

Sydney Symposium 2004

## E Publishing

Sharon Bradshaw

### ABSTRACT

*ProQuest Information and Learning is well known as an aggregator of current periodical journals. Our databases cross a wide range of subjects, including Business, Education, General Reference, Health & Medical, History, Literature, and Social Sciences. In addition to current content, the company has a long-standing history of microfilming content, and in recent decades has been creating retrospective digital conversion of content.*

### A LOOK AT PROQUEST INFORMATION AND LEARNING'S PRESERVATION ACTIVITIES

I'd like to run through a few of ProQuest Information and Learning's preservation activities, from microfilm to digital mediums. Specifically I will looking at three retrospective digital conversions, including Early English Books Online, the Historical Newspapers project, and the Dissertations Archiving & Access program. The paper reviews advantages of electronic indexing, extended opportunities for scholarly

research and the renewed application of these resources. It will also discuss commercial publishing programs, and provide examples of collaborative approaches to digitisation projects.

### HISTORICAL CONTEXT

Founded in 1938 by Eugene Power, originally known as University Microfilms (later known as UMI). Eugene Power pioneered microfilming technology, and was approached by the British Library to microfilm their rare books collections, which were under threat from the WWII bombing of England. This collaboration resulted in our first commercially available product the Early English Books series on microfilm, which included the earliest works printed in the English language. The UMI microfilm vault now contains more than 5.5 billion page images, delivers 500 years of information and is the largest commercially available microform collection in the world. Today this vault is

the foundation for our retrospective and current digital conversion.

In 1985 the company' expanded into electronic access, initially with CD-ROM products, followed by web-based services. The ProQuest Information and Learning suite of products, which includes current and retrospective content, is accessed by millions of higher education, school, and public library users globally.

#### EARLY ENGLISH BOOKS

Early English Books includes images of titles indexed within *The Short-Title Catalogue* (Pollard & Redgrave, 1475-1640); *The Short-Title Catalogue II* (Wing, 1641-1700); *The Thomson Tracts*, a compendium of broadsides on the English Civil War printed between 1640 and 1661; and *The Early English Books Tract Supplements*.

Microfilm conversion is still ongoing. Microfilmed works are released in units. We produce one unit of Early English Books I material and four units of Early English Books II material each year. Due to the rarity of the remaining material we estimate that it will take 5-10 years to complete the microfilm project. We are currently filming on four continents and managing relationships with over 125 contributing libraries.

Early English Books Online (EEBO) is the digitised version of the Early English Books series. Launched in 1998, the collection includes works printed in English and its colonies and any item printed anywhere in the world in English between 1473 – 1700. When completed EEBO will be the largest database of its kind with 125,000 titles comprising 22,500 pages.

The collection covers a variety of subject, such as:

- Literature - *Shakespeare, Spenser, Marlowe, Milton*
- Science - *Galileo, Harvey, Mercator, Vesalius, Newton*
- Religion - *Calvin, Luther, Knox, Laud*
- Exploration - *Drake, Raleigh, Hakluyt*
- Philosophy - *More, Bacon, Hobbes,*

*Locke*

- Politics - *Macchiavelli, Prynne, Lilburne*
- Law - *Littleton, Brownlow, Fitzherbert, Coke*

During initial development phase, we worked with librarians, and faculty members to access needs, determine key features, and look at how digital access could enhance primary research. The aim of the product / project was to provide a single source for research on Early English Books, including searchable bibliographic citations, full-page representations of all images and ASCII – encoded text (which is an example of a collaborative approach to content creation).



#### HISTORICAL NEWSPAPERS – NEW YORK TIMES

ProQuest Historical Newspapers is a digital archive of key US newspapers from 1<sup>st</sup> issue, to 1990s or 2000s, end dates will be extended year upon year. Titles currently available under this program include The New York Times (1851-2001), The Wall Street Journal (1889-1987), The Washington Post (1877-1988), The Christian Science Monitor (1908-1991), The Los Angeles Times (1881-1984) and the Chicago Tribune (1849 – 1984). We are now looking to broaden the geographical scope of the program, and introduce collaborative publishing agreements.

Each title includes cover to cover imaging of all photographs, advertising and articles. Titles sit on the same product

platform (server), so searches can be run across historical titles or current newspaper files. In the case of the newspaper products, they also include fully searchable ASCII text. Images are sectioned into Article view, or Page Map view, which provides a broader context of position within the page.

The Historical Newspapers program is born out of material within our microfilm vault. Digital images are captured from microfilm. The digitisation process includes de-skewing and cleaning the microfilm images, which improves the accuracy of the OCR process. Microfilming each of these titles is still ongoing.



## DISSERTATIONS ARCHIVING & ACCESS PROGRAM (DAAP)

Dissertations are published under the UMI® imprint. Microfilm Abstracts, the forerunner of Dissertation Abstracts, with 15 dissertations from five universities was published in 1939. Dissertation Abstracts now includes; titles from 1861 to the present, over 2 million citations, 1.8 million of these are available as dissertation copies and 450,000 plus titles are in digital (PDF) format. Approx 700 US institutions submit to the dissertations program, which equates to about 55,000 titles each year.

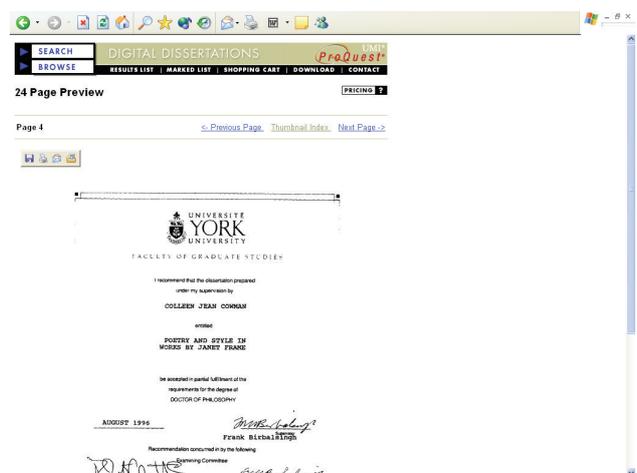
The UMI digital dissertations collection is the official Library of Congress offsite repository for US dissertations and theses in electronic format.

The aim of the Dissertations program is to provide a permanent, accessible record of an institution's contribution to higher education; there three streams to the Dissertations program Publishing, Preservation, Access.

*Publishing*, current or retrospect dissertations are submitted in paper or digital format. Bibliographic information is created and added to the Dissertation Abstracts database.

*Preservation*, the work is microfilmed and stored within the UMI vault.

*Access*, where a work has been submitted in paper, a digital copy is created from the microfilm. Works can be located and ordered (in paper or PDF) via bibliographic information, which is accessed via the Dissertation Abstracts database or the Dissertations Express service. Or an institutions own digital copies are available through a Current Research @ site.



## AN EXAMPLE OF COLLABORATION

### EEBO

Early English Books Online (EEBO) currently only provides searchable citations, which link to images of the works themselves. The product does not contain searchable text, and it wasn't a viable option for ProQuest to create searchable text, as there was no automated means for capturing text. The Academic community saw searchable text as the next stage of researching these works. The Text

Creation Partnership (TCP) was created to capture the EEBO ASCII searchable text. The TCP is a collaboration between the academic community and ProQuest Information and Learning. The aim of the TCP is to provide ASCII searchable text for the EEBO images. Members of the TCP contribute an annual membership fee to fund the conversion, the project is targeting to convert 25% in the first 5 years, and members will share in joint ownership and access at the end of that period, with the hope that the project will become self-funding.

#### *Historical Newspapers*

We are looking to expand the Historical Newspapers program, but realistically need to review investment, against market opportunity. Now that we have a successfully method for publishing these titles, and an established market, we'll look to partner with state and national institutions and / or publishers in order to create an historical archive of regionally based newspaper content. We are currently considering Australian newspaper opportunities.

#### *Dissertations*

By working on retrospective conversion of dissertations from paper or microfilm, to digital formats we've become an outsourcing arm for an institution thereby reducing physical and electronic storage space and costs, and expanding access to forgotten research.

The dissertations service is the designated off-site digital repository for the Library of Congress. This represents the first time the Library of Congress has recognized an external site as a repository for a key collection. The Library of Congress uses Dissertation Abstracts database for access to US dissertations materials. There is a contract in place for ProQuest to provide electronic copies of all materials to the Library of Congress, should ProQuest no longer provide access.

Authors receive royalties for any paper, microfilm or PDF copy ordered through the service.

#### *Advantages of electronic access and indexing*

Online availability of each of these products presents some common advantages of moving into a virtual environment, including global access to titles, consolidation of titles - bringing the collection together in the one place. Virtual access provides a safe environment for display and protects original works and promotes the rediscovery of original material.

#### *EEBO*

Online access to EEBO introduces advantages not previously available with microfilm collection including;

- A searchable electronic index of STC I and STC II, and a detailed index of illustrations.
- The option to skip pages, rather than scroll.
- The ability to take images outside of the database, via a durable / permanent link facility.
- The ability to resize images, so that users are able to focus in on detail, such as printer's marks, and illustration detail.

#### *Historical Newspapers*

The Historical Newspapers material includes searchable ASCII text (searchable – not available to view). Prior to the project it was not possible to search across the text of newspaper articles electronic.

Articles are segmented so that a user can search for a particular article reference, or look for material under the following article categories.

Article, Display Ad, Obituary Death Notices, Real Estate Transfers, Wedding Engagements, Editorials / Opinions, Comics, Reviews, Classified Ads, Birth Notices, Weather Reports etc.

#### *Dissertations*

Dissertation Abstracts, and digital (PDF) copies provide a global audience for the material and a centralized access point for dissertation content.

Bibliographic (MARC) records are available for each institution's titles, which provides easier locating of local materials.

#### EXTENDED OPPORTUNITIES FOR SCHOLARLY RESEARCH AND RENEWED APPLICATION OF RESOURCES

Aside from the deterrents that microfilm materials present, electronic access is taking the content to a new level.

#### *EEBO*

Previously the Early English Books series was purchased for and used by a select group of academics, or researchers. The electronic access has extended the use of the content. In a teaching environment, sixteenth and seventeenth-century texts have become required reading. In a research environment, scholars are now able to view 1<sup>st</sup> and 2<sup>nd</sup> editions side by side on screen. Or using the TCP text, can search for the references to specific works across works.

#### *Historical Newspapers*

- Users can create an historical time line of events.
- Cross match events across multiple newspaper sources.
- Uncover previously unfound, with the ability to search ASCII text.

#### Dissertations

The Dissertation Abstracts database is the most widely used PhD/master's thesis research tools in the world. Through the database research topics are presented to an international audience. Digital access to titles revitalizes research by making older material accessible to new audiences.

#### *Commercial Publishing programs*

Products come with a high price tag! Most are purchased on a perpetual access model, so that the institution or organisation owns the content. There is a license agreement, which accompanies any purchase to ensure that the purchasers and publishers meet commitment to ongoing access.

As a commercial company we prepare business models for projects like these,

and then have a sales channel to ensure that we meet the costs of bringing the product to market. Once the content has been digitised the content then presents us with new opportunities, we can 'Slice and dice' content and present it within our various publishing streams, curriculum based products are a good example of this.

As technology changes, equipment costs go down (and up) and techniques change in line with new opportunities. For example ORC technology continues to evolve or the skill set of people within the professional also develops in line with the no. of projects now underway.

#### SUMMARY – CONSERVATION AND ACCESS

##### *Digital continuity: access, preservation*

Microfilm, has been and continues to be a valuable medium for the company, as a preservation medium and a source for digitisation of content.

We have been innovative with the digitisation of materials; this has ensured access is always intertwined with preservation.

Digital content is moving into the next phase, new technology and protocols present our user communities and us with new avenues for access and delivery.

#### AUTHOR BIOGRAPHY

Sharon has been involved with electronic database products since their early inception; and has first hand experience with the adoption and the direction the environment has taken, both in Australia, New Zealand and overseas. Her current role Marketing, Training & Support Manager, ANZ involves marketing and managing products from sale - to implementation - to successful use within each education institution. She is currently undertaking a Graduate Certificate in Museum Studies to add theoretical knowledge to a personal and professional interest in the provision of electronic information to cultural institutions.

#### **Sharon Bradshaw**

Marketing Training & Support Manager, ANZ  
ProQuest Information and Learning  
PO Box 181, Drummoyne NSW 1470  
[email:support@anz.proquest.com](mailto:support@anz.proquest.com)  
OR [sharon.bradshaw@anz.proquest.com](mailto:sharon.bradshaw@anz.proquest.com)  
AUST Free call: 1 800 11 6660

NZ Free Call: 0800 443429  
Mob: +61 0414 669 026  
Tel: +61(2) 99116660  
Fax: +61(2) 99116652