

Theses Alive!

FINAL REPORT

Project Acronym		Project ID	
Project Title	Theses Alive!		
Start Date	Oct 2002	End Date	Oct 2004
Lead Institution	The University of Edinburgh		
Project Director	John MacColl		
Project Manager & contact details	Theo Andrew Edinburgh University Library George Square, Edinburgh EH8 9LJ		
Partner Institutions			
Project Web URL	http://www.thesesalive.ac.uk/		
Programme Name (and number)	FAIR		
Programme Manager	Balviar Notay/Rachel Bruce		

Document

Document Title	Final Report		
Reporting Period	<i>n/a</i>		
Author(s) & project role	Theo Andrew (Project Officer) John MacColl (Project Director)		
Date		Filename	
URL	<i>if document is posted on project web site</i>		
Access	<input type="checkbox"/> Project and JISC internal		<input type="checkbox"/> General dissemination

Document History

Version	Date	Comments
0.1	December 2004	

Table of Contents

Executive summary	p.3
Background	p.4
Aims and Objectives	p.4
Methodology	p.5
Implementation	p.6
Outputs and Results	p.9
Outcomes	p.9
Conclusions	p.10
Implications	p.11
References	p.11

Acknowledgements

The Theses Alive! project was supported and funded by the JISC Focus on Access to Institutional Resources (FAIR) programme. A number of organisations were extremely helpful during the course of this project, particularly those involved in other FAIR projects: The University of Glasgow, the University of Nottingham, Robert Gordon University, the University of Southampton. Particular thanks are also due to the original project partner sites- University of Cambridge, Cranfield University, The University of Leeds and Manchester Metropolitan University.

Executive Summary

The *Theses Alive!* project has developed on a pilot basis a distributed system for the management of electronic theses and dissertations (ETDs) in the UK, in order to take advantage of the availability of the networked digital environment. The system we developed provides both a metadata submission system, and a full-text online repository for those universities wishing to use it. To achieve these aims we initially identified six objectives that the project needed to focus on to be successful; i) Develop a digital thesis submission system for use by interested universities, ii) Develop an international standards compliant digital infrastructure which enables e-theses to be published online, iii) Develop and support a generic metadata format capable of delivering metadata to a number of relevant metadata repositories for UK thesis information, iv) Test the value of a national support service for e-theses creation and management in the UK, v) 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, and vi) Produce a 'checklist approach' for universities to use as they develop e-theses capability.

The project was based around a core team of three staff at the University of Edinburgh- a Project Director, Project Officer and Systems Developer. The Project initially consulted with a group of pilot universities with the aim to take delivery of the *Theses Alive!* software for use in their institutions. As the project proceeded user feedback and product evaluation was gained from a larger community of institutions as part of the worldwide development and user group of our chosen software package.

Initially, the project carried out an extensive evaluation of the current open-source digital repository software available to the HE community. With a suitable underlying software platform chosen (DSpace) work then concentrated on building a bespoke digital repository and thesis submission system suitable for the requirements needed for the UK HE community. This software package, called *Tapir* (Theses Alive Plug-in for Institutional Repositories), is freely available to download as a self-contained add-on to the core DSpace code from the *Theses Alive!* web site, along with supporting installation documentation. The *Tapir* had been downloaded and installed by several institutions, whose feedback has been instrumental in upgrading the software to newer versions. This development work from the *Theses Alive!* project, along with major input from the SHERPA project, has culminated in the creation of an Institutional Repository for the University of Edinburgh- the Edinburgh Research Archive (ERA).

Concurrently, a standard metadata schema for ETDs in the UK has been developed and subsequently has been implemented in the DSpace platform as part of the TAPIR software. This schema was developed in conjunction with the RGU *Electronic Theses* project, the GUL *Daedalus* project and representatives from the British Library.

During development of the software, and afterwards, the Project worked to provide a general information and user support service on ETDs. This service took the form of a mediated deposit service and ETD creation support. Practically this consisted of providing guidance for postgraduate students and supervisors on suitable file format types, scanning resolutions, conversion and system administration, through web-based technologies (email/web pages) or telephone support. This user support service was successfully piloted at Edinburgh University; however feedback from consultation with the pilot institutions indicated that a national user support service of this nature would not be appropriate. Instead we have successfully concentrated on disseminating our project findings primarily through the *Theses Alive!* website, published journal articles and conference papers.

In addition to these original aims and objectives set out in the project plan it has been necessary to investigate the effects on intellectual property rights (e.g. copyright and patents) and other legal implications (e.g. the Freedom of Information Act 2002) which arise when publishing research material online. These unforeseen problems proved to be a significant barrier to the progress of the project and the development of electronic thesis programmes in general. The solutions delivered by the *Theses Alive!* project have been published by the JISC Legal service and have already proved to be extremely valuable to the HE community. In conclusion, the E-theses service piloted by the *Theses Alive!* project has been warmly welcomed in Edinburgh. However it is clear that dedicated support from home institutions, in the form of changing the current thesis regulations to include provision for electronic submission, is required if e-theses programmes are to be successfully adopted.

Background

At its most mundane an electronic thesis is a digital image of the print distribution object, being still firmly grounded in the traditions of print. However, the digital format offers a unique opportunity to create an electronic document unrestricted by conventional limitations. It is now possible to author a document that contains both aspects of multimedia and the dynamic presentation of large data sets that previously were unattainable in print format. In addition to these benefits, the digital format takes full advantage of the networked computing environment to deliver the thesis literature, which has for too long been considered intractable, online to a global audience.

The practice of making theses and dissertations available online is growing internationally. Repositories of electronic theses and dissertations are now common in universities in North America, Australia and in many European countries. Conversely, the UK has generally been relatively slow in adopting electronic theses. However, a number of initiatives have tried to promote the electronic thesis agenda within the UK HE context. The foremost of these initiatives was the UK Theses Online Group (UTOG).

In October 2001 the SELLIC team at Edinburgh University Library (EUL) presented a report to the UK Theses Online Group on the results of our Edinburgh University Library Doctoral Theses Digitisation Project. This work was generously funded by UTOG. This report contained certain recommendations for the future of the management of 'born digital' theses at Edinburgh. It 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.

The Theses Alive! project presents 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.

Aims and Objectives

The Theses Alive! initial objectives as stated in the project plan were:

1. To develop a thesis submission system for use in all participating universities
2. To develop and support a generic metadata format capable of delivering metadata to a number of relevant metadata repositories for UK thesis information
3. 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
4. To test the value of a national support service for e-theses creation and management in the UK
5. 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.
6. To produce a 'checklist approach' for universities to use as they develop e-theses capability.

Methodology

A core group of three staff at the University of Edinburgh formed the project team, which consisted of a Project Director, Project Officer and Systems Developer. At a top level the project investigated two main strands; technical development and advocacy/liaison. These discrete areas were primarily investigated by the Systems Developer and Project Officer respectively, under the management of the Project Director. However, each strand was closely interrelated to the other, allowing feedback to influence and shape the development of the associated work packages in each strand. The two strands were further subdivided into five work packages, each under the guidance of the relevant project staff member. The work packages were as follows; i) WP1: Pilot Administration, ii) WP2: Building the System, iii) WP3: Advocacy, iv) WP4: User Support, and v) WP5: Project Management. Each work package has its own approach to methodology described briefly below.

WP1: Pilot Administration

An initial aim within the project plan was to work with a set number of additional HE institutions to help test and develop the proposed e-theses management system. It was decided that a dedicated work package should be set aside to support this aim. It was envisaged that the Project Officer would arrange and liaise with a number of pilot institutions to take delivery of the Theses Alive software, to gather feedback about the system and to help provide installation and end-use support. Primarily these objectives would be achieved during face-to-face site meetings and interviews by the Project Officer, with follow up telephone meetings arranged where necessary.

WP2: Building the System

Initially, this study needed to consider which of the popular open source institutional repository software packages would be suitable for providing the infrastructure for an e-thesis management system. It was felt that a formal evaluation of the most commonly used repository platforms would provide us with the most robust approach and would eventually yield the most comprehensive and meaningful results. These evaluation results would then feedback into the design process for developing a system suitable for use in the UK context. Actual coding of the system would follow the evaluation step using a lightweight iterative process of software development.

Throughout development we were aware of the need to comply with international standards, especially with regards to software development and metadata. In the absence of appropriate standards, e.g. a UK metadata standard for theses, we felt that it was necessary to co-ordinate activities with other initiatives and projects in order to develop a meaningful outputs and results. Where internationally adopted standards were already available and in commonplace use, e.g. the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), we would try to adopt and support these common protocols.

WP3: Advocacy and WP4: User Support

Due to related content, these two work packages shared similar methodologies, thus they are described together for brevity. Both these project strands comprised working with, sharing or gathering information from end-users. This involved meeting research students, supervisors,

administrators, and library staff, initially from the University of Edinburgh, but later on from the pilot partner institutes identified in Work Package One.

To ensure that we developed and tested high-quality user support methods, the project needed to systematically gather a critical mass of current content from postgraduate students. Embarking from the project plan, which provided no strategies for gathering content, we felt that a well-planned pilot study within the UoE would be the most logical starting point for content collection. Initially this would require collaboration with the targeted academic units to set up the framework for the pilot study. Subsequently, as the content was being deposited a web- and telephone-based user support service would be required to aid the submitters.

Additionally, both these work packages are heavily involved with wider communication, to the general public and other interested researchers from the worldwide library and information science community. The most effective way to do this would be to take a multi-functional approach, targeting precise dissemination through a variety of forums, for example international conferences and publications, including both traditional (print journals) and contemporary journals (open access). Integral to this would be the establishment of the Theses Alive! website. It was a primary aim from the start of the project to make the website functional, dynamic and well-used, through regular updates and management, good presentation and abundant content which users would find beneficial.

Implementation

Software evaluation to choose an appropriate platform began early on in the project lifetime. This extensive study looked at the major open-source digital repository software packages available to the HE community at the time (early 2003), which included DSpace, ETD-db and ePrints.org. This comparison looked at some of the common elements between these packages and drew conclusions on which is the best in each field. In addition, it looked at how difficult it would be to modify each of the packages to provide an E-theses service for the UK. This analysis was considered alongside the medium-term future of each of the packages as they are developed as well as the scope for expansion that each package has within the library and also the university itself. The main part of the study considers elements particularly relevant to E-Theses such as the metadata elements and submission flow, as well as essential areas such as security and administration. A full description and detailed analysis of the packages can be found in Jones' article in *Ariadne* (Jones, 2004a).

After installing the major repository platforms and testing/evaluating them for a number of months the decision was made to use DSpace as our software platform of choice. With this decision made it was necessary to begin determining the necessary specifications for developing the extra functionality an e-thesis management system would require. In the first instance, these evaluation results fed back directly into this design process. To further identify the system and user requirements the project staff sought advice, in the form of individual face-to-face interviews, with stakeholders at the University of Edinburgh. This stakeholder group consisted of University administration (Registry), library representatives (Special Collections), Faculty (selected School Postgraduate Study Directors and research supervisors), and postgraduate students.

With a comprehensive design specification it was possible to begin the actual software development. The software itself formed a modular add-on to the DSpace core code, and has

been distributed under an open source software license, from the Theses Alive website and the SourceForge Open Source Software repository. Following a lightweight iterative process, the Theses Alive add-on software is currently on release version 0.4 (beta 1). From version 0.3 we adopted the name *Tapir*, which stands for Theses Alive Plug-in for Institutional Repositories.

The current version of *Tapir* provides the ability within DSpace to operate a supervised authoring facility, allowing Thesis and Dissertation Supervisors to observe the ongoing work by their student on their project, to comment and to even make changes. This comes with an addition to the DSpace administration area to manage the supervising groups and their access policies to the student's work. It is envisaged that although developed specifically with ETDs in mind, that this software may also find other applications. In addition, two submission interfaces (one for E-prints and other documents, and one for E-theses) are now supported, with the option to choose between them. Each of these submission interfaces provides custom metadata collection and licencing options for submissions (Jones 2004b).

Meanwhile, during the development of the *Tapir* we felt it would be beneficial to perform a baseline survey of research material already held on departmental and personal Web pages in the ed.ac.uk domain. Such a survey would be constructive on a number of levels; i) it would provide a qualitative view of Web usage across different subject areas, something that at the present time was poorly understood, ii) it would aid the initial population of the repositories by identifying ready material and willing scholarly contributors, and iii) such a survey would provide an invaluable baseline upon which progress of the project could be measured during evaluation (Andrew, 2003). The main benefit of the baseline study for the Theses Alive project was the identification of willing academic departments willing to take part in pilot electronic theses projects.

With initial contact made in suitable academic departments, the project arranged and embarked upon a six-month pilot e-theses service for two schools within the College of Science and Engineering; GeoScience and Informatics. The two Schools were chosen to represent as fully as possible a wide range of disciplines, which could have an impact on the type of e-thesis submitted. The School of Informatics, to some extent, already had a culture of producing and electronic theses, but lacked an efficient way of dissemination via the internet. The School of GeoSciences, however, had no previous experience in creating electronic theses, but were willing to embark on a pilot project. To encourage submission, we felt that an incentive was particularly needed for the GeoScience postgraduate students. To meet these aims we decided the project would pay for one thesis hard-copy to be bound for every e-thesis submitted.

The School of GeoScience includes the Institutes of Earth Sciences (Geology, Geophysics), the Institutes of Ecology and Resource Management (Ecology, Atmospheric Sciences), Geography (Human and Physical) and Meteorology. Typical theses from GeoSciences include features that could be problematic to represent in e-theses; for example, large fold out inclusions, high diagrammatical content and large auxiliary data sets. By including these types of thesis in the pilot study hoped to directly assess the impact on students and also for the repository. In total, during the 6-month pilot study 20 students completed their doctorate theses and submitted an electronic version to the Edinburgh Research Archive. The GeoScience study primarily set out to investigate and test the ERA submission procedure using content from current, or recently finished, postgraduate students. A significant

component of this study was also dedicated to providing end-user support for postgraduate students and supervisors, via telephone and web-based technologies.

The second school in the Edinburgh pilot study was the School of Informatics, which conducts research in Computer Science, Cognitive Science, Computational Linguistics and Artificial Intelligence. During the six month pilot study the project retrospectively gathered 136 electronic theses, and received 11 submitted electronic theses from recently completed postgraduate students. In contrast to the GeoScience study, the focus of the Informatics study was very much concentrated on investigating and developing a sustainable strategy for high volume ingest, including topics such as providing efficient workflow and format conversion.

Prior to the system launch, a second round of interviews occurred. A group of pilot universities were approached with the aim to take delivery of the Theses Alive! software for use in their institutions. The Theses Alive! project initially liaised with five other HE institutions, consisting of The University of Cambridge, Cranfield University, Leeds University and Manchester Metropolitan University. The initial project concept was to develop repository software to be used by these universities, and to test a national support service for this group of pilot institutions. Being unfunded these institutions did not form a formal consortium, thus any participation was purely on a voluntary basis. This scenario led to an uncomfortable situation where the project agenda did not fit with individual institutional aims and objectives, for example, different choice of repository software package. Given the project timescale it was necessary to progress ahead with development without these pilot institutions fully on-board to test the end products. Although it was disappointing not to test the system at these institutions, much valuable information was gained through consultation and liaison with these sites.

As the project progressed it became apparent that a national e-theses support service was not entirely appropriate. Although it is necessary to help institutions build repositories and appropriate policies, it was felt that other types of support, for example student support or mediated deposit, would be best offered by the home institution where embedded staff would have detailed knowledge of current working practices and procedures. Certainly this was a common opinion voiced by the partner institutions during site visits. Further user feedback, product evaluation and testing was gained from a larger community of institutions as part of the worldwide DSpace development and user group. From June–November 2004, the *Tapir* was downloaded 31 times by a number of institutions worldwide. From this worldwide user group many feature requests and programming bugs were identified, which facilitated new version releases of the *Tapir*.

During the course of Theses Alive, the project staff attended a number of international conferences on electronic theses and dissertations. These conferences allowed researchers from all over the world to share common experiences. Early on it became apparent that many institutions had achieved successful electronic theses and dissertation programmes by mandating at a top level the electronic submission of postgraduate degrees. In another departure from the project plan we have investigated and are in the process of implementing revised thesis rules and regulations for the University of Edinburgh to permit submission of electronic theses and dissertations. This has involved meeting with and engaging the University administration, writing and presenting proposals to the Postgraduate Senatus committee, rewriting the Codes of Practice for supervisors and research students, rewriting the University Degree Rules and Regulations, and finally, designing the workflow infrastructure behind the scenes to support this.

For the interested reader, the project timeline giving a detailed work breakdown is available from the Theses Alive website (at http://www.thesesalive.ac.uk/ta_timeline.shtml)

Outputs and Results

For institutions worldwide the most recognisable output from the project is the development of the Thesis Alive Plug-in for Institutional Repositories (*Tapir*). During the project lifetime a number of institutions have downloaded and installed the *Tapir* to enable supervised authoring of dissertations and theses, and general submission to their archive of e-prints and e-theses. To illustrate the diversity of institutions who have found the *Tapir* useful, organizations such as the University of Bergen (Norway), the University of Glasgow and the Texas A&M University (USA), are all DSpace/Tapir users.

Through involvement with another JISC FAIR-programme funded project (SHERPA), we were able to develop and launch the Edinburgh Research Archive (ERA)- Edinburgh University's institutional repository. This open-access digital repository of research output from the University of Edinburgh contains full-text digital Theses and Dissertations, book chapters, journal pre-prints and peer-reviewed journal reprints. During the project lifetime we have gathered c.170 electronic theses and delivered them online through ERA.

A primary aim of the Theses Alive project was 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. As part of this objective we participated in the creation of the recommended UK e-thesis core metadata set, led by the Robert Gordon University, working in conjunction with the University of Glasgow and the British Library. In addition the project has developed best practice guidelines for institutions wishing to adopt electronic theses, which are reflected in the project documentation, stored online in the Theses Alive website and also where appropriate in ERA. We have actively helped to assess and explore different mechanisms for the disclosure and sharing of content, by evaluating open source software platforms for digital repositories (e.g. Jones, 2004a) through to assessing the cultural use and impact of digital media (e.g. Andrew, 2003).

Outcomes

The technical and cultural expertise we have garnered through developing and implementing Edinburgh's live institutional repository service can, and is in the process of being disseminated to the HE information and library services community. The hard-won lessons we have learned will make this process for other institutions a much more enriched one. In addition to the core project aims covered by project documentation, we have also addressed a number of critical side issues. The resolution of these issues, in particular Intellectual Property Rights, proved to be of paramount importance, not just for project completion but also for the wider community. We have delivered a report on IPR and electronic theses which was commissioned by JISC Legal (Andrew, 2004), which stemmed directly from work carried out by the project. Included in this report are sample use and deposit licences, which were developed by project staff, and advice on the hot topic of Freedom of Information implications.

The major impact, not directly implicit in the project plan, that this project has delivered to the research community is the provision of Open Access status to selected research and thesis

literature from Edinburgh University. This has a knock on effect for enhancing teaching and learning, in that source material, e.g. book chapters and research articles, are also increasingly being made available through the repository. This toll-free access to students is available constantly without the physical lending restrictions that are traditionally associated with the published literature. Finally, the internal thesis submission and subsequent management processes at Edinburgh have been updated enhancing the students teaching and learning environment. We have investigated and are in the process of implementing revised thesis rules and regulations for the University of Edinburgh to permit submission of electronic theses and dissertations.

Above and beyond the project expectations a number of exciting opportunities and outcomes have arisen. Primarily, the Theses Alive! Systems Developer, as a result of work carried out during the project, has become highly regarded internationally in the open source software community such that he has been given the prestigious role of ‘*committer*’ for the DSpace federation. A number of individuals have administrative control over the product development, and these administrators are referred to as committers; their role is to action any changes to the core code. Only a handful of trusted developers are given this access. This role of *committer* means that Edinburgh University Library has a direct significant involvement in the future development of DSpace. This is not only a prestigious role, but highly functional in that we can ensure that the needs and requirements of the UK community are fully represented.

Conclusions

In summary, the electronic-theses service was warmly welcomed in Edinburgh by the College of Science and Engineering where a 6-month pilot service was tested. It is apparent that dedicated support from home institutions is needed to succeed. This support should not just take the form of dedicated resources, but should ideally include a commitment to establish institutional policy change. Therefore, we are currently investigating the issues involved in changing the Edinburgh University’s current thesis regulations to include future provision for electronic submission. The *Theses Alive!* project, indirectly via the Library, has suggesting policy changes to the Senatus Postgraduate Studies Committee, who tentatively accept the need for future e-theses submission.

The model which we are promoting, and to which the Postgraduate Studies Committee appears to be receptive, gives the electronic version of the thesis the status of being the authoritative version, or ‘golden copy’. Printed copies are then made from it *by the Library*. If this is accepted, then procedures will change within the University such that electronic theses become the default submission route, even before electronic deposit is mandated by University regulations.

Through meeting postgraduate supervisors and examiners it is clear that restrictions to thesis access will be needed if ETDs are to be generally accepted and used within the academic community. This has implications for institutions if they wish to comply with the newly introduced Freedom of Information Act. Further complications and implications need to be considered by the host institution when delivering thesis literature content online, especially in terms of copyright and other intellectual property rights. These problems are being looked into and advice is available from sources such as the JISC Legal Information service.

Finally, it is clear from our experiences that when a project is working with other partner institutions, the responsibilities, aims and objectives of each institution should be made clear

at the start of the project and formalised in some capacity, otherwise the final project outcomes could suffer as a result.

Implications

The Theses Alive project showed, by building a proof-of-concept service, that an electronic theses programme is an extremely worthwhile endeavour, and critically, is a viable proposition for most UK HE institutions. The findings of this project are being carried forwards by the recent JISC-funded EThOS project, in which Edinburgh University Library is a developmental partner.

References

Andrew, Theo. 2003. *Trends in Self-Posting of Research Material Online by Academic Staff*. Ariadne Issue 37. (Originating URL: <http://www.ariadne.ac.uk/issue37/andrew/intro.html>)

Andrew, Theo. 2004. *Intellectual Property and Electronic Theses*. JISC Legal commissioned report. (Originating URL: <http://www.jisclegal.ac.uk/publications/ethesesandrew.htm>)

Jones, Richard. D. 2004a. *DSpace vs. ETD-db: Choosing software to manage electronic theses and dissertations*. Ariadne Issue 38 (Originating URL: <http://www.ariadne.ac.uk/issue38/jones/intro.html>)

Jones, Richard. D. 2004b. *The Tapir: Adding E-Theses Functionality to DSpace*. Ariadne Issue 41 (Originating URL: <http://www.ariadne.ac.uk/issue41/jones/intro.html>)