents collaboration on an international level, but the

collaboration potential between the two institutions is limited. CORC has a significant cost, and another concern is possible slow response times in using an Internet based system.

Standards

The group recommended that there be a two-tier approach to cataloguing non-serial Internet resources. Resources of lasting value would be given full cataloguing using AACR2 and LCSH, with records located in the library catalogue. A brief record standard would be used for resources of a more ephemeral nature, and this standard would be compatible with the MARC format. Where these records were stored would depend on the Model chosen. These brief records would have very broad subject headings assigned, and there would be no attempt to construct an in-house thesaurus. The brief record standard recommended was a subset of Dublin Core metadata in its unqualified form.

Internet Resources Project

At a meeting in early 2001 staff from both institutions met and made the decision to trial Model 2 (including a web form). Another working group was formed, consisting of the original four members, plus two Reference staff, one from each institution. The group decided on the Internet Resources Project (IRP) as a working name. The project had a number of facets: development of a collection development policy; detailing of the brief record standard and conversion to MARC; writing of user input guidelines; and construction of the web input form.

Collection Development Policy

A collection development policy for free Internet resources was developed, and can be summarised as follows:

Free Internet resources must meet the same criteria as those that are applied universally in the selection of materials. These include:

- Authorship or Authoritativeness (i.e. it is necessary to identify the credentials of the author/s and the publication to establish the authority and credibility of the information);
- Content or Coverage (i.e. it is necessary to obtain an overview of the coverage of the information source);
- Provenance (i.e. it is necessary to consider the origin of information);
- Accuracy or Bias (i.e. it is necessary to determine whether the information is factual or whether it contains opinion or bias).

In addition to universal criteria, criteria specific to the resource format need to be applied. These include:

- Design; Image Quality; User friendliness; Timeliness or Currency; Permanence, Durability or Reliability; Completeness; Quality of links to other sites; Value-added utility beyond the print version; Originating domain of the site; Downloading capability; Ready availability of needed add-on software for access; Uniqueness; “Reverse links”, i.e. the quality and quantity of sites that link to the resource.

Resources will only be included if the information they provide is directly relevant to the educational programs of each institution, and are at a level appropriate for higher education. Web-based resources that attract a subscription cost are governed by the policies at each institution. If the Internet resource requires the user to register, use special software, or engage with advanced WWW technologies (e.g. Java applications), mention of this will be made in the catalogue record. Preference will also be given to websites that are presented in English, though resources in other languages will be included as required for particular programs. The origin of the resource is not of consequence to selection, providing that subject matter and quality are fitting.

Examples of various types of Internet resources that are acceptable to both Libraries' collection policies include:

- Organisation web pages (e.g. National Gallery of Australia)
- Newspapers (e.g. Art Daily)
- Subject gateways (e.g. World Wide Arts Resources)
- Personal pages (e.g. artists such as Judy Chicago)
- Collections (such as images e.g. Picture Australia)
- Scholarly discussion lists (e.g. ARTCRIT)
- Other Grey Literature resources (e.g. research & technical reports, conference papers, annual reports, dissertations etc)
- New media works (e.g. web-based media)

Choosing Full or Brief Records

Another aspect of the collection development policy is determining whether or not IRP records should be given full cataloguing. Three aspects will determine whether records for free Internet resources should be left in brief form or given full cataloguing using AACR2 and LCSH by cataloguers. An exception is electronic journals, which are always given full cataloguing.

1. Provision of standardised access points

One example of a case where the provision of standardised access points may be an important factor is when the issuing corporate body is of paramount importance (e.g. official reports of committees and commissions; conference proceedings; primary legal materials). Corporate body headings are frequently complex and if the same form of heading is not used for all works issued by the body users may not retrieve all of them. Another case where standardised access points is important is when Library of Congress subject headings are required to provide access to complex concepts.

2. Integration with the rest of the collection

Examples of cases where integration may be an important factor include: the resource belongs to an important series and other titles in the series in print form have been given full cataloguing; other similar types of publication are in print form and have been given full cataloguing (e.g. Education Queensland curriculum material); and, the resource supplements or adds value to an existing print resource and consistency in description is important (e.g. a student's manual accompanying a textbook).

3. Content of the resource

Examples of cases where the content of the resource may be an important factor include: the resource has lasting significance (e.g. important government reports); the resource is important because of its coverage (e.g. uniqueness, local significance, and/or relevance to the needs of clients); the resource is a particular type of material (e.g. map, music) that may benefit from specialised treatment; and, the resource is an important reference tool.

Brief Record Format and Conversion to MARC Format

A brief record format was developed, based on unqualified Dublin Core, and a crosswalk for conversion from Dublin Core to MARC was devised. Some additional fields were added, for example fields for the name of the selector, and for an indication that full cataloguing is required.

Figure 1 is a table showing the fields used and corresponding MARC fields for each library catalogue.

Figure 1: DC fields and crosswalk to MARC

Note: Fields marked with an asterisk are mandatory on the web form.

DC ELEMENT	FIELD CONTENT	MARC FIELD FOR QUT (INNOPAC)	MARC FIELD FOR GU (GEAC ADVANCE)
***Title**	Title of the resource	245 00$a	245 00$a
Repeated title field	Alternative title	246 33$a	246 33$a
Creator/Contributor	Entities responsible for the creation of the content of a resource, in the form of personal or corporate names	720$a	720$a
***Broad subject heading** (repeatable)	Subject headings assigned from ASCED list	958$a	958$a
***Description**	Summary of resource content	520$a	520$a
Keyword (repeatable)	Additional terms to be assigned if not included in the summary	653$a	653$a
Publisher	Entity responsible for making the resource available	260$b	260$b
Date	Used for resources with a specific publication date, not ongoing resources.	260$c	260$c
***Identifier**	URL for the resource	856 40$u	856 40$u
***Language**	Language of the resource	008 (cp 35-37)	008 (cp 35-37)
Course code	Used by Griffith University only	Not used (excluded on load)	091$a
***Institution of selector**	QGU or QUT	040$a	040$a
***Name of selector**		950$a	091$f
Date of cataloguing	System assigned	008 (cp 00-05)	008 (cp 00-05)
Coverage (repeatable)	Geographic coverage of the resource	522$a	522$a
Full cataloguing required		951$a	951$a
IRP number	System assigned	035$a	035$a
GMD	System assigned	245$h[computer file]	245$h[computer file]

The Australian Standard Classification of Education (ASCED) (Australian Bureau of Statistics, 2001) narrow codes are used for the broad subject heading field. This scheme was chosen because it is Australian, authoritative and will be maintained by the Australian Bureau of Statistics. The ASCED descriptors will not be used by clients for retrieval, but rather to enable grouping of resources for review, or creation of pathfinders.

Input Guidelines

Input guidelines were developed to assist the Reference staff with the creation of brief records. These guidelines ensure that there is no ambiguity as to the content of fields or the format.

Examples of the IRP Input Guidelines include:

- Identifier/URL: Give the full form and record accurately: e.g. http://www.imaginary.com NOT www.imaginary.com.

- Title: Take the title from the most appropriate source. Usually this will be the title screen (e.g. the home page), but in some cases it could be a source like the HTML header. In case of doubt, give the title that the user would expect to be used.

- Creator/Contributor: Record personal names in inverted form (e.g. Smith, James). As a general rule record corporate names in the form found on the resource. The correct form of the name can be recorded if known, but adherence to AACR2 standards is not required. However, the word "Department" should always be abbreviated to "Dept." to be consistent with AACR2 usage.

- Description: The descriptive summary, along with the title, should provide enough information for users to decide whether to access the resource. Include terms or phrases that provide subject access to the resource. Include terms or phrases that provide information about the unique features of the resource (e.g. intended audience, format, etc.).

Web Input Form

The web form was designed to be as simple as possible, and is written in HTML and Javascript. Javascript is used to provide an authority list of the Broad subject headings; some basic error checking; and to display “Help” messages containing the input guidelines.

On submission of the web form a CGI program, written in Perl, processes the data into a defined format and appends it to a text file. On a regular basis this text file is retrieved, at which point the records are archived and the working file is cleared. The file containing the retrieved records is then processed using MarcMaker (Library of Congress, 2001) to produce MARC Communications Format. MarcMaker is a freely available program from the Library of Congress.

The records in MARC Communications Format are loaded into the catalogues using specially designed load tables which adapt the records slightly for local requirements.

If a record is flagged for full cataloguing, upon submission of the web form an email is sent to cataloguing staff at the requestor's institution. Cataloguing staff upgrade or replace the record in their own local catalogue. Upgraded records are periodically exported and loaded into the other library catalogue, where they bump the brief record.

Collaborative Aspects of the Project

A key aspect of this project has been its collaborative approach, which builds upon a long-standing cooperative relationship. For many years Griffith and QUT have offered special reciprocal benefits to each other's student populations.

The Libraries had been looking for a project in which they could collaborate when the InfoQuest concept was first developed, and despite the fact that the project was not entirely successful, the continuation of the collaboration when the project changed direction was never questioned. In fact, during the project the two libraries and their respective parent divisions, (the Division of Information Services at Griffith University and the Division of Information and Academic Services at QUT), significantly increased collaborative activities and entered into a formal agreement to support collaboration for mutually beneficial outcomes.

At the time of writing, 50 records have been identified, entered into the web form, and uploaded to both catalogues. The work has been limited to the members of the working party but the intention is that all Reference staff will identify resources and create records using the web form. The resulting records will be converted by Technical Services staff at one or other Library on a rotating basis, and the resulting MARC file loaded into the catalogue separately by each Library.

The climate of the Internet Resources Project is one where staff know each other and have common goals and visions, and as a result there has been virtually no conflict between the requirements of each institution. There are some differences in the technical requirements of each library catalogue system, but these have been easily accommodated.

In many ways the collaboration between Technical Services/Reference Services/Systems has been the more enlightening experience. Between institutions but within each functional area there was an immediately obvious strong alignment of interests and perceptions. For example, the two Technical Services staff members were extremely interested in and felt strongly about adherence to standards, while the two Reference staff members were most interested in client usability and agreed strongly on ways to achieve this. From the Systems perspective, the advantage was that coding and programming have been done once rather than twice, and the work was able to be allocated where the expertise was available. All involved have said that they have valued the opportunity to work on such a project, and have benefited from learning more about others' perspectives.

The only negative aspects of the collaboration have been the difficulties of physically gathering the participants together from diverse parts of the city for meetings, and that the unavoidable staffing issues, and competing sectional and institutional priorities have been compounded across six sections and two institutions. However, both libraries as large, multi-campus operations are all too used to these types of complications, and the goodwill among the participants has resulted in a valuable experience in collaboration.

Australian Bureau of Statistics. *1272.0 Australian Standard Classification of Education (ASCED).* Canberra: ABS, 2001. Online. Available: http://www.abs.gov.au/ausstats/abs%40.nsf/66f306f503e529a5ca25697e0017661f/f501c031bd9ac9c5ca256aaf001fca33!OpenDocument. 1 Oct. 2001.

Baruth, Barbara, "Is your catalog big enough to handle the Web?", *American Libraries*, Vol. 31, no. 7, Aug. 2000. pp. 56-60.

Gorman, Michael, "Metadata or cataloguing? A false choice", Journal of Internet Cataloguing, Vol. 2, no. 1, 1999, pp. 5-22.

Griffith University Library and Queensland University of Technology Library. *InfoQuest: Your Guide to Finding Information.* Hp. 2001 [last update]. Online. Griffith University and QUT Libraries. Available: http://roads.itc.gu.edu.au:8880/. 10 Nov. 2001.

Library of Congress. *MARC Specialized Tools*. Washington, DC: Library of Congress, 2001 [last update]. Online. Available: http://lcweb.loc.gov/marc/marctools.html. 1. Oct. 2001.

OCLC. *OCLC CORC*. USA: OCLC, 2001 [last update]. Online. Available: http://www.oclc.org/corc/. 1. Oct. 2001.

ROADS Project Team. *ROADS.* Hp. 2001 [last update]. Online. ROADS Project Team. Available: http://www.ilrt.bris.ac.uk/roads/. 1 Oct. 2001.

UNSW Library. *CORC Project (Cooperative Online Resource Catalog).* Sydney: UNSW Library, 2000 [last update]. Online. Available: http://www.library.unsw.edu.au/~eirg/corc/corc.html. 1 Oct. 2001.