

ELECTRONIC THESES ONLINE SERVICE (EThOS) – A NATIONAL UK PROJECT

Martin Wolf, University of Warwick, M.Wolf@warwick.ac.uk

1. INTRODUCTION

This paper outlines the background and work-to-date of the Electronic Theses Online Service (EThOS) project. The project's aim is to create **“a working, scalable prototype national service”** for the UK.

The EThOS service, which will be implemented from the prototype service, will enhance access to the UK's body of doctoral theses by providing a single Web-based service through which users can access theses that are both born digital and digitised from hardcopy format.

Information about the project will be presented in the following sections:

- Project background and approach – this outlines the current situation in the UK with regards to accessing doctoral theses, and how the EThOS project was established;
- Proposals and achievements – this outlines the proposals for a national e-Theses service and the project's achievements so far;
- Next steps – this outlines what further steps are necessary to reach the goal of a prototype service.

2. PROJECT BACKGROUND AND APPROACH

2.1 Access To Doctoral Theses In The UK At Present

Current access to doctoral theses in the UK can be a lengthy and cumbersome process. This process is summarised below:

1. an end-user first identifies the thesis they want (from personal recommendations, references in other papers, catalogues and indexing databases, etc.);
2. the end-user makes an Inter-Library Loan request for the item at their home institution (assuming they know this is possible);
3. the requesting institution's library must search for the thesis;
4. the requesting institution's library must then request the thesis either from the British Library or directly from the thesis' originating institution;
5. the British Library must locate the thesis then prepare a copy to be sold at cost to the requesting institution (cost dependant on whether the thesis is to be supplied as a bound photocopy or on microfilm) – alternatively, the originating institution must locate the item and make a decision on whether or not to loan the item;
6. the item is then mailed to the requesting institution, who either absorb the cost of making the request, or pass that cost on to the researcher.

This whole procedure can take up to 6 weeks, and involves a great deal of expensive administrative work on the part of requesting institutions, supplying institutions and the British Library, which administers the current service. For this reason, web-based alternatives to the delivery of UK doctoral theses are being pursued.

2.2 EThOS And Previous E-Theses Projects In The UK

The Joint Information Systems Committee (JISC) is a UK-wide body that supports the use of information and communications technology by further and higher education institutions, funded by all the UK's post-16 education funding councils. It funds research projects that are

grouped together by common themes into research programmes. One such programme, the FAIR programme (Focus on Access to Institutional Resources), funded three separate projects looking at different aspects of e-Theses. These were:

- Electronic Theses – based at The Robert Gordon University, this project explored methods of e-Theses production, management and use, in order to produce recommended models to encourage the take up of e-Theses by UK education;
- DAEDALUS – based at University of Glasgow, this project examined using different software solutions (for example, GNU eprints and DSpace) for storing and delivering different information types (e-Theses, e-prints, grey literature, etc.);
- Theses Alive! – based at University of Edinburgh, this project examined the software requirements for creating online theses submissions systems, and for creating additional functionality, specific to e-Theses, for repository software.

The EThOS project is bringing these different strands together into one project. The three institutions mentioned above are all involved in EThOS (with University of Glasgow acting as lead institution). The aim of the project is to create a **working, scaleable prototype national service** for UK e-Theses.

The main distinguishing elements of the EThOS project are:

- investigation of the digitisation of current print theses in addition to the storage and delivery of new theses created in a digital format (the prototype service envisaged will therefore provide access on-demand to digitised versions of print theses, as well as born-digital theses);
- creation of a “toolkit” that will provide detailed practical advice to aid institutions in establishing Institutions Repositories and contributing to the EThOS service (for further details, see Susan Copeland’s conference paper *The “EThOS” Toolkit*).

The project began in January 2005, though only began in earnest upon the appointment of a full time Project Manager in May 2005. The project is set to be completed in September 2006.

2.3 Who Is Involved In The EThOS Project?

The sponsors and funders of the project include the JISC (50% of funds) and CURL (Consortium of Research Libraries in the British Isles). The rest of the project’s funds come from the partner institutions. The partner institutions have a broad range of experience. In addition to the universities mentioned above, there are other universities with experience of e-Theses (e.g. Cranfield University) and those with no prior experience with e-Theses (e.g. University of Warwick). The British Library and the National Library of Wales are also heavily involved, having made significant contributions to the management and technical development of several aspects of the project.

2.4 How Is The EThOS Project Being Managed?

The project is being guided by a Project Board. On the board are leaders of some of the EThOS work packages (see below), the JISC project officer, and representatives of the stakeholder groups such as senior academics, university administrators and postgraduate students.

A full time Project Manager is in charge of managing the day-to-day running of the project. This is a complicated task given the range of separate elements the project is investigating (technical, legal, procedural, etc.) and the geographically dispersed nature of the institutions involved in the project.

The work has been split into 8 separate work packages, each led by one or two institutions, but often involving collaboration with other project partners. These work packages are summarised in the table below.

Table 1 Summary Of EThOS Work Packages

Work package	Lead Institution
1. Project Management	University of Glasgow
2. Building of British Library-based central host with single search interface	British Library
3. Building of interfaces to harvest data from institutional hosts	Cranfield University
4. Implementation of a pilot programme of retrospective and current digitisation	British Library
5. Development and integration of rights and permission procedures	University of Edinburgh
6. Production of a start-up e-Theses toolkit for institutions	The Robert Gordon University
7. Development of business models (service and digitisation) for long term sustainability	University of Glasgow and the British Library
8. Dissemination and advocacy programme	University of Birmingham and University of Warwick

3. PROPOSALS AND ACHIEVEMENTS

3.1 Service Delivery

After investigating the issues surrounding service models and service sustainability, intellectual property rights, technology considerations, and institutional considerations, the project has devised a series of proposals for a prototype national e-Theses service.

In summary, it is proposed that:

1. the service will provide a server (known as the Central Hub) to be located at the British Library – this server will host e-Theses deposited by institutions who have no Institutional Repository of their own, or who would welcome the opportunity to store copies of their e-Theses in more than one location;
2. there will be an EThOS harvester which will harvest metadata and e-Theses from Institutional Repositories, Consortia Repositories, and other sources of theses bibliographic data (it is hoped the service will eventually hold information on all UK doctoral theses, whether in print or electronic format);
3. end-users will search the EThOS web-based service, locate details of theses they want and then order them;
4. if the thesis the end-user requires is already available in an electronic format (held either on the Central Hub or the relevant Institutional/Consortia Repository) it will be available for immediate download – if not, the thesis will be requested from the

originating institution, digitised by the EThOS service, and uploaded to the Central Hub for end-user download;

5. additionally, end-users can ask for theses to be supplied in a range of physical formats (loose-leaf, soft bound, hardbound, CD/DVD) – they would be charged for supply in these alternate formats on a cost-recovery basis.

Regardless of whether a thesis is held on the Central Hub or on an Institutional/Consortia Repository, the EThOS web service will be able to deliver the required thesis to the end-user, who has to search only one resource.

The proposed service envisages the theses available as being either born-digital or scanned and digitised from hardcopies. Digitisation of hardcopy theses could be undertaken either by the theses' originating institution, or by the EThOS service.

The degree of Open Access (here defined as the free supply of intellectual content to the end-user) to the electronic version of a thesis will largely be determined by the thesis' originating institution, as explained in the next section.

3.2 Service sustainability

A cost-recovery model is necessary in order to sustain the service administration and infrastructure, and fund the digitisation of theses currently available only in hardcopy at the library of the originating institution. In order to maintain a viable service, the following model is proposed:

1. The service will be run on a cost-recovery only basis, i.e. the service will not be for-profit.
2. Participating universities will be offered a choice of two participation options: **Sponsor** and **Associate**.
 - a. **Sponsor institutions** would pay EThOS an annual subscription. Funds from this payment will be used to digitise theses from that institution in response to requests from end-users. If the cost of digitising requested theses during a year exceeds the subscription payment, Sponsors can choose either to fund the additional digitisation, or charge the costs of digitisation to the end-user. If the cost of on-demand digitisation is lower than the yearly subscription, the institution can then choose to have digitised further theses from their collections, up to the value of the subscription payment. Once a thesis has been digitised it would be freely available to any other user of the service.
 - b. **Associate institutions** would pay for the digitisation of theses in response to requests from end-users on a piecemeal basis (for example monthly or quarterly), with no upfront subscription fee. Again, once a thesis has been digitised it would be freely available to any other user of the service.
3. Additionally, the core services of metadata and/or content upload to the Central Hub, and metadata harvesting, will be open to all UK universities for no charge.

This service sustainability model provides a flexible set of options both to end-users (who can select the most appropriate supply format for their purposes) and to participating universities, who have the option to determine if their theses will be available on an Open Access basis.

The proposed model represents a fundamental shift from the present situation, where it is the end-user/institution *requesting* a thesis that pays for its supply, to one where the costs can (if they choose) be borne by the institution *supplying* a thesis. It is expected that, given the reduction in staff time and administration costs to which the proposed service should lead, overall costs of access to thesis content should remain static, or be reduced.

3.3 Achievements

The EThOS project has made a number of achievements on its way to creating a prototype national e-Theses service for the UK of the kind outlined above. These are detailed below.

1. Creation of a standard for UK e-Theses metadata
A qualified version of Dublin Core, these standards should ensure maximum interoperability and consistency between repositories holding data on UK e-Theses.
2. An Intellectual Property Rights report and resultant risk management strategy
The project investigated Intellectual Property Rights (IPR) in two major areas – the IPR in third party material (data, images, text, etc.) used within theses, and the IPR in the theses themselves (does IPR lie with the author or the author's institution?). It concluded that while to digitise a hardcopy theses without first obtaining the author's consent is a breach of copyright, to search for and obtain such permission for every digitised thesis would place an insurmountable strain on the proposed service. As such, it is proposed to digitise theses without first obtaining permission, with procedures in place for fast take-down should authors object to their work being made available in this way.
3. Model end-user and deposit licences
The IPR work package has also produced model licences both for the end-users of the proposed EThOS service (setting out what can and cannot be done with the content) and those institutions contributing content and/or metadata to the service. These licences will also be available for institutions to adapt to meet the needs of their own Institutional Repositories.
4. Standards and procedures for digitisation
Standards for digitisation have been developed. It is most likely that an in-house digitisation studio at the British Library will provide the most cost- and time-effective digitisation of theses. This will be investigated further as part of the evaluation of the project (see Next Steps, below).
5. Harvesting interface and mechanisms
The technical infrastructure for the harvesting of UK e-Theses metadata has been developed and tested. Standard procedures for participating institutions (outlining requirements for exposure of metadata, folder and file conventions, etc.) have been produced.
6. Core Central Hub system built and ingest mechanisms working
The hardware infrastructure for hosting e-Theses metadata and content has been fully tested and configured. The full application is currently being developed. Mechanisms for ingesting e-Theses into the Central Hub have also been fully tested and are working.
7. Advocacy activities
The project has been promoted through various methods, including press releases to the national education press, features in the professional literature, announcements on the mailing lists of stakeholder groups (such as the National Postgraduate Committee and the Association of University Administrators), and a successful promotional workshop.

4. NEXT STEPS

EThOS is nearing the end of its project lifecycle, and should soon deliver all the requirements for the establishment of a prototype national service for access to UK doctoral theses. There is still a small number of steps that need to be taken before the project's end. These include:

1. An evaluation of the project
The initial EThOS plan submitted to the JISC included provision for an independent evaluation of the project, to ensure the proposed service meets the needs of the UK higher education community. This evaluation will take place, and is due to report, in summer 2006.
2. Finalisation of end-user interface
Design of the look-and-feel of the eventual service to be rolled out to end-users is ongoing.
3. Possible involvement in future projects
The JISC has issued a call for projects investigating large-scale digitisation material currently only available in hardcopy format. It is possible the British Library may bid for funds to digitise a large number of print theses. If successful, such a mass digitisation could provide a useful critical mass of content to kick-start the prototype service.
4. Further advocacy events
Two future awareness-raising advocacy events before the end of the project's lifecycle are planned. Responsibility for promotion of the eventual prototype service, and the handover of advocacy from the project to the service, needs to be finalised.
5. Finalisation of the EThOS toolkit
The EThOS toolkit will be an interactive set of tools, accessed through a Web browser, that will provide institutions with all the information they would need in order to participate in the service. This will include elements such as the metadata standards required and the model end-user and deposit licences. Additionally it will contain advice on elements such as how best to ensure engagement with e-Theses at a senior level within institutions, and model procedures for moving towards mandatory submission of theses in electronic format. Further information on the toolkit is given in the paper *The EThOS toolkit* by Susan Copeland.

6. CONCLUSION

This paper has outlined the main features of the Electronic Theses Online Service project. It has covered:

- the background and structure of the project;
- the project's proposals for a future UK national e-Theses service and its achievements to date;
- the steps necessary to move from a project to a live prototype service.

The project is well on track to deliver a workable, scaleable prototype national e-Theses service for the UK by September 2006. It is expected that a live service will be launched, subject to infrastructure developments at the British Library, in 2007.

9th International Symposium on Electronic Theses and Dissertations
IX^e Symposium international sur les thèses et mémoires électroniques
June 7 – 10 Juin, 2006, Quebec City / Québec, Canada