# JISC Progress Report Template for FAIR Projects

February 2004

## *Overview of Project*

### Grant Statement

I confirm that the project is being conducted under the terms agreed with JISC in the letter of grant and the JISC Terms and Conditions attached to it.

## 2. Aims and Objectives

The original aims and objectives remain as stated in the project plan.

The targets set for this reporting period are set out in the following table:

| Targets | Description | Status |
|---|---|---|
| **Task 2.3** Schema development | Development of a metadata schema for use within the system | Met |
| **Task 2.4** System completion | Initial system development completed | Met |
| **Task 3.1** Site visits | Visits to pilot sites to meet with contributors | Met |
| **Task 4.3** User support set-up | Provision of a web- and telephone-based support service for pilot sites | Not Met |
| **Task 5.4** Reporting | Provision of biannual progress report | Met |

*Table 1: Targets set for reporting period.*

## 3. Overall Approach

The overall approach remains as outlined in the project plan, thus;

The project will be based around a core team of three staff at the University of Edinburgh- a Project Director, Project Officer and Systems developer. The Project will work with a group of Pilot Universities who will take delivery of the Theses Alive! Software for use in their institutions, and populate it with their own ETD data, assisting the Project team with product evaluation and user feedback. During development of the software, and afterwards, the Project will work to provide a general information and user support service on ETDs for UK HE as a whole.

## 4. Project Outputs

Since the last reporting date a number of milestones have been reached, outlined in the table above. In Workpackage Two (Building the System) tasks 2.3 and 2.4 have been completed. In the initial project plan these tasks were due to be completed in month 12 (October). Task 2.3, Development of a metadata schema, has been completed in collaboration with the RGU-led Electronic Theses Project, the Glasgow University Daedalus Project and the British Library. In December RGU hosted a meeting between representatives from these institutions to establish a standard ETD metadata schema suitable for the UK HE sector. This preliminary version of the schema was presented at the recent GUL-hosted EFAIR cluster meeting in December. Subsequently refinements from suggestions made at the cluster meeting have been made and a final version has been approved within the community. Task 2.4, Completion of initial system development, has been intrinsically linked with the development of this metadata schema. With an agreed standard ETD metadata schema available, we were able to produce an ETD submission interface for the DSpace software in month 16 (February). This interface automatically creates the metadata record from information supplied by the author during the ETD

submission process. This submission interface is available as part of the next EUL DSpace add-on, which will be released in March to our pilot partners.
Task 3.1, Visiting Pilot sites, is essentially an on-going process, originally planned for completion by Month 12 (October). Delays in system set-up (see Section 7: Risk Analysis) have meant that delays have occurred in some Workpackages. So far, three site visits to Cambridge University, Leeds University and Manchester Metropolitan University have taken place to assess the suitability of their inclusion in the *Theses Alive!* Project, and to start the initial installation process. Due to the non-funded nature of our pilot partners we need to work with institutions that are willing to support initial set-up costs, which has made finding suitable partners difficult.

## 5. Project Outcomes

To summarise the project's achievements, each of the original project objectives are evaluated with the progress currently accomplished in each area. Interim conclusions and future developments are then discussed separately.

***Objective 1: To develop an OAI-compliant thesis submission system for use in participating universities.***

Initially, the project carried out an extensive evaluation of the current open-source digital repository software available to the HE community. The results and recommendations from this study are published in Ariadne [http://www.ariadne.ac.uk/issue38/jones/]. DSpace was eventually chosen as the platform of choice for the *Theses Alive!* project. Work then concentrated on building a bespoke digital repository and thesis submission system suitable for the requirements needed for the UK HE community. Preliminary versions of this work are available to download as self-contained add-ons to the core DSpace code from the Theses Alive! web site [http://www.thesesalive.ac.uk/archive]. The latest version will be available from March.

The development work from the Theses Alive! and SHERPA projects has culminated in the creation of the Edinburgh Research Archive (ERA), built using augmented DSpace software. ERA will be launched in March and is capable of managing the University's digital research output, including e-theses and e-prints. Initial system development was due to be completed in Month 12 (October); however, a number of factors discussed in Section 7 (Risk Management), pushed back the launch date, which has somewhat delayed pilot partner participation.

***Objective 2: To develop and support a generic metadata format capable of delivering metadata to a number of relevant metadata repositories for UK thesis information.***

In conjunction with the RGU *Electronic Theses* project, the GUL *Daedalus* project and representatives from the British Library, a standard metadata schema for ETDs in the UK has been developed and has been applied to the DSpace software by the Theses Alive! project. With this in place we are now developing a crosswalk capable of delivering metadata to the British Library Theses Service, the Index to Theses and other local OPACs.

***Objective 3: To test the value of a national support service for e-theses creation and management in the UK.***

To achieve this objective we are considering a two-tiered approach. Initially, the support service will cover the Edinburgh e-theses pilot study (March-Aug), but will extend to the partner institutions once they are ready. Support comes in the form of a mediated submission service and ETD creation support. Practically this consists of providing guidance on format types, conversion and system administration, through web-based technologies (email/web pages) or telephone support. Support for Edinburgh postgraduate students and system administrators is already offered by the Project Officer and will extend to the pilot partners if needed.

***Objective 4: To develop an infrastructure which enables e-theses to be 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 2 years.***

Already in ERA we have a test-bed of circa. 30 e-theses. On top of this, over 100 retrospective theses have been identified and are available to be submitted to the digital repository. However, to properly test the full functionality of the *Theses Alive!* system we aim to capture 'born-digital' theses produced by current postgraduate students. To study this in an effective and controlled manner we have set up a six-month pilot e-theses service for two schools within the College of Science and Engineering. The two Schools have been 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 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 we can directly assess the impact on students and also for the repository. We envisage that GeoScience will have c.30 students submitting during the 6-month pilot study. The second School that will be joining the Edinburgh pilot study is the School of Informatics, which incorporates all aspects of computer science. It is estimate that circa. 50 students will submit during the pilot study. In summary, total estimates for the project end suggest circa. 200 e-theses from Edinburgh only. Cambridge University have indicated that 100 retrospective theses are already available. Pilot studies with the Departments of Engineering and Computer Science will add more born digital theses to this total. Data from the other pilot partners is not available yet.

***Objective 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.***

To date the *Theses Alive!* project is playing an active role internationally in the latest e-theses developments. Through knowledge gained whilst developing software for this Project, Edinburgh University is fast developing an international reputation for expertise in DSpace software development. Through formal and informal meetings, we are building a close relationship with the MIT-led DSpace organisation, and hopefully we will be in a position to help assist determine the future direction of DSpace development.

In addition to software development, we are disseminating our project findings through attending international conferences and meetings. To date we have presented papers at ETD-2003 in Berlin and the 3rd OAI-workshop in Bath. We have also been invited to speak at the forthcoming DSpace federation meeting in Massachusetts and are presenting two papers at the ETD-2004 conference in Kentucky. Furthermore, we are assisting other UK e-theses projects through on-campus ETD promotion. The *Theses Alive!* Project has aided the promotion of ETDs at other institutions by giving numerous presentations at local, e.g. Leeds and Glasgow universities, and national advocacy events, e.g. the RGU-led Electronic Theses Project '*Future of British Theses'* meeting in London.

All of the main conclusions from research carried out by the *Theses Alive!* project have been and will be published in appropriate venues to assist the research aims of other projects funded within the JISC FAIR programme. To date we have published three articles in Ariadne (see Section 18:Dissemination) and have plans to publish our results in other peer-reviewed journals.

***Objective 6: To produce a checklist approach of universities to use as they develop e-theses capability.***

At the present time we have advocacy material in print and on web site, e.g. ETD FAQs, but we plan to augment this with the production of an ETD adoption document compiled from the findings of the *Theses Alive!* project.

*Interim conclusions:*

- E-theses service warmly welcomed in Edinburgh by the College of Science and Engineering where a 6-month pilot service will be tested.
- Need dedicated support from home institutions to succeed. Therefore, we are 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, is suggesting policy changes to the Senatus Postgraduate Studies Committee, who tentatively accept the need for future e-theses submission.
- The Postgraduate Studies Committee is considering a request to change the submission regulations from next Academic Year (2004-2005) to mandate the submission of one electronic and one printed copy of all theses. This can only apply to students commencing their programmes from next Academic Year, and cannot apply retrospectively to students submitting theses over the next two or three years.
- 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.

In March Edinburgh will undergo a 6-month e-theses pilot project for the Schools of GeoScience and Informatics. In the longer term, e-theses are seen by EUL as a core service that will be offered to the whole University. Consequently, we now have a *hands-on* knowledge of the processes required, both technologically and culturally, in achieving electronic theses capability, which would be invaluable to institutions who are interested in going down this route.

## 6. Stakeholder Analysis

So far, in accordance with the project plan, we have been concentrating on interacting with the major stakeholders from the University of Edinburgh, both on an informal and formal basis. Prior to the pilot system launch at Edinburgh we have solicited views from future users by sending a representative to the annual School of GeoScience postgraduate conference, where postgraduate students and supervisors had a chance to voice their concerns and offer useful advice on the transition to electronic submission of theses. The Science and Engineering Library Committee and representatives from the University's Academic Affairs, Planning and Secretariat have been consulted and involved in the planning of the pilot study. Meetings have taken place with the respective chairs of Postgraduate Studies from the Schools of Informatics and GeoScience to discuss details of the pilot study. Finally, the Library, in conjunction with the Theses Alive! project, have approached the Senatus Postgraduate Studies Committee with a detailed proposal for the provision of future electronic submission of theses. So far communication with stakeholders from pilot partner institutions has been limited to initial meetings with relevant contacts within the partner institutions' libraries, as a prelude to further contact as the project progresses.

## 7. Risk Analysis

Early on in the work plan schedule the project suffered from minor delays in the hardware set up. Initially the plan was to run the joint e-theses and e-print service (ERA) from one dedicated server, however, as the project progressed it became apparent that a second server would be required for software development and disaster recovery purposes. Cumulative minor delays in the hardware acquisition, configuration and systems administration added up to a delay in service roll out of six months.

## 8. Standards

The project is employing the standards set out in the project plan, with the minor change that OAI-PMH v.2.0 is being employed in preference to v.1.1, which was available at the time of the project proposal being submitted.

## 9. Technical Development

There have been no significant changes in the technical development strategies from the original project plan.

## 10. Intellectual Property Rights

After discussions with postgraduate students and their supervisors, we have come to the conclusion that, in some cases, it would be desirable for access to electronic theses in ERA to be restricted. A number of legitimate situations for restrictions exist, for example highly sensitive research funded by industry or government sponsors. It is clear that although the aim of our project is to aim for 'Open Access' to thesis literature we must provide provision for thesis restriction at some level depending on demand. Currently, two to three theses are permanently restricted, out of a total of circa. 200 submitted in the College of Science and Engineering each year. By providing restrictions to electronic theses we would aim for a similar amount to be withheld yearly. We are currently investigating the types of restriction that would be desirable and functional for use in an electronic context, before reporting back to the University's Senatus Postgraduate Studies Committee. These restrictions will have to be applied in conformance with Freedom of Information legislation, and we are working with the University's Freedom of Information Compliance Officer to ensure that the relevant exemption clauses are also written in to the submission system.

# *Project Resources*

## 11. Project Partners

The original pilot partners institutions chosen to be included in the Theses Alive! project were Cambridge University, Cranfield University, Leeds University, Manchester Metropolitan University and Warwick University. The non-funded nature of the project partners meant that each institution had to be committed to fund their own hardware set-up and dedicate staff time to the project. After formal invitations were sent out, Warwick University were hesitant to commit, so were left out of the final plans. This will not have a detrimental effect on the project schedule.

After an initial visit to Manchester Metropolitan University there remain some concerns about the suitability of the *Theses Alive!* pilot study being hosted there, in terms of limited resources offered by the partner institution and the potential narrow scale of the pilot study. Currently we are considering the options available and the implications for the project, both in terms of timescale and objectives, before taking action.

## 12. Project Management

Work on the project has increasingly overlapped with the work we are doing within the SHERPA Project. Theo Andrew is Project Officer for both, and Richard Jones is providing systems support for SHERPA essentially as an allied activity of his work in Theses Alive!, since Edinburgh has chosen to use the DSpace platform for its SHERPA repository.

## 13. Programme Support

The Theses Alive! project has been an active member within the FAIR programme, building close links with the other members of the eFAIR cluster and e-theses sub-cluster (Glasgow and RGU), with the goal of assisting each others research aims. To date the sub-cluster has formally met twice, but contact is more frequent as we meet regularly at FAIR programme workshops and other seminars. This informal contact is vitally important because it gives individual projects the chance to discuss progress and problems with other specialists within the field. This '*cross-fertilization*' has led to the development of a UK standard ETD metadata schema, and other forms of support, such as presenting talks at advocacy seminars at other institutions, e.g. Leeds University and Glasgow University. We have sent representatives to the recent Sustainability and QA workshops and the LEADIRs seminars.

## 14. Budget

**Total JISC Grant**: **£84,277** **Co-funding:** **£3,990**

| | **Forecast budget for this reporting period (from project plan)** | **Budget for this reporting period (including any underspend or overspend)** | **Spend for this reporting period** | **Balance for this reporting period** |
|---|---|---|---|---|
| **Staff** *(list all staff with FTEs and salary scale range)* | | | | |
| Theo Andrew, 0.5 FTE, AL2 | 6,563 | 6,563 | 7,584 | -1,021 |
| Richard Jones, 0.75 FTE, AD2 | 6,563 | 6,563 | 11,376 | -4,813 |
| John MacColl, 0.1 FTE, AL5 | 2,450 | 2,450 | 2,450 | 0 |
| Hardware & software | 375 | 375 | 375 | 0 |
| **Travel & Subsistence** | 1,545 | 1,545 | 1,741 | -196 |
| **Consumables** | 155 | 155 | 286 | -131 |
| Underspend from year 1 due to late recruitment and reprofiling of Richard Jones as 0.75 FTE throughout Project | | 15,177 | | 15,177 |
| **Total** | 17,651 | 32,828 | 23,812 | 9,016 |

# *Detailed Project Planning*

## 15. Workpackages

Key Workpackage activities scheduled for completion during this reporting period are listed below with a detailed progress report:

**Task 2.3 Schema development due Oct (Month 12)- completed Feb (Month 16)**

Work was delayed on this task to accommodate a meeting facilitated by Robert Gordon University to develop a standard metadata schema suitable for electronic theses produced in the UK. This meeting

took place in early December and was attended by project staff from the RGU-led *Electronic Theses* project, the Glasgow University *Daedalus* project, the Edinburgh-led *Theses Alive!* project and representatives of the British Library. The first version of the ETD metadata schema was initially presented at the eFAIR cluster meeting in Glasgow later on in December. Feedback from the cluster meeting was incorporated into a revised schema in January and was subsequently adopted by the *Theses Alive!* project.

***Task 2.4 Core System completion due Oct (Month 12)- completed Feb (Month 16)***

The initial DSpace installation was completed back in May (Month 7). Subsequent software development work has created the preliminary Edinburgh Research Archive (ERA). Ongoing population of the repository with e-theses and e-prints throughout this reporting period has tested the functionality of the core system, which is now ready to be released after minor revisions and innovations, which include a revised e-theses submission system developed for DSpace incorporating the new ETD metadata schema. This submission system forms the basis of the new EUL-DSpace add-on pack, which is available to our project partners and other interested parties from the *Theses Alive!* website. Additional set backs were also experienced due to delays in hardware set up (see Section 7: Risk Management). These additional delays have been accommodated by the necessary wait for the metadata schema developments; therefore, they have not been overtly detrimental to the overall project timescale.

**Task 3.1 Initial Site visits due Oct (Month 12)- status ongoing**

Delays in core system completion have pushed back this task; the rationale being that the *Theses Alive!* system needs to be complete before the involvement of the pilot partner institutions. Initial site visits were made in November (Month 13) to Leeds University and Manchester Metropolitan University to discuss preliminary logistics in hardware set-up. Cambridge University and Cranfield University were not visited in this round of site visits because they already had the infrastructure set up and did not require assistance. The completion of the EUL-DSpace add-on, which facilitates thesis submission in DSpace, enables the next round of site visits to proceed. In early February (Month 16), the *Theses Alive!* project staff visited Cambridge University to begin the process of setting up a pilot e-theses service there. A site visit to Cranfield University is now in the process of being arranged.

**Task 4.3 User support set-up due Oct (Month 12)- completed Feb (Month 16) for Edinburgh, not yet rolled out for Pilot partners**

Delays in setting up are support service for users is directly related to the problems in setting up a live service. With these problems now solved it is possible for the Theses Alive project to offer a support service for users and administrators in Edinburgh University. This support will be extended to the project partner institutions as soon as their pilot e-theses services start.

**Task 5.4 Reporting due Feb (Month 16)- completed.**

Minor adjustments were made to the project plan to bring the Biannual reporting periods in line with the rest of the FAIR programme.

## 16. Evaluation

In accordance with our Project Plan, formative evaluation will be undertaken once the partner sites take delivery of the software, and our role moves more into support than development. To date, there has been some informal feedback from dissemination acitivity, most of which has been positive. One unexpected endorsement recently has been the inclusion of the results of our study *Trends in Self-Posting of Research Material Online by Academic Staff* (Ariadne 37, October 2003) in the Publisher and Library/Learning Solutions (PALS) Pathfinder Research on Web-based Repositories: Final report.[1]

[1] Mark Ware Consulting Ltd. *Pathfinder Research on Web-based Repositories: Final report* (commissioned by Publisher and Library/Learning Solutions), January 2004 www.palsgroup.org.uk/

The project objectives for the next reporting period are set out in the table below:

| Workpackage Objective | Date due | Changes |
|---|---|---|
| 1.3 E-theses Workshop | Month 18 (April) | - |
| 2.2 System programming | Month 20 (June) | - |
| 2.5 System Launch | Month 20 (June) | - |
| 3.2 Site Visits | Month 20 (June) | - |
| 4.1 FAQ creation | Month 21 (July) | Completed in Month 11 on the recommendation of the Programme manager. |
| 4.2 ETD guide creation | Month 21 (July) | - |
| 5.5 Reporting | Month18 (April) | Moved back to Month 16 to fit with FAIR programme schedule. |

*Table 3:* Objectives for the next reporting period.

## 17. Quality Assurance

Quality assurance for the project outputs is being managed in a number of ways. All output, including the project website, is subject to accessibility testing using tools recommended by UKOLN (e.g. Bobby), as well as code validation by the W3C validator service (http://validator.w3.org/). In addition, the software has been released into the DSpace community, and feedback from testers is providing insight as to the direction of development and the location of problems. In the absence of employment of a large testing team from various disciplines and backgrounds, these methods are the best for 'quick win' problem fixing and identification of potential problems before they arise. As we move forward into the pilot stage of this project we are expecting valuable feedback from our own departments as well as from other institutions, and the students who will be using the service. In this way we can expect to rapidly identify and fix problems as they arise in this environment to be sure that the service, when provided as 'live', has been moulded into the correct form for its purpose.

All development is undertaken using the recommended guidelines for the software and languages used to increase the ease with which it may be used, and the likelihood that it can be maintained in the future.

## 18. Dissemination

Successful dissemination of the project's findings during the reporting period has occurred using a number of strategies at an internal and national level. In September the Theses Alive! project staff attended and presented a paper at the 3rd OAI-Workshop in Bath. Participation in the seminar and e-theses breakout sessions has helped to raise awareness of the Theses Alive! project and the FAIR programme in general. At a more local level, Edinburgh hosted a seminar to promote the Edinburgh Research Archive and raise awareness within the HE community about some of the current issues in Scholarly Communication. The seminar was well attended by academics from Edinburgh University who heard presentations from the *Theses Alive!* project staff, representatives from *SPARC Europe* and the *SHERPA* and *DSpace@Cambridge* projects. More recently, the project attended and presented a technical talk at the RGU E-theses workshop in London. The Project Director also delivered a brief paper on *The E-Theses project at Edinburgh.* The main findings from the project so far have been written up and published in *Ariadne* (See Andrew, 2003[2] and Jones 2003[3]). This work has been extremely useful for the OAI-community and has been cited a number of times already, for example, in the Publisher and Library/Learning Solutions (PALS) report on Institutional Repositories[4] (see Section 8.4), and in the forthcoming book *'How to find information: a guide for researchers'* published by the Open University Press.

[2] 'Trends in self-posting of research material online by Edinburgh University academics', *Ariadne*, **37**. [http://www.ariadne.ac.uk/issue37/andrew/].
[3] 'DSpace versus VT-ETD DB- a comparative evaluation'. Ariadne, 38. [http://www.ariadne.ac.uk/issue38/jones/].
[4] PALS Guide to Institutional Repositories: http://www.palsgroup.org.uk/

Dissemination targets for the next reporting period include presenting six papers at national and international conferences, including the forthcoming Eprints UK workshops (Oxford/Edinburgh), the ETD-2004 conference in Kentucky and the inaugural DSpace Federation Meeting in Massachusetts. In addition, we have been invited to submit a paper to ASSIGNation, the ASLIB Social Sciences Information Group and Network journal.

## 19. Exit/Sustainability

To ensure the continuation of the service developed under the remit of this project there are a number of provisions which will be required to keep the service running:

- ensure that a developer capable of performing the required software upgrades is available
- ensure that funding for the service can be secured
- ensure that the service has found a place within standard library procedures
- ensure that staff expected to use the service have sufficient access to support and training

For the University of Edinburgh and our Theses Alive! partners who are running the software, the exit strategy will necessarily include:

- investigation into how the service sits within a library environment, and continued work to make this the case, as well as the production of guidelines for use
- the production of both extensive user and technical documentation to support the project outcomes for the purposes of future development and use
- staff training in order that multiple members of the technical staff can use and maintain the service.

This project has been run on the basis of Edinburgh both developing the platform and service, and exploring the various tasks required to gain acceptance for the concept of an e-theses service in a large research-led UK university, and to change procedures and regulations within the university which already apply to the management of printed theses. This has proved to be a large undertaking, and our initial timescale for achieving these objectives was a little too ambitious (which partly explains why the roll-out to partners has not happened on the projected timescale). It is clear to us now that a sustained effort within universities is required in three major areas of library/IT business in order to effect the changes we wish to see within the overall aims of the FAIR Programme:

1. Technical: setting up the workflows within the customisable areas of the add-on DSpace software; understanding the operating environment and system maintenance routines
2. Content: implementing the UK standard ETD metadata schema
3. Liaison and advocacy: presenting the case to committees, Deans, senior university managers and individual academics; tailoring the case according to the working patterns of different disciplines; working with submitting students and their supervisors; dovetailing an ETD service to existing library routines for printed theses management, including binding procedures; making the case for the e-thesis as the ‘golden copy.’

The experience we have gained has been hard-won and very valuable, and we would be interested in any follow-on programme or extension to FAIR which allowed us to take not only the Theses Alive! software, based on DSpace, but also the ‘Edinburgh experience’ out to those universities which are still hesitant about committing to an ETD service. From recent dissemination events with which we have been involved, we believe that most UK universities are still reluctant to commit to such a service, even though many of them have embraced the concept of using institutional repositories for other types of research publication.