


Cover Sheet for Proposals			
Project Area: <input checked="" type="checkbox"/> a) Interoperability Demonstrators <input type="checkbox"/> b) Data Audit Framework <input type="checkbox"/> c) Pilot Implementations of the data audit framework			
Name of Lead Institution: Aberystwyth University			
Name of Proposed Project: The Deposit Plait			
Name(s) of Project Partner(s): Department of Information Studies Information Services Centre for Advanced Software and Intelligent Systems			
Full Contact Details for Primary Contact:			
Name: Dr. Christine Urquhart Position: Director of Research, Department of Information Studies, Aberystwyth University Email: cju@aber.ac.uk Address: Department of Information Studies, Llanbadarn Fawr, Aberystwyth, Ceredigion, SY23 3AS Tel: 01970 622188 Fax: 01970 622190			
Length of Project: 8 months			
Project Start Date: 1 st May 2008		Project End Date: 31 st December 2008	
Total Funding Requested from JISC: £49,579			
Total Institutional Contributions: ██████████			
Outline Project Description The five elements of the project aim to: <ul style="list-style-type: none"> • Investigate the metadata requirements of repositories and service providers of deposited items • Investigate what metadata can be easily extracted from the deposited document, and if useful, whether a web service can be built to provide such metadata from a file • Investigate what metadata can be obtained from online metadata sources and personal bibliographic management software • Investigate the possibility of building a ‘Meta-API’ using the open document format web service and bibliographic data sources • Investigate what metadata needs to be verified or provided by the depositor <p>The initial element of the project will document the metadata needs of repositories and service providers (e.g. Intute Repository Search, OAIster, Google Scholar) from a deposit engine. Following that, three strands of a potential plait of data providers will be investigated to see which can provide each of the metadata needs identified. The investigations will focus on what systems and software could provide this information, what interfaces there are into these systems, what interfaces would be useful, and what services could be built upon the interfaces. The findings will be written up in a final report.</p>			
I have looked at the example FOI form at Appendix A and included an FOI form in the attached bid (Tick Box)		YES ✓	NO
I have read the Circular and associated Terms and Conditions of Grant at Appendix B (Tick Box)		YES ✓	NO

JISC Circular 5/07 – Repositories and Preservation Programme Proposal under ‘Interoperability Demonstrators’ theme:

‘The Deposit Plait’

Introduction

1. This bid from Aberystwyth University seeks funding under the capital programme in JISC circular 3/07, specifically the Interoperability Theme.

The five elements of the project aim to:

- **Investigate the metadata requirements of repositories and service providers of deposited items**
 - **Investigate what metadata can be easily extracted from the deposited document, and if useful, whether a web service can be built to provide such metadata from a file**
 - **Investigate what metadata can be obtained from online metadata sources and personal bibliographic management software**
 - **Investigate the possibility of building a ‘Meta-API’ using the open document format web service and bibliographic data sources**
 - **Investigate what metadata needs to be verified or provided by the depositor**
2. The initial element of the project will document the metadata needs of repositories and service providers (e.g. Intute Repository Search, OAIster, Google Scholar) from a deposit engine.

Following that, three strands of a potential plait of data providers will be investigated to see which can provide each of the metadata needs identified. The investigations will focus on what systems and software could provide this information, what interfaces there are into these systems, and what interfaces would be useful. The findings will be written up in a final report.

The project will commence on the 1st May 2008 and run until the 31st December 2008.

Background

3. One of the main issues affecting institutional repositories today is the lack of content, and one of the common reasons expressed by academic staff is the resentment of re-keying data.
4. The concept of a ‘deposit engine’ as outlined in the call could help to address this problem by making use of interoperable systems that can assist by gathering metadata for a given paper, aiding the academic in easing the overhead of depositing, and thus increase the corpus of works in institutional repositories.

There are many systems and techniques that could be employed by such a deposit engine. This project will investigate two of these in detail:

- What information can be easily extracted from a version of the deposited document once converted to an ‘open format’ which can easily be read by software
 - What interfaces are currently provided to publisher (commercial, learned society, open access), funder catalogues and bibliographic software that can be queried to provide metadata.
5. Combining the three strands of metadata derived from within the document, metadata about the document from external databases, and input in the form of verification or extra metadata from the depositor should create a plait strong enough to perform a deposit with sufficient

metadata about the document. The investigations can provide information about possible interfaces a deposit engine could use to make the action of depositing easier.

Project description

6. The project will make use of the expertise of three departments at Aberystwyth University: Information Services (repositories), Department of Information Studies (library systems), and the Centre for Intelligent Systems and Advanced Software (intelligent software). By combining these three areas, an effective investigation will be undertaken to examine some interfaces that could be used by a deposit engine.
7. The repository expertise from Information Services will be used to first document the metadata needs, both minimal and ideal, of a deposit, from the perspective of the repository itself (taking into account standards such as SWAP) and from the perspective of service providers such as Intute Repository Search who harvest and re-use the data.
8. The library systems expertise held by staff in the Department of Information Studies will be used to investigate current and possible interfaces into publisher, bibliographic software and funder catalogues of metadata which could be used to provide some of the metadata required by repositories. Any obvious licensing issues will be included in the investigation. Systems and interfaces to be investigated include:
 - Web of Knowledge – ISI Proceedings
 - ArticleFirst from OCLC
 - ZETOC
 - Thomson's Dialog and Datastar
 - SWETS online / BIDS / OpenSIGLE / NTIS
 - EndNote, RefWorks, and open source bibliographic software

Project staff will seek to work closely with systems providers and will work within the terms and conditions of the typical institutional licences we have to these systems.

9. The Centre for Advanced Software and Intelligent Systems (CASIS) are experts in software and intelligent systems. CASIS will investigate existing tools that examine the document to automatically extract metadata, select candidate metadata for selection by the depositor, or ask depositors to highlight the metadata. Interfaces into documents (via software products or APIs) will be investigated to see what is easily possible with existing tools, or what might be feasible with further work. Software and APIs to be investigated include:
 - Xena¹ (Xml Electronic Normalising for Archives) is a piece of software developed by the National Archives of Australia to aid long term preservation by converting files to open formats.
 - OpenOffice² is a well known desktop office application, which has the ability to read and write files in an OpenDocument format, and a machine interface to allow the conversion of a document.
 - Metadata Aggregator³ is a tool which can read and extract metadata from documents in an OpenDocument file.
 - OpenXML tools⁴.
 - Other software and systems as recommended by the OpenDocument Fellowship⁵.

CASIS will investigate, and if possible build a web service that can perform metadata extraction from OpenDocument or OpenXML files. CASIS will also be responsible for demonstrating the interoperability of candidate systems identified by the Department of Information Studies.

¹ <http://xena.sourceforge.net/>

² <http://www.openoffice.org/>

³ http://books.evc-cit.info/odf_utils/aggregator.html

⁴ <http://openxmldeveloper.org/>

⁵ <http://opendocumentfellowship.com/applications>

10. Once candidate bibliographic information provider systems have been identified, and an OpenDocument format web service created, the possibility of a Meta-API can be investigated. Such a Meta-API, perhaps provided as a web service could in theory be used to query such systems using either sparse information provided by an author, or from a file in an OpenDocument format.
11. Finally the results of each of the investigations will be brought together to see what, if any, metadata might be missing which would need to be provided by the depositor, or from other interfaces with the deposit engine. The findings will be written into a five chapter report, documenting the five elements of the project.

Project plan

12. The project will be split into three phases, first defining the requirements for a deposit, followed by investigations of possible methods of metadata population, followed by a report writing phase. Each element of the project will be a workpackage.
 - **Workpackage one: Repository and service metadata needs:** An investigation undertaken by repository research staff from Information Services into the metadata requirements of repositories and search services.
 - **Deliverables:** A report into the metadata requirements for repositories and service providers, aimed at facilitating the evaluation of the other workpackages.
 - **Effort:** 2 weeks.
 - **Workpackage two: Metadata from information stores:** An investigation undertaken by staff from the Department of Information Studies into the availability, licensing issues and interfaces to metadata information stores owned by publishers and funders. Staff from CASIS will then create light-weight demonstrators of the most potential systems.
 - **Deliverables:** A report detailing the findings of the investigation into metadata that is available from information stores.
 - **Effort:** 4 weeks + 4 weeks interoperability testing by CASIS.
 - **Workpackage three: Metadata from OpenDocument and OpenXML files:** An investigation undertaken by CASIS staff into the possibility of easily extracting metadata from OpenDocument and OpenXML files, or by tools allowing easy selection of metadata by the depositor. If possible, a web service will be created to provide this functionality, or a definition of a suitable service or API defined.
 - **Deliverables:** A report detailing the findings of the investigation into metadata that can be derived from a document.
 - **Effort:** 6 weeks.
 - **Workpackage four: Meta-API: Investigating a Meta-API:** An investigation undertaken by CASIS in conjunction with the other partners into the possibility of creating a Meta-API to query bibliographic systems and to make use of an OpenDocument file format metadata extractor.
 - **Deliverables:** A report and if possible a prototype of a Meta-API.
 - **Effort:** 6 weeks.
 - **Workpackage five: Plaiting the strands:** An investigation undertaken by repository research staff from Information Services comparing the required metadata defined by workpackage one with that which can or could be made available by the methods investigated in workpackages two and three. Where relevant the project will work with CRIG and disseminate findings via CRIG.
 - **Deliverables:** A report comparing the findings of workpackage one with the findings of workpackages two and three.
 - **Effort:** 2 weeks.
 - **Workpackage six: Project Management:** Undertake the standard management and reporting functions required by JISC.
 - **Deliverables:** Reports, plans and attendance at meetings as required by JISC.

Benefits to the community

13. The main benefits to the community will come in the reports written for each element of the project:
- Report 1: A report into the metadata requirements for repositories and service providers, aimed at facilitating the evaluation of the other workpackages.
 - This will not only provide an evaluation mechanism for the later investigations, but will be a useful document in its own right for repository managers to know what metadata to collect.
 - Report 2: A report detailing the findings of the investigation into metadata that is available from information stores.
 - This will provide a useful overview of what systems may be able to easily provide metadata to depositors about the items they are depositing.
 - Report 3: A report detailing the findings of the investigation into metadata that can be derived from a document.
 - This will inform possible future work looking into metadata extraction and the structure of information within documents to support this task.
 - Final report: A report pulling together the contents of the first three reports, and subsequent investigations into the possibility of a meta-API.
 - This report may inform future funding calls in this area.

Project Scope

14. The project will restrict itself to the deposit of Scholarly Works in the form of research papers. Books, data sets, or presentations are examples of works not covered by the project.

Sustainability and IPR

15. All reports created by the project will be licensed under a suitable Creative Commons licence, and will be deposited in a local Open Access repository, and the JISC Information Environment repository.

Project team and collaboration

16. The project team will be made up of staff from the Department of Information Studies, Information Services and the Centre for Advanced Software and Intelligent Systems.

Department of Information Studies:

- Dr. Christine Urquhart (Head of Research)
- Ms Pauline Rafferty (Senior Lecturer)
- Ms Rhian Thomas (Research Officer)

Information Services:

- Dr Mike Hopkins (Director, Information Services)
- Mr Bill Hines (Assistant Director, Information Services)
- Mr Stuart Lewis (Project Manager)
- Ms Jackie Knowles (Repository Support officer)
- Dr Talat Chaudhri (Repository Advisor)

CASIS:

- Mr Neil Taylor (Project Manager, Senior Software Engineer)
- Mrs Suzana Barreto (Software Engineer)

17. Dr Christine Urquhart was a project director on the JISC funded JUSTEIS project (1999-2004) and Rhian Thomas was a Research Officer on this longitudinal study of the trends in uptake of electronic information services by staff and students in higher and further education in the UK. More recently, Christine Urquhart and Rhian Thomas have worked on e-library evaluations for NHS Wales, and Christine Urquhart has worked with colleagues in Sheffield and the Higher Education Academy Information and Computer Sciences subject centre on procurement of electronic content across the NHS and Higher Education. Christine Urquhart has supervised a postgraduate Master's dissertation on repository use, and is supervising a PhD student examining trends in open access journals in Greek biomedical research.
18. Staff from the Information Services department at Aberystwyth University have previously worked on, and are currently working on, a line of successful repository projects including the Repository Bridge project, the Repository Support Project (RSP), the ROAD (Robot-Generated Open Access Data) project, the SWORD (Simple Web Service Offering Repository Deposit), and the Welsh Repository Network, all directed by Dr Mike Hopkins and Mr Bill Hines. Stuart Lewis leads the Web Applications and Repository Projects team, is the chair of the UK & Ireland DSpace user group, and is a DSpace committer. Ms Jackie Knowles and Dr Talat Chaudhri are well known in the UK repository world and are both closely involved with UKCoRR.
19. Staff from CASIS have previously worked on metadata projects with the National Library of Wales, working on systems to collect data to be used for preservation of digitised artefacts and to create an editing tool to allow library staff to edit and maintain METS (Metadata Encoding & Transmission Standard) documents. The METS work also involved developing solutions to process METSRights, MIX, MODS and Dublin Core information within specified sections of the METS document. CASIS has also developed a demonstration client tool as part of the JISC SWORD (Simple Web Service Offering Repository Deposit) project. Neil Taylor is a project manager within CASIS and has worked on these projects as manager or software engineer. Suzana Barreto is a software engineer and is working on XML processing software.
20. The project team will work closely with the CRIG (Common Repository Interface Group) and the WO-CRIG (Wisdom Of CRIG) project, and will attend relevant meetings. Where work in this area on similar interfaces is taking place, the team will endeavour to work alongside these initiatives in a complementary fashion. Particular importance will be placed on working with the ODF and OpenXML communities. If it is considered useful, models of potential APIs can be published as SUMS (Service Usage Models) for the eFramework.

Project risks

21. As with any project of this nature, there are risks associated. These have been summarised below:

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staff recruitment and retention	1	3	3	Staff recruitment should not be an issue as the project proposes to use internal pre-existing staff.
Staff retention	1	3	3	Due to the short nature of the project, staff retention should not pose an issue, or other departmental staff could be used.
Technical challenges	3	3	9	If some of the technical issues do not provide useful outcomes, this will still provide useful knowledge in the area.
Changing environment	3	3	9	The repository environment is a constantly changing landscape and staff will use their professional knowledge to stay abreast of these changes throughout the project.

Appendix A

FOI Withheld Information Form

We would like JISC to consider withholding the following sections or paragraphs from disclosure, should the contents of this proposal be requested under the Freedom of Information Act, or if we are successful in our bid for funding and our project proposal is made available on JISC's website.

We acknowledge that the FOI Withheld Information Form is of indicative value only and that JISC may nevertheless be obliged to disclose this information in accordance with the requirements of the Act. We acknowledge that the final decision on disclosure rests with JISC.

Section / Paragraph No.	Relevant exemption from disclosure under FOI	Justification
N/A	N/A	N/A

Appendix B

Letters of Support



Gwasanaethau Gwybodaeth
Information Services

Hugh Owen Library
Aberystwyth University
Penglais Campus
Aberystwyth
SY23 3DZ

11 January 2008

Dear Sir/Madam,

I write to confirm my support for The Deposit Plait, a bid being submitted to JISC by Aberystwyth University in response to the Call for Projects in Digital Repositories made in JISC Circular 5/07. As is indicated in the proposal Aberystwyth is currently managing several JISC funded projects in respect of repository development and has built up a pool of specialist expertise in this area that will act as a firm foundation for a project of this kind. The Deposit Plait is a natural extension of the work already undertaken and will enable experienced staff to undertake investigations that will be of considerable value to the repository community at large.

I consequently give this project proposal my wholehearted support.

Yours sincerely

Dr Mike Hopkins
Director of Information Services

Centre for Advanced Software
and Intelligent Systems (CASIS)
Department of Computer Science
University of Wales Aberystwyth
Penglais
Aberystwyth
Ceredigion Wales SY23 3DB



CENTRES OF EXCELLENCE
CANOLFANNAU RHAGORIAETH

Y Ganolfan Meddalwedd Dyrchaif
a Systemau Deallgar (CASIS)
Adran Cyfrifiadureg
Prifysgol Cymru Aberystwyth
Penglais
Aberystwyth
Ceredigion Cymru SY23 3DB

14th January 2008

JISC
University of Bristol
3rd Floor, Beacon House
Queens Road
Bristol
BS8 1QU

Dear Sir / Madam

JISC Circular 05/07: JISC Capital Programme
Digital Repositories - Strand A - Interoperability Demonstrators

I write to give my strong support for the JISC Deposit Plait bid submitted by Aberystwyth University with our colleagues from Information Services and the Department of Information Studies.

From our past experience of working in the repository environment, including the recent SWORD project, we fully appreciate the benefit of simplifying the process of depositing publications into digital repositories.

I am consequently pleased to support the project and CASIS's involvement with it.

Your sincerely

Peter Simmonds
Commercial Manager
Centre for Advanced Software and Intelligent Systems