


Cover Sheet for Proposals <i>(All sections must be completed)</i>	
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:	King's College London
Name of Proposed Project:	EIDeR - Enhanced Ingest to Digital e-research Repositories
Name(s) of Project Partner(s):	
Full Contact Details for Primary Contact: Name: Karen Stanton Position: Chief Information Officer & College Librarian Email: karen.stanton@kcl.ac.uk Address: Information Services & Systems King's College London, ISS Directorate Waterloo Bridge Wing Room 4.1 Franklin-Wilkins Building 150 Stamford Street London, SE1 9NH Tel: 020 7848 4482 Fax: 020 7848 4457	
Length of Project: 12 months	
Project Start Date: 1 st April 2008	Project End Date: 31 st March 2009
Total Funding Requested from JISC:	£38,024.86
Total Institutional Contributions: ██████████	
Outline Project Description EIDeR aims to develop a demonstrator that implements enhanced deposit and ingest for institutional repositories (IRs). The demonstrator will not be a production system – however, it will be implemented within a real-life repository environment at KCL, and the implementation will address genuine needs, in particular by reducing the need for manual intervention and providing a more streamlined deposit process. In the course of this, EIDeR will demonstrate a number of machine-to-machine interfaces referenced or implied by the scenario in Para. 18 of the Call, wherever possible building upon existing work and services. These interfaces will include preservation, automatic metadata extraction, determination of access rights/restrictions, and deposit/capture. The main focus will be on access rights and restrictions, as less prior work has been done in this area. Specific objectives of the project will be:	

- To develop detailed scenarios and use cases for enhanced ingest.
- To examine services and tools that may be used to support this functionality.
- To create a demonstrator integrating these services to implement the identified workflows.
- To evaluate the results of the project, and in particular the extent to which the available services and tools, and associated standards, support the requirements.

The deliverables of the project will include:

- A demonstrator system.
- A case study.
- A shorter report co-authored with the CRIG Support Team.
- e-Framework components, including detailed scenarios, SUMs, Service Expressions.

I have looked at the example FOI form at Appendix A and included an FOI form in the attached bid (Tick Box)	YES	
I have read the Circular and associated Terms and Conditions of Grant at Appendix B (Tick Box)	YES	

FOI declaration

Nothing in this proposal need be withheld from disclosure under the Freedom of Information Act 2000.

Proposal: EIDeR (Enhanced Ingest to Digital e-research Repositories)

Submitted under: JISC Circular 05/07: Digital Repositories

Submitted by: King's College London

1. Appropriateness, Fit to Programme Objectives, Overall Value to JISC Community

1.1 Project timescale: 12 months, from 1st April 2008 until 31st March 2009.

1.2 The aim of the project is to develop a demonstrator that implements enhanced deposit and ingest for institutional repositories (IRs). The demonstrator will not be a production system – however, it will be implemented within a real-life repository environment at KCL, and the implementation will address genuine needs, in particular by reducing the need for manual intervention and providing a more streamlined deposit process.

1.3 Specific objectives of the project will be:

- To develop detailed scenarios and use cases for enhanced ingest.
- To examine services that may be used to support this functionality (although a potential set has already been identified).
- To create a demonstrator that integrates these services to implement the identified workflows.
- To evaluate the results of the project, and in particular the extent to which the available services and tools, and associated standards, support the requirements.

1.4 The demonstrator will exercise a number of machine-to-machine interfaces to services that provide relevant functionality. Specifically, the enhanced ingest demonstrator will integrate services that support preservation, automatic metadata extraction, determination of access rights and restrictions for copyrighted material, and deposit/capture. The major part of the effort is likely to address the implementation and enforcement of access rights & restrictions, as less work has been done in this area.

1.5 The demonstrator addresses a substantial set of interfaces, exceeding what is requested by JISC in Para. 21 of the Call. However, the project will for the most part utilise and integrate services that are either already available or are currently being developed, thus offering JISC and the community good value for money, while making effective use of other JISC-funded developments.

1.6 The environment in which the demonstrator will be implemented comprises two Fedora repositories at KCL; however the service interfaces and implementations to be used are independent of the particular repository software, and the results (and services) will be applicable to repositories based on other repository software at other institutions.

1.7 Throughout the project, we will liaise closely with the CRIG and the CRIG Support Team, including participating in any WoCRIG events, in order to share issues and experience both with the CRIG and with other projects funded by the call.

2. Quality of Proposal and Robustness of Workplan

2.1 Project Background

2.1.1 The context for the demonstrator is the recent and ongoing work at King's focussing on institutional repositories. An IR based on EPrints v2 was developed in 2004 as part of the JISC-funded SHERPA-LEAP project, with the aim of storing research publications and pre-prints, and is now live.

2.1.2 However, in 2006 a decision was taken to implement a replacement IR based on Fedora. Although initially this repository too was to focus on scholarly publications, the longer-term aim was to utilise the flexibility of the Fedora framework to create a repository that was able both to handle research data and more complex digital material, and to provide the framework for a digital preservation system for all of the digital material produced by research across College.

2.1.3 The first stage of this, the creation of a Fedora-based IR for scholarly works, together with an online submission facility for researchers, is now complete. The internally-funded project to extend the IR to manage other categories of digital material started in late 2007, and is building on preliminary work carried out by KCL staff working for the Arts and Humanities Data Service (AHDS) Executive, now incorporated within KCL's newly instituted Centre for e-Research.

2.1.4 In parallel with this, as part of the JISC-funded *From Entry to EThOS* project (1 Sep. 2007–31 Dec. 2008), KCL is developing an EThOS-compliant e-thesis repository, also based on Fedora, together with an e-thesis creation and submission interface. In the live system, the intention is to use the same instance of Fedora as the research repository.

2.2 Overview of Proposed Work

2.2.1 The proposed demonstrator will be developed in the repository environment described in Section 2.1. It will demonstrate an enhancement of the existing deposit and ingest process. This scenario arises from authentic requirements, and will demonstrate a number of the interfaces identified in the call, together with others not mentioned in the call. Although the demonstrator will be built on top of a production repository environment, it will not itself be a production system.

2.2.2 Some of these interfaces demonstrated are more innovative (e.g. automated determination and enforcement of access policies, extraction of resource discovery metadata); in other cases (e.g. deposit, preservation actions) significant work has already been carried out. In all cases, however, we will build upon any existing work, and the demonstrator will wherever possible make use of services that are either already available or are currently being finalised (see Section 2.2.4). In particular, it will use several services that are being developed by JISC-funded projects at KCL's Centre for e-Research, where the development of the demonstrator will be undertaken. The main focus of the project effort will be services that implement and enforce the publishers' access conditions obtained from SHERPA/RoMEO, as less work has been done on this topic..

2.2.3 The overall scenario that will be addressed by the demonstrator will contain the following main steps. Those components that will be the main focus of the project are marked in **bold** - the other components to a great extent will require integrating services that are already being developed:

- 1) Deposit/capture: An object (publication, pre-/post-print, thesis) is captured by the system, either by user deposit or by machine-to-machine interface. We will in particular address **capture of multi-file resources (e.g. where a table or image from a paper is submitted as a separate file)**.
- 2) Preservation: Appropriate preservation actions are taken, e.g. the capture of preservation metadata (based on PREMIS) and the migration of ingested files to preservation formats.
- 3) Access rights, restrictions and conditions: Published papers, as well as pre-prints (or post-prints) of papers that are going to be (or have previously been) published, may be subject to conditions imposed by the publishers. The SHERPA/RoMEO API (<http://www.sherpa.ac.uk/romeo.php>) will be used to capture this information, which may affect how the paper can be stored in the repository, or how and when it may be made available. For example, a publisher may specify an embargo period, or require that the publisher's pdf version may not be used, or that the repository must incorporate a link to the publisher's version, or that certain permissions must be sought before making the work available. These conditions may result in updates to the metadata (e.g. in the case of an embargo), and may require certain restrictions to be enforced. We will select a number of the conditions supplied by SHERPA/RoMEO and implement services to verify and/or enforce them in the repository. We will attempt to automate this as far as possible – however, in many cases it will be necessary to prompt depositors or authors for additional information.
- 4) Resource Discovery Metadata: Currently, any resource discovery metadata is created by the depositor. We will incorporate services to extract such metadata automatically. One or more services (selected from a service registry) may be invoked, depending on the nature of the input, and the depositor will be allowed to review the data before it is lodged in the repository.

2.2.4 The demonstrator will incorporate the following existing services and tools (although other options will be examined in the initial stages of the project):

- The workflow will be constructed using the open source jBPM workflow tool, which is currently being applied at KCL (by the Arts and Humanities Data Service and the Centre for e-Research) to create ingest and preservation workflows, as part of the JISC-funded SOAPI project.
- Services to carry out automated metadata extraction have been developed as part of the JISC-funded MetaTools project, and will be used in implementing (4). Although MetaTools is not due to finish until the end of September, the required services will be available for use well before June, when WP 3 is due to start.
- The SHERPA/RoMEO project has developed an API (<http://www.sherpa.ac.uk/romeo/api.html>) to obtain publishers' access policies, which will be used in implementing (3).
- Services developed by SOAPI will also be used to implement the preservation functionality in (2) above. These services in turn incorporate third party tools (e.g. JHOVE, format conversion tools) and external registries (e.g. PRONOM), with an emphasis on generic interfaces.
- The SWORD project has developed a deposit API, which will be used in implementing (1).

2.2.5 It is anticipated that the greater part of the effort will be required to address the access rights component of the demonstrator. Although SHERPA/RoMEO provides a simple URL-based API for capturing metadata about publishers and journals, significant effort will be required to interpret the XML document returned, to map restrictions onto the repository metadata, and to verify that any conditions have been satisfied.

2.2.6 The demonstrator will primarily address scholarly works (including pre- and post-prints as well as published versions) and e-theses. Other types of digital object may be considered, although they may not be supported by the resulting demonstrator.

2.2.7 The services used will implement generic interfaces that are not dependent on the specific type of repository software used, and the results of the project will be applicable to other repository environments.

2.3 Workpackage Breakdown

Workpackage 1: Project Management

2.3.1 This workpackage includes all management activities, including planning, coordination with JISC and CRIG, reporting, and assessment of risks and opportunities as the project progresses. It also covers project advocacy and dissemination. The project will be managed in accordance with the JISC project management guidelines.

2.3.2 Deliverables will include:

- Detailed project plan; progress and risk assessment reports; website & dissemination.

Workpackage 3: Scenario and Use Case Development

2.3.3 The rough scenario outlined above will be further developed and fleshed out, and a number of use cases will be developed. This will involve liaison with selected stakeholders (see Section 3.1) and in particular with the CRIG Support Project. The use cases will take into account multiple types of digital material that may be ingested into the repository, including pre- & post- print research publications, supporting research data, and e-theses.

2.3.4 Deliverables:

- Scenarios and use cases.

Workpackage 4: Demonstrator Development

2.3.5 This workpackage includes examining available tools and services supporting the required interfaces, integrating the selected existing components to create an automated workflow, and developing any additional software (or enhancing existing software) necessary to achieve this integration. We will liaise with the CRIG Support Team regarding the development.

2.3.6 Deliverables:

- Demonstrator
- Service Usage Models and Service Expressions for the e-Framework
- Other documentation

Workpackage 6: Testing and Evaluation

2.3.7 Testing and evaluation will in the first instance be carried out by the project team. The team will subsequently work with a sample of selected researchers and other stakeholders, monitor their use of the demonstrator, and gather feedback. The processes followed by these stakeholders and their outcomes will be monitored and evaluated.

2.3.8 Deliverables:

- Testing & evaluation documents (describing tests carried out, test outcomes, issues raised).
- Modifications to demonstrator software in line with feedback.

Workpackage 7: Final reports

2.3.9 Produce a Final Report that incorporates a case study addressing the issues raised by the project, evaluation results, and recommendations based on the experiences of the project. Produce a shorter report in association with the CRIG Support Team. Produce final versions of other documentations, including SUMs, Service Expressions and other e-Framework components.

2.4 Summary of Main Deliverables

- a) Project plan.
- b) Demonstrator

- c) Final Report incorporating case study; shorter report co-authored with the CRIG Support Team.
- d) e-Framework components, including detailed scenarios, SUMs, Service Expressions

2.5 IPR

2.5.1 IPR in all reports and other documents produced by the project will be retained by the authors and host institutions but made freely available on a non-exclusive licence as required by JISC. Any software created or modified during the project will be made available to the community on an OSI-approved open-source basis on the GPL licence. We will respect the licence model of all third party software used during the project, most of which is made available under open source licences.

2.6 Risks

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Difficulties recruiting and retaining staff.	1	4	4	The nominated staff are already employed at KCL, and there is a broad knowledge of the technologies involved within KCL. Spread expertise throughout the project, document work so that knowledge is not lost.
A complete solution cannot be implemented within the project time constraints	1	3	3	The absence of a complete solution is not an indication of failure, as one aspect of the project is to investigate potential problems. The project report will address the issues that could not be resolved.
Dependencies on other ongoing projects (e.g. SOAPI, MetaTools, From Entry to EThOS).	1	2	2	Most of the software required from these projects is already available, at least in a Beta form sufficient for work to start. Final production versions will be available well before the end of the proposed project. Liaise closely with the teams producing this software, to ensure timely availability of any required software. These projects are also based at KCL and have close links with the proposed project staff. The SOAPI & MetaTools projects are actually co-located with the proposed development work, at KCL's Centre for e-Research.
Failure to meet project milestones.	2	3	6	Produce project plan with clear objectives. Continuous project assessment and rescheduling when necessary.

2.7 Outline Project Timetable

WORKPACKAGES	Month	1	2	3	4	5	6	7	8	9	10	11	12
		A 08	M	J	J	A	S	O	N	D	J	F	M 09
1: Project Management													
2: Scenarios/Use Cases													
3: Demonstrator Development													
4: Testing and Evaluation													
5: Final Reports													

3. Engagement with the Community

3.1 Engagement with and evaluation by project stakeholders

3.1.1 To ensure that the project outputs meet the needs of the stakeholders, we will follow a user-driven approach to development, involving incremental cycles of implementation and evaluation in collaboration with targeted stakeholder representatives.

3.1.2 The primary stakeholders of the project are:

Stakeholder	Interest / stake	Importance
Researchers who use institutional repositories	Potential users of enhanced deposit processes.	High
Centre for e-Research at KCL	Carrying out research and development into research infrastructures, of which digital repositories will be a major component.	High
Information Services & Systems staff at KCL.	Responsible for running future live services at KCL, based on project outputs.	Medium
The wider JSC, CRIG and repository communities.	Interest in demonstration of repository & service interfaces and their interoperability.	High
EThOSnet Advisory Board	The interfaces & workflows demonstrated will be applicable to EThOSnet and other EThOS projects.	Medium

3.1.3 Targeted representatives will include:

- Research and institutional repository staff at both KCL and other institutions (including institutions running other repository software such as DSpace and EPrints)
- EThOSnet Project staff
- Repository R&D staff at the Centre for e-Research, KCL
- ISS repository and management staff.

3.2 *Liaison with other initiatives*

3.2.1 We will liaise closely with other related projects and groups (such as standards bodies), both within the JISC Digital Repositories programme and elsewhere. In particular:

- We will participate in JISC programme-wide evaluation and support activities, in particular with a view to engaging the research communities from which potential users will be drawn.
- We will liaise, and share results and issues, with the CRIG and the CRIG Support Team.

3.3 *Sustainability*

3.3.1 Although the demonstrator is not itself sustainable, and will not form a production service, it will be developed in a real-life repository environment and will address authentic requirements using available tools and services. After the project is complete, the project outputs and experience gained will serve as a basis for further developments (that will in turn be made freely available).

3.3.2 All specific software that we intend to use is open source, and any additional software developed will be made available to the wider community on an open-source basis, in accordance with the Policy on Open Source Software for JISC Projects and Services. We will liaise with OSS Watch as regards the appropriate way of making the source available.

3.4 *Dissemination and take-up*

- Carry out all work in close coordination with the CRIG and the CRIG Support Team, as well as other members of the repository community.
- Participate in and present at related JISC workshops/programme-level activities.
- Make all documentation publicly available, e.g. on the JISC website, DigiRepWiki, and project website.

4. Budget / Value for Money

Directly Incurred Staff	Year <07-08>	Year <08-09>		TOTAL £
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
Directly Allocated				

Staff	£	£	£	£
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
Amount Requested from JISC	£6,554.14	£31,470.72	£	£38,024.86
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]

Nature of Institutional Contributions

Directly Incurred Staff				
Post, Grade & % FTE	£	£		£
Directly Incurred Non Staff				
Hardware/Software etc.	£	£		£
Directly Allocated				
Staff, Estates etc.	£	£		£
Indirect Costs				
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]

of the project's dissemination activities, thus no separate dissemination cost has been given.

5. Previous Experience of Project Team

5.1 Project Director, based at Information Services & Systems, King's College London

Role: ensure high profile of project internally and externally, take reports from Project Manager; monitor targets and delivery.

Karen Stanton – Chief Information Officer and College Librarian

Karen is Chief Information Officer, College Librarian at King's College London and service director for the Information Services and Systems (ISS) Directorate. As a member of the King's management team her responsibilities include the strategic development of a College-wide Knowledge Management Strategy encompassing the King's information environment. King's is an active partner in CURL and RUGIT collaborations and is home to a number of existing JISC projects.

5.2 Project Manager: 0.1 FTE, 12 months, based at Information Services & Systems, KCL

Role: general project management; acting as project advocate; reporting; liaison with JISC/CRIG; liaison with KCL's Centre for e-Research, which will be developing the demonstrator.

Russell Burke – Electronic Resources Co-ordinator

Russell Burke is the Electronic Resources Co-ordinator at King's College London. He manages a team responsible for the acquisition, management and delivery of the e-resources collection (including licence management and negotiation). Russell represents King's on the SHERPA-LEAP Steering and Field Officer groups, the University of London eTheses Group, the JISC Ebooks Working Group and is a project board member of the JISC National E-books Observatory Project. Russell is currently a Project Manager of the JISC funded *From Entry to EThOS* project.

5.3 Software Engineer, 0.5 FTE, 12 months, based at the Centre for e-Research, KCL

Role: Creation of scenarios/use cases (in discussion with stakeholders); design, development, integration & testing of software; documentation (including e-Framework components); liaison with CRIG and CRIG Support Team on technical issues.

Vijay Albuquerque – Software Engineer

Vijay Albuquerque has an M.Sc. in Software Engineering, and 8 years experience of software development, particularly in an agile environment. He is the senior Software Engineer at KCL's Centre for e-Research, where he has specialised in digital repositories, service-oriented architectures, and web services.

5.4 In addition, the project will be able to draw on the knowledge and expertise of KCL's Centre for e-Research, which is carrying out several projects concerning repositories and associated services.