EThOS: progress towards an electronic thesis service for the UK

The EThOS (Electronic Theses Online Service) project is building on previous e-thesis initiatives, and co-ordinating the work of some of the key players in the UK to develop a service for finding, accessing and archiving digital copies of doctoral theses produced in UK higher education institutions.

Key issues for the project are the development of a sound financial basis for a successful service, the provision of advice needed by authors and university staff on handling intellectual property rights, and the protection of legitimate needs for confidentiality. EThOS will also establish workable and standards-based procedures for populating e-thesis repositories with current and retrospectively acquired digital versions of theses and associated metadata. These developments must also fit with universities' own internal admini-strative arrangements and regulations.

The project aims to deliver an e-thesis infrastructure that is both technically and financially sustainable, together with a full supporting toolkit of guidance, standards and procedures.

JILL RUSSELL
on behalf of the EThOS project team

Introduction

PhD and other research theses produced in UK universities have long been recognized as a rich source of research findings, yet for anyone wishing to read them, they are often the hardest material to obtain in full text. There has been a steady growth in interest amongst researchers in improving the speed and convenience of access to theses. This practical necessity chimes well with the expectation or requirement from an increasing number of funders (such as Research Councils UK[1] and the Wellcome Trust[2]) to make research findings publicly available via open access repositories.

Accessibility of theses in the UK: the current situation

Metadata for theses that have been accepted for research degrees in the UK is fairly readily available. Subscribers to the *Index to Theses*[3] can search a database of thesis metadata with excellent coverage of the output of UK higher education institutions (HEIs). This is the most comprehensive discovery tool for UK theses but they are also traced through other routes such as library catalogues, or via university web sites or special interest listings, or recommendation to researchers by colleagues or supervisors.

However, having established the existence of a thesis, the present arrangements for actually reading it are often slow, expensive and inconvenient, based on manual methods and media established some decades ago. Although there are some good commercial e-thesis publishing services, these cover only a small proportion of UK work and can be expensive.

On an individual basis, some authors of theses do make their own work available, sometimes freely through personal or employers' web pages and institutional repositories (IRs), sometimes through commercial services. Internationally, Australia, the USA and many parts of Europe are developing co-ordinated e-thesis services (also known as Electronic Thesis and Dissertation or ETD services) and, arguably, we in the UK have lagged behind.

Where the UK does have a great strength is in the service provided by the British Library Document Supply Centre (BLDSC) which acts as a central focus for receiving requests for UK theses, and providing copies or loans through its microfilm production and printing facilities, or by

co-ordinating requests via the inter-library loan system from individual university thesis collections. However, photocopy and microfilm technology is no longer popular, inter-library loans restricted to use within library buildings are not convenient, and in an age of desktop delivery for other types of information, readers expect an equivalent level of service for theses.

EThOS aims to combine the strengths of the present UK system, i.e. good metadata and a good central requesting service working with a network of libraries providing theses, with the convenience of modern delivery methods and a more streamlined process for gathering metadata, to provide a one-stop service for discovering and accessing theses.

The case for e-theses

The case for ETDs has been made many times before, often in the same breath as that for self-archiving of papers. Case-studies are presented annually at the International Symposium on Electronic Theses and Dissertations, organized by the Networked Digital Library of Theses and Dissertations (NDLTD).[4,5] To summarize briefly, there are many potential gains for all parties concerned (authors, researchers, supervisors, library/archive staff, university administrators, funders of research) from making theses easily accessible. These include:

- raised profile for individuals, universities and funders from wider circulation and citations of their work
- a more visible virtual community of researchers
- metadata visible to search engines (in addition to thesis databases)
- possibility of flexibility in expressing research results
- greatly reduced delivery time for theses, and time saved by eliminating extra visits to library to consult restricted volumes or use microfilms
- improved customer satisfaction, and use of the service is easier to measure
- savings on cost of binding and printing and possibly storage
- savings on costs of lending and handling theses.

Why EThOS?

Among the EThOS partners are the British Library and the libraries that provide the highest number of theses currently requested via BLDSC. The partners and project team also include institutions and individuals who are already at the forefront of UK e-thesis developments. Copeland[6] summarizes the achievements of past projects and how the balance is tipping in favour of achieving the remaining work needed. With its mix of key players and expertise, and drawing on the experience of previous initiatives, EThOS is in a strong position to pull together the work of diverse organizations, and bring the service into being. A number of other libraries and organizations who are not partners in EThOS have already expressed a strong interest in joining the service once it is established, both as potential users of the service and as suppliers of theses.

EThOS differs from commercial ETD services in two key aspirations: 1) to make not only current theses available, but also to generate a critical mass of retrospective theses, covering the areas most regularly requested via the present inter-lending system, and 2) to make downloads of UK theses free to researchers at the point of use.

How do we get there?

There are, inevitably, one or two thorny problems along the path to a full ETD service. Much practical work has been done on implementing technical, metadata and other standards, and on the practicalities of digitizing theses and entering them into a repository. Two key issues that will determine the success of the project are intellectual property rights (IPR) and financial sustainability.

Intellectual property rights

The ownership of IPR varies greatly from institution to institution, and may be held by the author of the thesis, the university, a sponsor of the research or some combination of shared rights. The key issues are summarized in a JISC briefing paper[7], and the project is making good progress with producing guidance for authors and universities on how best to widen access to theses while still protecting these rights, through licensing

models and fair use agreements. The guidance and agreements can be fairly readily applied to current theses, and staff involved with handling theses will need to understand their local circumstances.

A separate set of procedures is also under investigation for handling the retrospective digitization of theses. Again, the situation varies from place to place, as some existing agreements between university archives/libraries and rights holders will allow retrospective digitization of their work, but in other cases it may be necessary to seek permission. In a minority of cases there are in addition legitimate needs for confidentiality and/or protecting commercial and other sensitivities. A further area that must not be overlooked is where significant third-party material has been included in a thesis. Use of this material may be legitimate for the purpose of the thesis as an examination document, but reproducing it on the web would stretch most interpretations of fair use.

Getting more for our money

The present thesis loan or copying arrangements incur costs in many ways, with a mixture of upfront and hidden costs. These include:

- direct charges by BLDSC or individual libraries for inter-library loans or copies
- staff time spent retrieving volumes from library archives or stores each time a thesis is requested
- charges for heavy tomes to be transported securely around the country
- staff time involved in handling thesis loans and requests, including in many cases time spent contacting thesis authors for permissions
- reprographic costs
- above all, the hidden costs to researchers in time spent waiting for theses to be supplied (which can sometimes take several weeks), and extra visits to library buildings to consult restricted volumes or to use microfilm reading facilities.

Providing an e-thesis service will of course incur other costs in obtaining metadata and content, administering the service, and the maintenance of the platforms on which metadata and content will reside.

EThOS aims to redirect the time, money and effort needed to support the present cumbersome thesis lending and copying operation into supporting a digital service that will add value by not only improving the convenience of access to theses but also by enabling more frequent or concurrent consultations of theses without libraries incurring further costs. The business model for this will be a key outcome of the project. EThOS will recommend models that can save HEIs money overall, and can take advantage of economies of scale offered by standardized processing, maintenance and preservation, yet allow HEIs and researchers some flexibility of choice.

The aim is to make access to the centralized database of metadata (i.e. the discovery tool) and downloads of theses free to end-users. Researchers who prefer to receive theses delivered on CD/DVD or paper will have the option to do so, charged on a cost-recovery basis.

Institutions will have a choice of hosting their own e-thesis repositories or (particularly in cases where an institution does not have its own IR) using a central server as the host platform for digital versions of theses, or may choose a combination of the two, so long as the metadata is contributed to the central database, and accessibility of the full text meets the agreed standards. Centralized hosting will be charged on a cost-recovery basis.

The largest cost element in supplying a thesis to a researcher is in the initial stage of acquiring the thesis and making it available on the web. Cost per use is then minimal. In an e-thesis environment, acquiring theses 'born digital' is relatively cheap, but retrospectively digitizing older theses is expensive.

Acquiring current e-theses

There will be two broad options:

a) HEIs who already have an IR (therefore have already invested in the startup and maintenance costs) may opt to make local arrangements for populating their own e-thesis repositories, complying with EThOS standards.

b) HEIs may opt to populate a centralized repository with their e-theses, incurring a maintenance fee for the service.

Retrospectively acquired theses

Financing the costs of retrospective digitization or retrospective acquisition of digital versions will be

the greatest challenge to the project. Again, there will be two broad options:

a) HEIs who have their own mechanisms for acquiring e-versions or for digitizing in-house may choose to make local arrangements for acquiring older theses.

b) HEIs may choose to outsource the digitization of theses to the centralized service, paying an appropriate fee for this work. There are likely to be economies of scale achievable via a large centralized digitization service operating on a non-profit, cost-recovery basis.

Applying the 80/20 rule to thesis supply suggests that obtaining e-versions of, or digitizing, theses up to approximately 13 years old would add the most value to a national thesis supply service. However, for some subjects, particularly in the humanities, there is also steady demand for older material. Therefore, it is likely that EThOS will propose a combined approach with an overall aim of mass-digitization of theses to generate a critical mass of recent retrospective work, combined with on-demand supply in response to individual requests.

University regulations and procedures

Having established a sustainable service, it needs to be sufficiently flexible to allow individual universities to take advantage of the collaborative aspects of ETDs, yet still work with the grain of local policies and practices. Different universities are at different stages of ETD development: a few have local procedures in place for voluntary or mandatory e-submission of theses (in addition to paper copies), a few have not started thinking about ETDs at all, and most are somewhere in between. Adopting ETDs will require discussion between libraries, academic departments, registries or university administrators, and computer service staff. It may in some cases require a lengthier process of changing regulations and obtaining high-level agreement for this move. There will be a requirement for minimum technical metadata standards to ensure interoperability of thesis discovery and supply mechanisms, and for future-proofing the service against obsolescence of digital formats.

To these ends, the project will be developing a toolkit of guidelines and standards that will provide a start-up kit and reference manual for universities providing e-theses. The toolkit will cover the key IPR issues, which institutions can use to inform local decisions about how best to handle rights in their ETDs, and will provide model documentation for changes that may be needed to procedures and regulations, model workflows, and required technical and metadata standards.

Input from UK HEIs

If you are reading this as a member of a PhD-awarding university in the UK, EThOS would like to encourage you to reflect on your local situation and whether your university would benefit from participating in the EThOS service. If so, you are welcome to make use of the toolkit and other documentation that will start to appear on the project website (http://www.ethos.ac.uk). You will also find there news of EThOS events in Wales, England and Scotland in the spring and early summer of 2006, where the EThOS team intends to present its draft recommendations, guidelines and standards, and receive comments and feedback from members of the academic research community together with library, administrative and computer service staff. This feedback will inform the final outcomes of the project and the prototype service, which is intended to be complete by the end of May 2006. EThOS aims to launch the 'live' service by the end of September 2006.

References

(Note: All URLs referenced here were checked on 20 December 2005)

1. Research Councils UK – position statement on access to research (draft as of 16 December 2005, final statement expected to be available from this web site "early in 2006"): http://www.rcuk.ac.uk/access/index.asp

2. Wellcome Trust – position on open and unrestricted access to the outputs of published research: http://www.wellcome.ac.uk/node3302.html

3. *Index to Theses* published by Expert Information: http://www.theses.com/

4. Links to NDLTD events and papers associated with them are at: http://www.ndltd.org/community.en.html

5. Papers from the 8th Symposium, EDT 2005:
    http://adt.caul.edu.au/etd2005/etd2005.html

6. Copeland, S., Penman, A. and Milne, R., Electronic
    theses: the turning point, *Program*, 2005, 39(3)
    185–197.
    URL for subscribers to *Program*:
    http://www.emeraldinsight.com/10.1108/
    00330330510610546

7. Andrew, T., briefing paper on Intellectual Property
    and Electronic Theses:
    http://www.jisclegal.ac.uk/publications/
    ethesesandrew.htm

■ **Jill Russell**
(Member of the EThOS project team)
Academic Support Team Manager
Information Services
Main Library
University of Birmingham
Edgbaston
Birmingham B15 2TT, UK
E-mail: j.c.russell@bham.ac.uk

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=19&issue=1&spage=32

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=19&issue=1