

Repositories and Preservation Proposal Cover Sheet

Cover Sheet for Proposals (All sections must be completed)		JISC Capital Programme
Name of Capital Programme: Repositories and Preservation Programme		
Bid for Call Area : (Please tick ONE BOX ONLY, as appropriate)		
Tools and Innovation (Strand B)		
<input type="checkbox"/>	Call Area I – Tools and Innovation Projects	Please specify area of proposed project eg <i>'metadata generation and validation'</i>
Discovery to Delivery (Strand C)		
	Call Area II – Discovery to Delivery Projects	<input type="checkbox"/> a) Version identification framework <input type="checkbox"/> b) Persistent identifier interoperability demonstrator <input type="checkbox"/> c) Federated access management and repositories <input type="checkbox"/> d) Semantic interoperability demonstrator
Repository Start-Up and Enhancement (Strand D)		
	Call Area III – Repository Start-Up and Enhancement Projects	<input checked="" type="checkbox"/> a) Repository start-up projects <input type="checkbox"/> b) Repository enhancement projects
Digital Preservation and Records Management (Strand H)		
	Call Area IV – Digital Preservation and Records Management Projects	<input type="checkbox"/> a) Digital preservation across the lifecycle <input type="checkbox"/> b) Models and implementation of preservation services <input type="checkbox"/> c) Preservation tools development
Shared Infrastructure Services (Strand I)		
	Call Area V – Shared Infrastructure Services Projects	<input type="checkbox"/> a) Pilot implementation of licence registry <input type="checkbox"/> b) Pilot national name and factual authority service <input type="checkbox"/> c) Scoping an architecture to support digital policy management <input type="checkbox"/> d) Scoping a terminology registry
Name of Lead Institution: University College Falmouth		
Name of Proposed Project: The Storage Space (http://storagespace.falmouth.ac.uk)		
Name(s) of Project Partner(s): Partners will be identified in the early phase of the project. Potential partners are identified within the Planning stage of the submission.		

Information in Support of JISC bid for Repository Start-Up Funding

Planning for the repository

1. University College Falmouth is a small specialist college of Art, Design and Media based in a geographically isolated location. Some prior work investigating repositories has been carried out following the identification of such technology as a solution to several local needs.
2. The College has a strong track record of partnership and collaboration, most significantly in relation to our ongoing work as a lead partner within the Combined Universities in Cornwall (CUC). No partners have been established so far for this project, although we would actively welcome this, and it is expected that several will be found within the following communities
 - Specialist Art, Media and Design colleges.
 - Academic institutions within the region including those CUC partners which we already collaborate with on projects such as the National Arts Learning Network (NALN).
 - Institutions using the same VLE software (Moodle).
 - Institutions using the repository software selected.
 - Additionally it is expected that collaboration with the University of Exeter will increase. This is currently well established at the CUC Tremough site where University of Exeter in Cornwall and University College Falmouth share resources. In particular it is expected that collaboration on the areas of access and authentication would build on earlier involvement in the SWISH project.

Aims, Outcomes & Objectives

3. We have identified a need for an Institutional Repository to meet a number of diverse objectives. An outline of the four most significant of these follows:
 - We require a location for the storage of digitized and copy right protected texts. Authorisation to enable access to particular texts will be dependent on membership of the appropriate course group. There needs to be a way of linking these objects to the institutional VLE (Moodle).
 - We require a single institutional location for staff to store key documents. Some method of document version control needs to be implemented, whether it is an IT method or an organisational method for achieving this.
 - We wish to archive student dissertations and theses in digital format. Currently, these are stored in the Library in printed format and are available to read on a reference only basis. Since we only own a single copy of each, they are vulnerable to theft and damage.
 - The Slide Library has an ongoing project to capture student degree show work. We currently have 3 years of student degree show exhibitions in storage. We would like to make this electronically available within the institution and also to the wider public.
4. We are proposing that these four projects be used as 'Early Adopter Communities' to use DSpace terminology (as outlined in website "Building an IR with DSpace"). From now on, in this bid we will refer to discrete repository units as 'communities'. It is also planned that Learning Objects (LO) be stored within the repository for use by the VLE but our current use of these is not sufficiently developed that we would consider them as a community for this project
5. The methods and policies that are thrashed out in implementing these four communities will provide a good background for assessing additional needs as they arise. It is planned that the needs of LO storage be considered and it is expected that other archival / repository needs will be identified by groups within the College as a result of this project.

Stakeholder & Operational Needs Analysis

6. Part of this bid will be to work out the policies and infrastructure needed to make the repository as a whole and the four 'early adopter' communities work. Important issues that will need addressing include:

Stakeholders and their roles

7. The cooperation, enthusiasm and involvement of many groups will have to be engaged if this project is to be a success. We need to identify the roles that these groups will perform. Roles include:
 - Developers
 - IT Technologists
 - Submitters – may be administrative staff submitting content on behalf of others
 - Content creators – could be those who submit the content also
 - Reviewers – there will need to be basic safeguards against abuse of the system
 - Metadata editors / creators – may need to use cataloguing expertise to improve record quality
 - Administrators
 - Users
8. Some of the groups we have initially identified include:
 - Principal stakeholders will be the Systems Librarian and the Learning Technologist who will form the early core development team.
 - Other major stakeholders will be the Director of Academic services and the Library staff who will have to administer the collections.
 - Lecturing Staff and to a lesser degree students will also be important in that the system developed must be attractive and easy for them to use.
 - Finally IT Services will have to offer technical support to the systems, provide data storage and meet backup and security requirements.

University College Falmouth has a converged central Academic Service structure and this will work to our advantage if we are successful with this bid. Library, IT and Learning Support staff are well used to working both strategically and operationally on projects together.

Metadata

9. For each community we will need to decide on the necessary metadata schemes such as Dublin Core, Qualified Dublin Core or VRA Core. We may need to convert, map and import existing data e.g. crossing from MARC format in our Library catalogue to Dublin Core in the repository.
10. Currently student dissertations are catalogued on the Voyager Library Catalogue. If we were to store our dissertations in the repository, students would lose the functionality of finding dissertations on our OPAC. Until the functionality to search external repositories is included in our Library catalogue, we will need to duplicate records on both systems, with links from the Voyager catalogue to the repository to provide this functionality. The duplicate records would need to be created automatically to save labour and expense.
11. See Table 1 below.

Intellectual Property Rights and Policies

12. We will need to determine the detail of any policies regarding Intellectual Property Rights, especially for material created by students. We may need to get legal advice or at least consult experts in copyright law with regards to these policies. For example, we will probably need to agree the terms of a deposit licence and a distribution licence as defined below:
13. Deposit licence: An agreement between the creator and the institution giving the repository the right to distribute the work.

14. Distribution licence: An agreement between the creator and the end user governing the uses that can be made of the work.
15. Material for which access rights have been purchased will need to have their rights protected through appropriate authorisation and authentication schemes. Dissertations and theses would ideally be made available on an 'Open Access' basis.

Authentication and Authorisation

16. The system we develop will need to protect copyright cleared material and possibly sensitive institutional documents. We will need to be able to authorise access at the individual and group level and also allow open access to some materials to all users. To achieve this, we will need to develop links between the repository system and other institutional systems.
17. We want to enable users of the repository to be able to access protected materials in the repository directly from other systems such as Moodle or the Voyager Library Catalogue without necessarily being too aware that they have entered a different environment or having to log in multiple times.
18. We will need to create appropriate links to other systems that allow the population of our chosen repository system with user accounts and authorisation / access groups.
19. To enable this we propose to implement the necessary infrastructure to be able to join the access management federation (Shibboleth) and internally a single sign on system (SSO). The details of these proposals follow in subsequent sections.

Table 1 Preliminary requirements analysis of four communities

'Community'	Digitized Texts	Institutional Documents	Dissertations and Theses	Digital Slide Archive
Data Formats	PDF Word	Word Excel PDF	Word PDF	JPG TIFF
Submitters	Library staff	Faculty Admin Academics	Library staff Faculty admin Students	Slide Library staff
Content Creators	Authors and publishers	Admin staff Academic/Technical Services staff Academics	Students	Slide Library staff Photographic technicians
Reviewers	Library staff	Senior admin staff	Library staff	Slide Librarian
Metadata editors	Library staff	Senior admin staff	Library staff	Slide Librarian
Administrators	Learning Technologist Librarian	Registry administrator	Librarian	Slide Librarian
Audience	Students enrolled on Moodle courses Authorised users	Authorised users	Open Access	Open Access
Metadata	Dublin Core	Dublin Core	Dublin Core	Qualified Dublin Core / VRA Core
Intellectual Property Rights	Publisher's Copyright cleared for groups or institution by CLA licence	University	Open Access	Open Access
Authentication for end users	Required	Required	Not required	Not required
Authorisation for end users	By group	By group	Not required	Not required

Choice of Software

20. Some preliminary work has already been carried out investigating open source repository systems such as DSpace and EPrints. In this work DSpace appeared promising. It is compatible with Microsoft Windows which is the server platform favoured by the Institutional IT Strategy.

Table 2

IT Strategy	EPrints	DSpace
MS Windows Compatible	No plans to release Windows version	DSpace 1.2.2 and above
Development platform	Perl	Java
Web technology	CGI	Tomcat

21. This work will be taken forward by establishing a small project team and obtaining appropriate hardware for a pilot study.

Specifying Enhancements

22. The project will build on some earlier work by the institution on evaluation repository systems. This work suggested that DSpace was the best match to institutional needs and this project will start by re-examining this area. If the choice of DSpace is confirmed by this then the work will move on to establish appropriate authentication links with other institutional systems and developing easy ways to insert links to resources within Moodle. Any customizations that are deemed of value to the wider community will be made available and conform to the JISC policy on Open Source Software.
23. It is anticipated that the authentication link will require establishing both a Single Sign On (SSO) apparatus within the institution and Shibboleth services. By establishing both Identity Provider (IDP) and Service Provider (SP) systems the institution will be able to be both an origin and target. Allowances have been made for servers to host these and for staff and consultants time to integrate the repository and the various existing institutional systems. Additionally some IT services time will be needed to establish the EduPerson schema and to properly populate it.
24. It is also anticipated that the link to resources within the repository will be presented as a URL. This URL could simply be copied and pasted into the institutional VLE Moodle although the IT skills required to do this albeit not major could prove a barrier to use. It is planned to create an interface to make such transfers as simple as possible. This customisation if successful will be released both to the Moodle community and to the wider community.

Risk Assessment and Risk Management

25. No existing system is being replaced so the major risk to the institution is that the time of the core development team is lost on a project which fails to deliver. After the first phase when material from the VLE and final degree shows starts to be stored the risk of lost time increases.
26. Once phase three is reached then the failure of the system will have an impact on the VLE and potentially lose final degree material. Adequate backup and fail back procedures will need to be in place to minimize this risk until such time as confidence in any new system is as high as that for existing systems.
27. The risks will be minimised by using hardware which matches the institutional server policy. Blade servers are used and in the event of a failure spare blades can be repurposed within a short space of time to cover the missing device. For storage the institutional Storage Area Network (SAN) system which has a high level of reliability and fault tolerance will be employed. It is also backed up independently as part of routine system procedures. The minimum amounts of storage which can be added to this system is large at 5Terrabyte but it should allow for several years growth in data before additional capacity is required.

Sustainability

28. It is intended that the phase 3 team be drawn from current library staff and that academic staff and students be trained to use the system. In this way it is expected that the workload will be reduced after the end of the project as the major expected benefits also begin to be experienced. There is sufficient capacity within the proposed system to cover the first few years growth and after this the hardware will be built into the institutional replacement programme. Growth in capacity when it is needed will be identified in advance and built into budgets as with any other resource.

Evaluation

29. There will be evaluation at the end of the initial phase of the work and a report will be produced identifying advantages, disadvantages and proposing a way forward. At the end of the project the extent to which the aims set out earlier have been met will be assessed and a summative report will be written. Finally once the system is in use a number of user evaluations of the system will be carried out and as well as informing future development these may if appropriate be included in an appendix to the final report.

Dissemination

30. As the project progresses we plan to update staff on developments in the Institutional newsletter. We will need to clearly document any policies and procedures that should be followed. Material promoting the advantages in using of the repository will be made available
31. Training and support in the use of the repository will be given to the necessary stakeholders:

Library staff:

- General procedures, understanding the service goals, etc.
- User interface, adding content
- Metadata procedures
- Search methods

Academics and Academic/Administrative Staff

- General procedures, understanding the service goals, etc.
- User interface, adding content
- Metadata creation

32. We will work with JISC and other institutions both locally and of a similar nature to us (a small Art, Media and Design institution). We will also work with institutions using similar systems (Moodle VLE, Voyager Library system and possibly DSpace repository). These links will be established early on in the project as we seek to identify background information and they will be developed during the project as we share experiences.
33. Finally the reports from our work will be offered to the wider community.

Work Packages

34. The work will be carried out in three phases by teams with some common membership.
35. In Phase 1 the current work on evaluating DSpace will be extended to test whether it meets all of the institutional needs in the most appropriate way. If this is not the case then alternative systems will be investigated. The issues surrounding interfaces with other institutional systems will be scoped in this phase.
36. Before Phase 2 of the project commences the repository system will have been selected and the nature of the interfaces which need to be designed will have been determined. A report from Phase 1 will be written which will set out the workflow for Phase 2.
37. In Phase 2 several tasks will be undertaken.
 - An SSO system will be set up for the institutional systems involved in the project. This will allow seamless interfacing between for example the library catalogue the VLE and the repository in the institution.

- Shibboleth IDP and SP systems will be established allowing appropriate access to the repository system to be granted to any federation member.
 - The EduPerson schema will be setup and populated
 - Required modifications will be made to the existing Library Catalogue system, the VLE the selected repository system and any other institutional system identified in Phase 1.
38. During Phase 2 material will be added to the repository to enable testing of functionality and links with other systems. Procedures for entry of material, for tagging it with appropriate metadata and for releasing it for general use will be developed and tested. Training materials will be developed to go along with this.
39. At the end of Phase 2 we expect to have a sparsely populated but fully working system and to have in place systems and procedures which will enable us to move on to the third phase.
40. In Phase 3 the repository will be populated. Initially this work will be carried out by a Learning Resources Officer and other members of the library staff. Once there is some significant material within the repository then it will be launched to the institution and training in its use both natively and as an adjunct to the VLE will be provided. It is anticipated that following this other staff will take over the population of the system and library staff will take the responsibility of met tagging the added material.

Budget

Phase 1 (estimate 3 months)

--	--	--

Phase 2 (estimate 6 months)

--	--	--

Phase 3 (estimate 12 months)

--	--	--

Internally funded	£27450.00
Externally funded (* above)	£27450.00
Total	£54900.00

Key Personnel

41. System Librarian Steve Pellow BA PgDip PGCE
Steve has been the Library Systems Administrator at University College Falmouth for five years and prior to that worked in a number of library roles over a period of eight years. He has been responsible for the conversion from Dynix to Voyager and the integration of data from Camborne School of Mines. He is the Athens administrator for the institution and oversaw the change to Athens DA setting up the necessary scripts to achieve this. He also headed the recent internal pilot study into repositories.
42. Learning Technologist Steve Power BSc MSc PGCE MInstP CPhys MBCS
Steve joined University College Falmouth last year following a long career in teaching and Local Education Authority advisory work. He taught Computing and IT for twenty years before taking over the management of an LEA IT support team where he was responsible for implementation of various wide area network projects as part of the National Grid for Learning. Since being with the institution he has researched and successfully piloted a VLE system which he continues to develop and integrate with other institutional systems.

References

JISC Repositories and Preservation Programme Briefing. Sept. 2006. JISC. 17 Nov. 2006.

Building an IR with DSpace. 2002. MIT Libraries & Hewlett-Packard Company. 17 Nov. 2006 <<http://dspace.org/implement/index.html>>.

Barton, Mary R. *Creating an Institutional Repository: LEADIRS Workbook*. 2004. MIT Libraries. 17 Nov. 2006 <<http://dspace.org/implement/leadirs.pdf>>.