

**Cover Sheet for Proposals****Focus on Access to Institutional Resources (FAIR)**

Name of Lead Institution/Organisation	Edinburgh University Library	
Name of Proposed Project	Theses Alive!	
Full Contact Details for proposal leader		
Name:	John MacColl	
Address:	Darwin Library The King's Buildings University of Edinburgh Mayfield Road Edinburgh EH9 3JU	
Tel:	0131 650 7275	
Fax:	0131 650 6702	
Email:	john.maccoll@ed.ac.uk	
Programme Area(s) of Proposal (Please indicate which programme area your proposal relates to using the categorisation in paragraph 9 of the Circular) Support for disclosure of institutional assets (digital information and learning resources) through the use of other relevant protocols.		
Length of Project and cost to JISC over its Life	24 months.	£84,277
Cost of Proposal to JISC in each academic year	Yr 1: £49,058	Yr 2: £35,219
Proposed Project Start date	1 July 2002	
Outline Project Description		
<p>This project will examine the use of OAI-compliant software designed to handle electronic thesis metadata. The software is the Virginia Tech-Electronic Theses &amp; Dissertations (VT-ETD) package, which is used in the Networked Digital Library of Theses &amp; Dissertations. This software permits individual institutions to disclose metadata about their institutional theses and dissertations. The Theses Alive! Project will seek to adapt and develop the software for use across the UK. It will develop a pilot national service which will have the aim of supporting electronic theses (e-theses) creation and management for UK universities. We recognise that there are particular characteristics to the UK environment for thesis metadata (largely because of the well-known services of the British Library and of the <i>Index to Theses</i>) which suggest that a national approach may be the most effective way of enabling universities to produce multi-purpose thesis metadata, to accept the need to develop procedures and systems to accommodate e-theses, and to ensure that the advocacy work required to generate e-thesis content is deployed with maximum efficiency. The Project will therefore work both to create the submission software required to make the disclosure of theses metadata as simple as possible, and to secure the take-up of the system and – through it - of electronic theses production and management generally in UK HE. The Project will complement the work of the proposal led by the Robert Gordon University.</p>		
Name and Contact Details of Project Evaluators		
<p>John Akeroyd, Director of Information Services, South Bank University Neil Dumbleton, University Librarian, St Andrews University</p>		

## Edinburgh University Library

Proposal response to JISC Circular 1/02 Focus on Access to Institutional Resources (FAIR) Programme

### Theses Alive!

#### 1. Summary

This project proposes a pilot national service which will have the aim of supporting electronic theses (e-theses) creation and management for UK universities. The pilot service will be based upon the software which underpins the Networked Digital Library of Theses & Dissertations (NDLTD), an international network which promotes e-theses submission, creation and management within a web-based virtual database environment. While any UK university is at present free to join the NDLTD, we recognise that there are particular characteristics to the UK environment for thesis metadata which suggest that a national approach may be the most effective way of both enabling universities to accept e-theses, and to ensure that the advocacy work required to generate e-thesis content is deployed with maximum efficiency.

While all theses now produced are in fact 'born digital', there are several good reasons for moving gradually to an environment in which the full digital texts – and their digital supplementary material – become freely available to the international research community. Security of the intellectual property of authors and institutions must be ensured, and since theses represent the fulfilment of a rigorous research process within the rules of intellectual enquiry for a given discipline, it is essential that only the final, successfully defended thesis is published with a freely-available corpus. The initial attention of *Theses Alive!* will therefore be to the metadata which describes theses. A system will be developed to allow metadata to be originated within universities with multiple routes for its publication. As full e-theses content becomes available in due course, so the metadata in its various locations will increase in value to the scholarly community, being merely a click away from the thesis content, rather than, as has been true of thesis access within the UK for many decades, the means to an order for an interlibrary loan or sale copy of the work to which it refers.

#### 2. Background

In October 2001 the SELLIC<sup>1</sup> team at Edinburgh University Library (EUL) presented a report to the UK Theses Online Group (UTOG) on the results of our *Edinburgh University Library Doctoral Theses Digitisation Project*. This work was generously funded by UTOG. Our report contained certain recommendations for the future of the management of 'born digital' theses at Edinburgh. We concluded that universities were moving into a digitally networked environment which has the potential to transform the current system for providing access to theses, by making them freely available on the web through the Networked Digital Library of Theses & Dissertations (NDLTD).

In the Executive Summary of that report, we state the following:

'The wish of academics for themselves and their students is for rapid, unfettered access to scholarly material. In the course of our work we have discovered that several academic departments are therefore already beginning to create theses repositories on web servers, while continuing to follow the traditional systems the Library has put in place for the management of the archival, print copy of theses. We therefore see a dual system developing, with the useful version being the electronic version, located on a

---

<sup>1</sup> Science & Engineering Library, Learning & Information Centre

departmental server, and the archival version held in print by the Library. Because of the requirements of digital preservation, as well as effective metadata and interoperability with other theses repositories worldwide, it is important that the Library becomes the manager of such a service ...

Our involvement in the UTOG pilot has also made us aware of the potential of a university-federated network of theses repositories to provide the same facility as the British Thesis Service, but on an international scale. We have decided to direct our energies towards assisting that ambition, and would be delighted if other UK universities joined us in this, whether via the British Thesis Service or independently.'

We believe that there is now sufficient interest in the potential of networked digital theses for a UK-wide national service to be established, and therefore would like to propose that JISC fund a project to deliver a pilot UK-wide service, to be called *Theses Alive!* (we are grateful to Fred Friend and the JISC Scholarly Communication Group for suggesting this name). We believe that such a service can be built from a number of existing components, viz:

- The NDLTD
- The *Index to Theses*
- The British Thesis Service at the British Library

### 2.1. The Networked Digital Library of Theses & Dissertations

The NDLTD is an international federation of digital theses archives, developed and coordinated by Virginia Polytechnic Institute (Virginia Tech). The NDLTD provides freely available open source software to allow any university to set up their own theses archive, together with a central web-based service allowing access to all archives in the network. This software is known as Virginia Tech Electronic Theses and Dissertations (VT-ETD). The software has recently become compliant with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) which means that all compliant archives can provide their metadata for harvesting by a search provider, thus permitting the searching of all theses archives in the network as though they were a single database of theses. The aim of the NDLTD is to encourage its member sites to move towards the provision of full-text theses online, though this is not an immediate requirement of membership.

### 2.2 Index to Theses

In 2000, the JISC Scholarly Communication Group awarded a development grant to Expert Information Ltd, publishers of the *Index to Theses*<sup>2</sup> to test the feasibility of automating the current theses submissions process, so that data may be submitted for publication via the Internet. It was envisaged that this would result in a more efficient, timely and richer service by both accelerating the publishing process and increasing the amount of relevant information made available on each thesis. Also, by sharing the information with the British Library's British Thesis Service, it would streamline the submission procedure currently used.

### 2.3 British Thesis Service

The BL Document Supply Centre's British Thesis Service (BTS) provides access to more than 140,000 theses submitted for doctoral degrees from over 100 UK universities from the late 1960s to the present day. The BTS supplies microfilm or bound copies of full-text theses for loan or purchase to Document Supply Centre customers on demand.

The British Library plans to develop a theses service to include electronic access and has

---

<sup>2</sup> [www.theses.com](http://www.theses.com)

recently become a member of the NDLTD on behalf of its BTS members. It intends to create a 'back collection' of its microforms by converting them to digital form, and to move to digitising in place of microfilming, working with 'born digital' theses. It wishes to do this in the most effective way so that universities currently participating in the British Thesis Service continue to see the benefits of the 'one stop shop' approach.

Edinburgh University Library (EUL) withdrew from the BTS in 1992, since it felt that it could provide its own thesis loan and sale service to better effect independently. Our recent involvement in the UTOG pilot has made us aware of the potential of a federated network of theses repositories in universities to provide a service on an international scale, and we wish to join and contribute to this network as our first priority. Nevertheless, we realise that for many university libraries the BTS provides a very useful means for ensuring that their theses are accessible to researchers internationally.

### **3. Partners**

The Project will be based in Edinburgh University Library. However, it is intended that it will operate in close conjunction with a complementary project proposed by a consortium led by the Robert Gordon University, and involving the University of Aberdeen, Cranfield University, the University of London and the British Library. This project will aim to produce a study which examines a number of models for the most effective support of e-theses creation and management by UK universities. The *Theses Alive!* model, delivered by the Project proposed here, will allow for the close examination of one particular model, and will therefore form the centrepiece of the RGU study. Since it will develop a real pilot e-theses creation and management system in the course of two years, it will allow the Project Officer based at RGU to examine in great detail the problems and successes of the particular model which informs *Theses Alive!*. The recommendations of the RGU pilot to JISC at the conclusion of its work will play a significant part in influencing the future direction of e-theses development for the UK. It will consider whether the *Theses Alive!* model is the most effective, and merits further support from JISC or support from other sources, or whether an alternative model is more appropriate.

### **4. Vision**

Through its British Thesis Service, the British Library has for many years offered UK universities a service which was national in scale. However, it has not yet taken advantage of the availability of the networked digital environment, and we consider that JISC, representing UK Higher Education, has an opportunity at the present time to promote the adoption of this environment by universities in order that theses can be provided online for the benefit of international research and scholarship. JISC can coordinate a national system which caters to the needs of researchers and meshes with the hybrid library environments currently being developed by university libraries through the provision of a new service, *Theses Alive!* We recommend that this system be based upon a distributed model, with the VT-ETD software being made available to universities for the provision both of a metadata submission system, and a full-text repository for those universities wishing to use it.

Our vision of *Theses Alive!* is of a service which adopts the following principles.

#### **4.1 A UK Electronic Theses Support Pilot Service is Established**

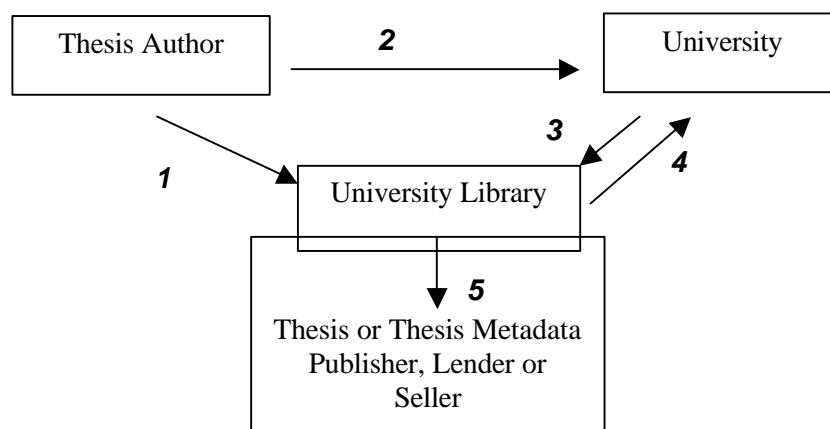
Universities in the UK will require assistance to move from their present arrangements for theses provision, to an e-thesis-based service. A central, JISC-funded support service could provide that assistance over the next few years, until such time as submission systems have coalesced into a standardised form, and alterations to university regulations to permit e-theses provision have become widespread. The service would provide:

- Technical assistance in installing and maintaining the VT-ETD software, acting in a liaison capacity with Virginia Tech.
- Support to the community as it makes the transition (documentation and e-mail support).
- Advice on how to join the NDLTD, and an agency service to support UK university library membership.
- Organisation of a national one-day workshop on e-theses, including a representative from the NDLTD.
- UK-wide advocacy and support for campus advocacy initiatives.
- Promotion and publicity for the move to e-theses, using the professional library, archive and general HE professional press.
- Advice on digital preservation of electronic theses.

There is an associated stage required before such a service can be launched, however. The technical development of an open-source e-theses submission system, based on work already done by the Virginia Tech team and by Expert Information, is also required. Once the system has been developed, it also needs to be maintained, and therefore there are costs associated with development and support which it is recommended should be included within the Project.

#### 4.2 The University Library is a Trusted Intermediary

We envisage the University Library having a role as the key university agent in the thesis publishing process. In this role, it would provide supportive documentation to postgraduate students, via their departments, at the commencement of their dissertations and theses, drawing on the support of the national pilot service. It would ensure that theses authors are given training in the use of thesis submission software several months before they were due to submit. It would receive submissions once they had been ‘signed off’ by the relevant ‘registrar’ – a term which may stand for whichever officer of the university is authorised to release the thesis – whether this is at departmental, faculty or central university level. The signing off is of course the last stage in the academic validation process, and follows on from the successful defence of the thesis by the student and the award of the postgraduate degree. The library therefore takes the role of ‘trusted intermediary’ in what is essentially a triangular relationship, thus:

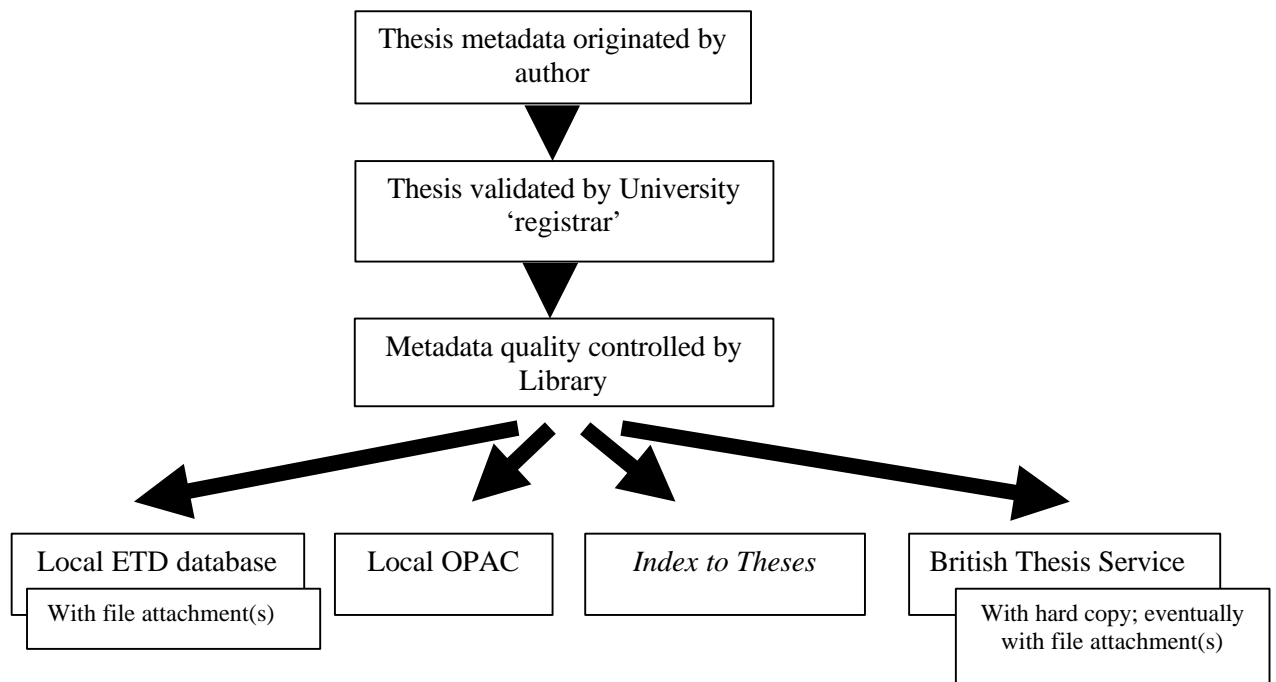


1. During the course of their research, the thesis author sends the Library the basic metadata for their thesis. The Library retains this.

2. The thesis author then submits the full thesis to the University.
3. The University submits the successfully defended thesis to the Library.
4. The Library matches thesis to metadata, and ensures that metadata is complete and that the University validation has taken place.
5. The Library finally releases metadata and, if appropriate, the full text of the thesis – including supplementary digital material – to publishers, lenders or sellers. These would include the Library itself, acting as a publisher (via the Virginia Tech EDT database software), lender or seller on behalf of the University; the British Thesis Service; the *Index to Theses*; and, potentially, other services.

#### 4.3 Metadata is Generated Once but is Multipurpose

Once the Library is happy with the metadata, it is released by the system, and the *same set* of metadata is used multiply by the various agencies providing publication, loan or sale services. It is not likely that the identical set of metadata will be required by each of these agencies, so the system operated by the Library should accommodate a superset, from which appropriate subsets can be generated for the requirements of agencies. It is recommended that the metadata employ an appropriate Dublin Core-based metadata set, using a Document Type Definition suitable for theses and dissertations, and marked up in XML to allow ease of repurposing.



#### 4.4 BTS and non-BTS members are Accommodated

The majority of UK university libraries are members of the British Library’s British Thesis Service, but a significant minority are not members. An E-theses Pilot service must be able to provide support for universities whether or not they are members, in order that the important work of gathering metadata on theses both hard copy and digital can be done comprehensively across the UK.

#### 4.5 All University Users are Committed to the Provision of Full-Text Theses Online in the Medium Term

Membership of the NDLTD involves a commitment to move towards the provision of full-text theses online. An E-Theses Pilot service for the UK will assist universities in this objective by identifying the stages involved in reaching the objective, and the steps required to reach each stage, in the form of a checklist which universities can apply. It should be possible to orchestrate progress for groups of universities so that several are able to move through the checklist simultaneously, and reach the objective at the same time. This would represent an efficient means of generating *Theses Alive!* in the UK, and would be an important benefit of the approach.

### 5. Objectives

The Project's objectives are:

- To test the value of a national support service for e-theses creation and management in the UK
- To develop a thesis submission system for use in all participating universities
- To develop and support a generic metadata format capable of delivering metadata to a number of relevant metadata repositories for UK thesis information
- To develop an infrastructure which enables e-theses to be published on the web to the extent that a minimum of 500 e-theses exist within the UK segment of the NDLTD after two years
- To work with other e-theses developments internationally, and in particular to assist the research aims of other e-theses projects funded within the JISC FAIR Programme.
- To produce a 'checklist approach' for universities to use as they develop e-theses capability.

### 6. Key deliverables

#### 6.1 Thesis Metadata

The Project will propose and develop a Document Type Definition (DTD) appropriate for UK theses, encompassing both printed and digital manifestations of the work. This DTD will be based upon the use of Dublin Core metadata. It will be both generic (so accommodating the various metadata needs of thesis metadata agencies currently in the UK) and extensible, so that future adaptations are possible.

Both Virginia Tech, with its Electronic Theses and Dissertations software, and the publishers of *Index to Theses*, Expert Information, have grappled with the particular problems of thesis metadata. These occur in the case of special characters – e.g. symbolic notation, or foreign language characters with diacritics. The template designed for *Index to Theses* proposed a solution based upon the option to attach files in a range of supported formats which would allow the special characters to be properly rendered. However, this is a solution which is not appropriate for metadata, which requires to be indexable and searchable. Expert Information are currently looking at a solution already provided by VT-EDT, though not currently in a user-friendly way, which is to provide an interface which supports the insertion of special characters in HTML format.

## 6.2 Submission System

While a considerable amount of work has been done on submission system design, in the development of VT-ETD and in the application of that system by many of its users (particularly in the US), and by Expert Information in designing their submission template, it is clearly desirable to rationalise this development into a single supported system for use in the UK. Our recommendation is that a submission interface to the VT-ETD be developed for use by UK universities, drawing extensively on the work of Expert Information, and on the advice of the VT-ETD team at Virginia Tech, in order to provide a single, user-friendly system which meets the requirements shown above, and meets current criteria for usability and accessibility. This system will be designed with a simple web interface and incorporate 'portlet' technology so that it can be included in university library web sites. It will be based on the assumption that these authors will provide some of the relevant metadata – in conformance with the DTD described above – and that this will be quality-controlled and supplemented with metadata provided by library staff. It will accommodate a range of digital object formats in order to permit theses as far as possible to retain their 'born digital' native formats upon submission. It will allow the generation of PDF versions from as many native formats as possible. As with the DTD itself, this system will be extensible and will continue to be developed after launch for the remainder of the Project.

## 6.3 Multimedia Content

One of the advantages of providing theses in digital form is that they can be linked to supplementary material of the sort which would not normally be possible in the traditional presentation of a thesis in either print or microform. It is thus possible for theses authors to include links to tables of statistics, or to animated or streamed content such as applets or video files. They may then be linked and opened within a browser or other viewer or player window, or they may be apparently embedded within the body of the digital thesis, but stored in a separate file, and converged at the point of display. Using the VT-ETD software, these materials can be attached to the relevant thesis metadata alongside the file containing the body of the thesis. The system developed in the *Theses Alive!* project will be supportive of a wide variety of commonly used file formats for the playing of multimedia files of all types, as well as of the common file formats for storing textual material.

## 6.4 Digital Preservation

There is a considerable challenge presented by the need to provide a digital preservation service for e-theses, and costs involved in providing such a service. It is recommended that the Pilot service should equip itself to provide general advice to universities on practical means of preserving theses files in their own institutions, and on the options for using third party 'data vaults' for preservation. The service should also consider the options for a national digital preservation service for electronic theses in the UK, looking at the benefits of a centralised data vault service; institutional data vaults and the possibilities of a shared mirroring approach such as that developed by the LOCKSS<sup>3</sup> service. It should make recommendations to JISC on the optimum solution for UK electronic theses.

---

<sup>3</sup> lockss.stanford.edu

## 7. Project outline

- Months 1-4** Project start-up: creation of *Theses Alive!* web site.
- Months 5-12** Production of Version 1 of support documentation, principally a *Theses Alive!* Guidance Document for UK university libraries. Development of submission system. Development of DTD.
- Months 11-12** *Theses Alive!* support service launch; submission system and DTD launch; publicity campaign; national workshop; introduction of email-based support service.
- Months 12-18** Development and maintenance of email- and telephone-based support service; one-on-one support to universities where necessary; development of FAQ service; production and dissemination of promotional materials.
- Months 19-24** Maintenance of support service; Version 2 of Guidance Document; survey of UK universities to identify adoption and progress rate; production of a report to JISC on *Theses Alive!* progress and estimate of timescale required to achieve 50%, 75% and 100% coverage of 'born digital' theses in UK universities by *Theses Alive!* service.

## 8. Track record

Edinburgh University Library has a long history of management of theses material, and is responsible for processing, cataloguing and housing the several hundred theses produced by the University each year. It was a member of the British Thesis Service of the British Library, but left the Service in 1992 when it felt that autonomous management of the loan and sale copies of University of Edinburgh theses was better undertaken internally.

The SELLIC Project has been working in the area of digital library and learning technology research and development since 1998. In 2001 it was commissioned to carry out an important study on behalf of the UK Theses Online Group (UTOG), *Edinburgh University Library Doctoral Theses Digitisation Project*. In addition to the experience gained by the SELLIC team in exploring the potential for e-thesis production and management at Edinburgh, the Library's digitisation service acquired valuable expertise in thesis digitisation as a result of this project.

## 9. Project management

The Project would be managed by John MacColl, Sub-Librarian, EUL, on the basis of 10% of his time over the two years of the Project.

## 10. Project staff

Project staff would consist of a Software Developer (18 months), and 2 \* 0.25 Project Officer posts to provide user support, promotion and advocacy. It is likely that these roles would be conflated in a single 0.5 FTE post. A potential post-holder has already been identified from within Edinburgh University Library, and could be seconded to the Project.

## 11. Project governance

It is strongly recommended that the Project shares governance with the RGU project, so that both projects benefited from a single Advisory Board.

## 12. Dissemination

We hope to work in close conjunction with the RGU e-thesis project submitted also to the JISC FAIR Programme. The following dissemination activities will be undertaken by the

Project. Working with the RGU Project will permit these activities to be shared to the benefit of both projects.

- **Project web site:** the web site will be regularly updated by the Project Manager and Project Officers and will include news of new developments. All public project documents will be presented on the web site. In addition, the web site will include links to related sites.
- **Project leaflet:** a leaflet or flyer will be produced explaining the aims and objectives of the Project. This will be appropriate for distribution within institutions and at conferences and meetings.
- **Email announcements:** important milestones will be announced on email discussion lists such as lis-link.
- **Publications:** publications will be written for the professional and academic literature.
- **Conference presentations:** various key conferences will be targeted for dissemination opportunities.

### 13. Evaluation

Collaboration with the RGU Project will be particularly effective in the evaluation activity of the Project, and we would expect to work very closely with RGU and its project partners. Formative evaluation will consist of ongoing analysis of the Project's performance against its objectives, and formal consultation with users (library and academic staff in the partner sites). Progress against Project targets will also be reported in each six-monthly report to JISC.

The summative evaluation will employ a range of measures, including interviews with thesis authors and a statistical evaluation of the impact of the Project. The study will assess the success of the Project in meeting its own objectives and targets, as well as its wider impact on research and scholarly publishing.

### 14. Summary budget

The cost of an e-theses support pilot service as described above is shown in the following table. We would recommend that this service be funded for at least two years, and continued funding thereafter be subject to a review of the service. In fact, it is likely that a major transition to the provision of electronic theses across the UK would take up to five years to achieve. A considerable impetus could be given to the process in two years however.

We have included costs for travel and dissemination to allow for some international travel associated with training in the use of the VT-ETD system, and for the cost of attending conferences and workshops.

	<b>Salary</b>	<b>Year 1</b>	<b>Year 2</b>
Project Management	0.1 FTE @ AL5	4,634	4,819
System Technical Support and Development	1 FTE @ AD2	24,083	13,125
Hardware & Software		5,000	750
User Support Service	0.25 FTE @ AL2	6,021	6,563
Promotion and Advocacy	0.25 FTE @ AL2	6,021	6,563
Travel & Subsistence		3,000	3,090
Consumables		300	309
<b>Totals</b>		<b>49,058</b>	<b>35,219</b>

### **15. Key Contact for Proposal**

John MacColl  
Sub-Librarian  
Edinburgh University Library