EThOS: An Electronic Theses Online Service for the UK

Susan Copeland
Senior Librarian, The Robert Gordon University, Aberdeen, Scotland

Simon Bevan
Information Systems Manager, Cranfield University, Cranfield, England

ABSTRACT

In 2002 the UK Joint Information Systems Committee (JISC) funded fourteen projects as part of its 'Focus on Access to Institutional Resources' Programme. Three of these projects related directly to ETDs: two reached completion in October 2004 and the third is scheduled to finish in July 2005. The key results and recommendations from these projects will be highlighted in this presentation, which covers topics ranging from advocacy and Intellectual Property Rights to software for institutional e-theses repositories and a proposal for a national model.

Many of the proposals arising from the above (the 'Electronic Theses', 'Theses Alive!' and 'Daedalus' projects) are now being taken forward as part of another JISC funded project: the 'Electronic Theses Online Service' ('EThOS'). The 'EThOS' project began in January 2005 and it will last for a period of 18 months. It aims to deliver a fully operational prototype UK e-theses online service based on a hybrid I.T. infrastructure which will include a central host, based at the British Library (for Higher Education Institutions that wish to have their theses stored centrally for dissemination or preservation purposes) and a mechanism to harvest data and metadata for the supply of theses that are stored in institutional repositories. The project will develop procedures to deal with digital rights management, and it will develop a viable and sustainable business model. It will include a programme of retrospective digitisation and it will develop a dissemination and advocacy programme. The conference presentation will describe the 'EThOS' project in some detail and provide information about the findings to date.

KEYWORDS

EThOS, JISC, ETDs, e-theses, UK

1. INTRODUCTION

Interest in the development of electronic theses and dissertations (ETDs) in the UK dates back to the mid 1990s. However, detailed research, and significant practical work to test ideas, commenced in 2002 when the Joint Information Systems Committee (JISC) provided funding for three related projects as part of its 'Focus on Access to Institutional Resources' (FAIR) programme (http://www.jisc.ac.uk/index.cfm?name=programme_fair). The three projects examined a wide range of issues associated with ETDs, including cultural, legal, administrative and technical matters. As the research and development work reached a conclusion in 2004 / 2005, guidelines were produced and advice was disseminated to the UK Higher Education community to encourage as many institutions as possible to start advocacy work. The 'Electronic Theses' project consortium, led by The Robert Gordon University

(RGU), also suggested models for the storage and provision of ETDs at national level (http://www2.rgu.ac.uk/library/e-theses.htm).

In 2004 JISC issued an 'Invitation to Tender' to 'undertake a test-bed implementation project to deliver an infrastructure to support the deposit, access and use of research theses for the UK Higher Education (HE) sector' (http://www.jisc.ac.uk/index.cfm?name=funding_etheses). The successful proposal, for an 'Electronic Theses Online Service' (EThOS), was submitted jointly by the Consortium of Research Libraries in the British Isles and the British Library in partnership with the University of Glasgow, the University of Birmingham, Cranfield University, the University of Edinburgh, The Robert Gordon University, the University of Southampton, the University of Warwick, the National Library of Wales, and SHERPA (a consortium partner led by the University of Nottingham). Many of the recommendations from the three earlier projects are being incorporated into the current work, which commenced in January 2005.

EThOS aims to deliver a fully operational prototype UK e-theses online service that will enable end-users, via a single Web interface, to access the full text of electronically stored theses after selection from a UK Database of Theses. In line with the findings of the 'Electronic Theses' project, that some institutions are intending to store their ETDs in their own institutional repositories and some would value the opportunity to deposit their material in a national collection, the EThOS project aims to develop and implement a hybrid IT infrastructure. The latter will combine a central host, based at the British Library, for those wishing to have their theses stored centrally, and interfaces for the British Library to harvest the necessary data and metadata for the online supply of theses that are held only in institutional or consortial repositories.

2. RECOMMENDATIONS FROM THE 'FAIR' PROJECTS (2002-2005)

2.1 Cultural issues

All three of the JISC funded projects relating to ETDs that were undertaken between 2002 and 2005 concluded that a significant amount of advocacy work is required to create the culture change that is required if supervisors and senior managers in UK Higher Education institutions are to fully support the concept of electronic theses submission. The 'DAEDALUS' project, based at the University of Glasgow, involved both advocacy and service development work and recognised that the major challenges involved in creating an institutional repository containing various types of research output, including e-theses, are cultural and organisational rather than technical (http://www.lib.gla.ac.uk/daedalus/). Similarly, the 'Theses Alive' project team, based at the University of Edinburgh, appreciated the need to persuade key stakeholders to understand the benefits of ETDs (http://www.thesesalive.ac.uk/). This project therefore also considered relevant cultural issues alongside the technological aspects that were the main focus of its investigations.

The 'Electronic Theses' project led by RGU involved a consortium comprising the British Library, the University of Aberdeen, Cranfield University and the University of London Library. The focus of this project was broad: to evaluate a wide range of existing practices of e-theses production, management and use in order to produce models for use within the UK information environment. Cultural issues, necessarily, formed a major part of the project's investigations. The key finding of the project in this area was the need to differ the emphasis when undertaking advocacy work with different 'market segments': students, supervisors, university senior management, and librarians. For example, students often see the main benefit of electronic submission as the opportunity to express their research results in more creative ways, supervisors value the improved access to theses when they are available on the Web, senior management are influenced by statistics showing that the research output of the university is being more widely consulted, and librarians appreciate the chance to provide

an improved service while also saving shelf space. Advocacy work needs to be planned to be most effective and the EThOS project is placing emphasis upon this topic, alongside the technical and organisational issues, in order to ensure that emerging examples of best practice continue to be widely disseminated.

2.2 Administrative and legal issues

The 'FAIR' projects considered many of the administrative issues associated with ETDs, concluding that institutions embarking upon a change in their regulations need to be aware of the potentially lengthy time scale involved. The committee structure in universities often results in delays if questions are asked that cannot be answered immediately, paperwork may need to be checked by legal officers, students may need to be consulted, etc. The 'Electronic Theses' project has made example regulations relating to ETDs available on its Web site, but each institution will have its own requirements and individuals are encouraged to make use of examples of good practice from a wide range of universities that have now endorsed the practice of electronic theses submission.

In the UK, the Library community has been assisted in its attempt to encourage the adoption of e-theses by the support of the UK Council for Graduate Education (UKCGE). Speakers representing the 'Electronic Theses' and 'Theses Alive!' projects were invited to address an audience of postgraduate administrators etc. at a UKCGE event in London in March 2005 and this provided an opportunity to demonstrate the benefits of ETDs to key personnel from university registries and graduate support offices. The area of greatest concern amongst this audience related to legal issues: copyright and Intellectual Property Rights (IPR). A talk by a representative from the 'Theses Alive!' project, who is the author of a JISC Legal Information Service briefing paper on IPR and e-theses provided much valuable information on this subject, and the argument that students benefit from gaining a better understanding of this complicated issue could not be denied (http://www.jisclegal.ac.uk/publications/ethesesandrew.htm). A body of knowledge is gradually being built up in this area, and the availability of plagiarism detection software etc. is of assistance (http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/Jiscpas/site/detect.asp). However, there are many legal issues relating to ETDs and, again, the EThOS project aims to further development work in this area through a work package dealing with IPR and permission procedures.

2.3 Technical issues

2.3.1 Software

A number of technical issues were addressed by the 'FAIR' projects and details are available on the project Web sites. One of the key decisions related to the evaluation of a variety of software packages which could be used to create an institutional repository containing ETDs. All three of the JISC funded projects considered a range of products and independently reached broadly similar conclusions. The 'Electronic Theses' project assessed selected software packages against a set of criteria, such as availability, likely sustainability, and interoperability, and recommended 'DSpace' (http://www.dspace.org) and 'EPrints' (http://www.eprints.org) as highly suitable. The 'Theses Alive!' project team concentrated particularly on 'DSpace' and have produced an add on submission system relevant to e-theses: the 'Theses Alive Plug-in for Institutional Repositories' (TAPIR) (http://www.thesesalive.ac.uk/dsp_home.shtml).

2.3.2 Metadata

Representatives from the 'FAIR' projects were also involved in detailed discussions about metadata and the agreement of a recommended core set

(http://www.rgu.ac.uk/library/guidelines/metadata.html). The latter is available on the 'Electronic Theses' project Web pages and it is currently serving as the basis for further work in this area by the EThOS project team. Members of the EThOS project team are aware of international developments relating to recommended metadata sets and consideration will be given to further developments in this area over the coming months.

2.3.3 Infrastructure

The 'Electronic Theses' project investigated possible models for the provision of theses in electronic format at national level. Feedback at dissemination events etc. indicated that some institutions are keen to establish their own institutional repositories, storing ETDs alongside other research output, such as e-prints. However, other institutions indicated that the cost of maintaining such a repository might be prohibitive, and that they would welcome the chance to deposit their e-theses in a central, national collection so that the work of their research students was equally easily accessible to potential readers. Some librarians are in favour of the 'LOCKSS' principle and favour storing a copy of their students theses in both locations. The FAIR projects have already provided a considerable amount of guidance for those who wish to establish an institutional repository containing ETDs. The aim now is for the EThOS project to progress work on the development of a national infrastructure.

3. THE EThOS PROJECT (2005-2006)

3.1 Project Outline

The EThOS project has funding from JISC for a period of eighteen months and it will provide a working prototype for a national scheme by summer 2006. As indicated above, the EThOS project will develop and implement a hybrid IT infrastructure, combining both a central collection and an interface for harvesting data and metadata for the online supply of theses that are stored across the UK in institutional or consortial repositories. A major element of the service will be a database of UK theses which users will be able to search in order to select and locate all indexed theses as well as view those which already exist in (born-digital and digitised) electronic format.

The project will develop a viable and sustainable business model which takes account of the need to run on a cost-recovery basis while favouring a free-at- point- of-use policy. The project will also address a range of important associated issues including digital rights management, IPR, royalties and permissions, and advocacy work. There will be a programme of retrospective digitisation work to ensure critical mass and there will be continued emphasis upon advocacy and dissemination work in order to encourage institutions to make theses available in electronic format in future.

The EThOS project comprises eight separate, but closely linked, work packages. The work package leaders come from different institutions within the consortium, and a full time Project Manager ensures that there is good communication between the related teams. An enthusiastic Project Board, chaired by the Project Director, includes a university Registrar and a Faculty representative, amongst others, so the project is benefiting from a diverse range of views and opinions.

3.2 Project Work Packages

Work package 1: Project Management. For such a large project, with many inter-dependent work packages, the project management strand is of central importance. It is being undertaken using PRINCE2 methodology which does have an overhead in terms of time and resource but which was deemed appropriate for the size, budget and importance of the project.

Work package 2 is the Central Hub development – building a British Library based central host with a single search interface using the 'e-prints' software. This work package will set up a fully functional host service using standard BL technical infrastructure components, metadata developed by the EThOS partnership and content from partner institutions. The central host will accept, load and store standard format e-thesis metadata from institutions which wish to supply their own content, as well as providing a one-stop shop 'identify and locate' function in a single search interface. The system will also administer rights, royalty and permission requirements and payments if and where appropriate.

Work package 3 will build interfaces to harvest data from institutional hosts. It will set up fully tested and functional interfaces to harvest and gather metadata and content from institutional hosts using a number of scenarios: direct submission of metadata and content, OAI harvesting, dynamic cross-searching of distributed repositories and the physical loading of CDs/DVDs. These interfaces will then be integrated into the BL based host system.

Work package 4 - Digitisation – is the implementation of a pilot programme of retrospective and current digitisation. This package will manage the digitisation of around 400 volumes of paper-based theses (bookscanning) and around 1500 volumes of microformed theses (microfilm scanning) in order to develop processes and inform the business model for a sustained programme of digitisation.

Work package 5 runs across many other work packages and covers the IPR and permission procedures: the development and integration of rights and permission procedures. It will produce best practice procedures related to rights, royalties and permissions prior to the development of the national host service and the institutional toolkit. The three strands will cover IPR and digitisation, author IPR and end user IPR.

The development of an institutional toolkit is the focus of work package 6. It will develop a start-up e-theses toolkit for institutions to support the setting up of an e-theses system within an individual institution and its integration within the UK-wide online service. It will cover hardware/software specifications, metadata, workflows for processes within institutions, procedures for IPR, royalties and permissions, preservation, advice on training authors, and a support network.

Work package 7 will recommend and develop a viable business model (service and digitisation) which takes account of the need for long-term sustainability. The service will be designed to run on a cost-recovery basis but it aims to adopt a free-at-the-point-of-use policy. It will be underpinned by a programme of retrospective digitisation to ensure critical mass.

The advocacy and dissemination programme which is the subject of work package 8 aims to prepare and pilot a template of advocacy measures that will support users in persuading their institutions to accept digital submission of theses and to participate in the EThOS service. It will disseminate information about the proposed EThOS service to appropriate target groups within and outwith Higher Education institutions in order to achieve an awareness of the service. A number of specific activities are planned, such as the identification and interviewing of stakeholder groups, as well as the design and production of publicity material which will include scenarios demonstrating the benefits of a successfully implemented service.