

## The "Electronic Theses" Project: A view from the consortium led by The Robert Gordon University

Andrew Penman

The Robert Gordon University

[a.penman@rgu.ac.uk](mailto:a.penman@rgu.ac.uk)

<http://www.rgu.ac.uk/library/e-theses.htm>

Keywords: Electronic Theses

### Abstract

"Electronic Theses" is a JISC funded FAIR (Focus on Access to Institutional Resources) project spanning 2 years from July 2002 until July 2004. The project, led by RGU, consists of members from the following institutions:

- The Robert Gordon University
- University of Aberdeen
- Cranfield University
- The University of London Library
- The British Library

The main aim of the project is to evaluate a wide range of existing practice and methods of e-theses production, management and use against a set of criteria in order to produce recommended models for use within the UK information environment.

The project builds particularly on recent work undertaken by the Networked Digital Library of Theses and Dissertations (NDLTD). It is also considering the potential developments and opportunities offered by the production of 'born digital' theses, and ways of creating an electronic theses resource base at national level.

This paper will highlight the work undertaken by the project and outline many of the findings and key issues that have surfaced along the way.

- Levels of interest within the UK
- Learning from others; success and established SYSTEMs transferable to the UK
- Which SYSTEMs are best suited to specific demands
- Potential problems and obstacles to progression
- Opportunities
- Unified progression? Distributed, institutional and centralised SYSTEMs

### Who is involved?

"Electronic Theses" is a JISC funded FAIR (Focus on Access to Institutional Resources) project spanning 2 years from July 2002 until July 2004. The project, led by RGU, consists of members from the following institutions:

- The Robert Gordon University
- The University of Aberdeen
- Cranfield University
- The University of London Library
- The British Library

The project is also closely linked with the E-theses cluster group consisting of the following members:

- Electronic Theses: The Robert Gordon University led consortium
- Theses Alive!: The University of Edinburgh
- DAEDALUS: The University of Glasgow

Liaison has also taken place on a wider scale with the E-Prints Cluster, consisting of: the following, in addition to the 3 members mentioned above:

- e-prints UK The University of Bath
- SHERPA The University of Nottingham
- TARDIS The University of Southampton
- HalRST The University of Strathclyde

### Project Aims

The main aim of the project is to evaluate a wide range of existing practice and methods of e-theses production, management and use against a set of criteria in order to produce recommended models for use within the UK information environment.

The project builds particularly on recent work undertaken by the Networked Digital Library of Theses and Dissertations (NDLTD). It is also considering the potential developments and opportunities offered by the production of 'born digital' theses, and ways of creating an electronic theses resource base at national level.

The research can be broken down in three main stages:

1. Information Gathering
  - NDLTD Members Website
  - General WWW Search
  - Questionnaire / UK Library survey
  - Conference attendance and liaison
2. Evaluation
  - Breakdown of sites of interest via criteria
  - Software Options available - Install
3. Recommendations and Conclusions
  - Proposed model/s
  - Creation of suggested guidelines

## Information Gathering

### NDLTD Members Website

#### Learning from others: success and established SYSTEMs transferable to the UK

The initial stage of the project was focused around the 152 members of the NDLTD. This provided a concise list of each university and institution which has an archive, pilot or interest in ETDs. A SYSTEMatic search through each member's listed URL enabled a database to be created containing information corresponding to the following set of criteria:

- **Subject and Numbers Stored**  
This was chosen to provide an insight into the most established archives and identify the number of ETDs that could realistically be expected in an archive. This helps us to better understand hardware requirements and how to effectively manage a collection containing a variety of subject areas and formats.
- **Sites with Usage Statistics**  
The majority of sites did not contain specific information (such as that from web logs) but from the ones that did a very useful impression was offered as to potential traffic volume and interest in ETDs. Again, this helped us to understand server requirements and the sort of bandwidth that could be required by differing sizes of institutions.
- **Production and Requirements**  
As ETDs are still very much in their infancy in the UK, this was an area that was viewed with high interest. It is important to offer students advice and support for creating 'born digital' documents and also for converting conventional word processed documents into suitable formats for electronic archival. This guidance needs to cover a surprisingly large amount of information as it is important to allow those students who may not be very IT literate, to create and express themselves with a variety of media, thus the complexity and need for explanation can greatly increase. Interestingly, the majority of sites provided this information and offered useful institutional guides, if not links to external pages of interest.
- **Usage of XML / SGML**  
Again, this area was somewhat limited in that very few sites actually mentioned or explained the workings of the backend to their archive, but it is most likely that XML will be used for metadata purposes. However by following various links branching from the NDLTD site it was possible to gain a better idea of where and how XML and SGML were being used. On the document side of things, very few archives had the actual documents in XML format, merely the abstract and metadata section. The overwhelming majority still adhere to the pdf or basic html formatting.

- **Access and Sustainability**  
Are archives only available to students on the institutions campus? (via intranet or password protected) Or is access available to anyone as a free service? Free open access is an important point to consider when formulating a national model. There will inevitably be debate as to whether or not theses should be a provided free or at cost with perhaps the ability to view abstracts or even chapters before purchase. The majority of archives viewed did in fact offer theses free of charge, with a small number providing free access to those only on the institutions campus network. With regard to sustainability, we looked at how existing archives or projects achieved funding and then, once established, had managed to maintain an archive and the associated costs.
- **Security, Copyright and Administration**  
Copyright still remains very much an area of concern and debate, so it was interesting to see how various institutions handled the matter with regard to content and also ownership of the theses.
- **Institutional Summary - noteworthy points**  
This field held a brief summary of any site that had information added. Anything noteworthy or of interest outside the aforementioned criteria was entered here. This proved useful to highlight examples of best practice regarding many aspects of ETD production and archival, no matter how great or small.

Once each URL had been checked, the resulting database could be queried to compare or list relevant information and statistics. It is worth noting that inevitably, not all 152 members had information entered due to a variety of reasons such as: outdated URLs or the lack of a pilot or archive available or simply because the information provided on archive sites did not cover the chosen criteria.

### General WWW Search

A large scale search covering many aspects of ETDs was also carried out via the use of search engines and the following of links from institutional sites. The extensive 'UNESCO Guide for Electronic Theses and Dissertations' was also consulted as it offered presentations and papers dating back several years covering almost every conceivable topic concerning ETDs.

This in total provided a great deal of information to augment our specific institutional findings and projects.

### Questionnaire - UK Library Survey

A brief questionnaire was distributed to chief librarians throughout the UK via IIS-SCONUL to obtain information about levels of interest and activity in e-theses production.

The questionnaire was kept deliberately brief in an attempt to keep response high, so only two basic questions were asked:

*Is your Library/Institution currently utilising or developing a SYSTEM for the production and management of electronic theses?*

If 'No':

*Would you consider doing so if a suitable national model and guidelines are provided?*

If 'Yes':

*Is this organised centrally (e.g. by the Library) or undertaken selectively by departments?*

*Is access restricted to staff and students of your own institution?*

*Is there a Web page with information?*

To date just fewer than 30 replies have been received. Many very insightful and encouraging comments were also received, indicating that although few electronic theses repositories exist at the moment, there is clearly a high level of enthusiasm in the area and also more notably, a desire to obtain guidance for creating such archives.

### Conference Attendance and Liaison

Since the beginning of the project, the RGU team have attended several conferences and meetings covering a variety of topics concerning ETDs:

30<sup>th</sup> October 2002: Arts and Humanities Digitisation Service (AHDS), Glasgow

31<sup>st</sup> Oct - 1<sup>st</sup> Nov 2002: JISC IE Joint Programme Meeting, University of Nottingham

6<sup>th</sup>-8<sup>th</sup> December 2002: Open Archives Forum 2<sup>nd</sup> Workshop, Lisbon

23<sup>rd</sup> January 2003: British Library Document Supply Centre, Boston Spa

26<sup>th</sup> February 2003: JISC EFX Workshop, Manchester Metropolitan University

In addition to this, frequent meetings have taken place at both Edinburgh and Glasgow to allow for coordination between the Theses Alive! and Daedalus E-theses projects.

## Evaluation

### Breakdown of sites of interest via criteria

Once the initial investigation was completed, a list of some 20 sites, federated and independent, were chosen for further evaluation. This list was an array of international institutions selected for a variety of factors. These included, design based aspects such as user interfaces, graphical design and layout, ease of use and navigation, to more technical issues such as their usage of metadata and search and retrieve capabilities.

It is hoped that we can learn from these established projects and create a model that addresses these instances of best practice and innovation.

Similarly, since the beginning of the investigation, sites that are were not user friendly or appealing have been witnessed and evaluated. This is of equal importance as it allows us to understand why these sites do not appeal as much as others, and thus hopefully enable us to propose a more valued and functional UK model.

This evaluation is very much an ongoing process within the project, with new developments and work being carried out until the end of the project plan.

## Software Outline

### Open source vs. commercial

From early 2003, research into available ETD software options has been carried out.

In the initial stages of the project we looked at several commercial packages and document management SYSTEMs. However, at present it seems more appropriate to concentrate on the open source side of things as many of the commercial options are closely linked with existing library SYSTEMs and are purchased as a module to be added on to such packages and usually with a high level of cost. Thus, to recommend one over another is unlikely to be the most viable or welcomed solution as different institutions will be utilising their own preferred and established library SYSTEMs.

Therefore, the RGU project initially concentrated on three main open source software options for the creation and management of an institutional electronic theses, or document, archive. These offer a standardised and interoperable solution for institutions to adopt without having to create entire software suites or web interfaces by themselves.

This paper will not highlight detail the technicalities and features of each option as this has been covered before elsewhere and because of the open source nature, extensive documentation may be found on the respective websites.

- DSpace: [www.dspace.org](http://www.dspace.org)
- Eprints: [www.eprints.org](http://www.eprints.org)
- VT-ETD: <http://etd.vt.edu>

It is worth mentioning that all three have been implemented in the last few years, and thus have all proven to be functional and effective for the creation of an electronic archive. So, with that in mind, the main issue here is surely that of unification and whether it is advisable to decide on a single option for adoption within the UK or if different packages can interoperate with each other via the use of OAI-PMH standards and such like.

It should also be noted that the installation and configuration of these pieces of software is not a simple task

and requires a degree of knowledge with UNIX and either Perl or Java and each institutions install will inevitably create unique problems whether relating to hardware / software or tailored institutional requirements. This is perhaps the first, or at least most apparent, negative aspect of the open source route.

## Conclusions to date

2002 and 2003 have witnessed a big increase in ETD development throughout the UK. This is primarily due to the JISC FAIR projects funding for extensive research and development

As we have experienced during the project, the amount of information and knowledge that can be learnt from other international ETD archive projects is vast and there to learn from..

It is now important that, in the UK, we don't start reinventing the wheel, instead use this information to arrive at an appropriate and valued model that may be adopted.