

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 'metadata generation and validation'
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 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 checked="" 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:	Arts and Humanities Data Service	
Name of Proposed Project:	SOAPI (Service-Oriented Architecture for Preservation and Ingest of digital objects)	
Name(s) of Project Partner(s): Arts and Humanities Data Service		
Full Contact Details for Primary Contact:		
Name: Dr Mark Hedges		
Position: Technical Manager		
Email: mark.hedges@ahds.ac.uk		
Address: 26-29 Drury Lane (3rd Floor), King's College London, London WC2B 5RL		
Tel: 020 7848 1970		
Fax: 020 7848 1989		

Length of Project:	19 months	
Project Start Date:	01/03/2007	Project End Date: 30/09/2008
Total Funding Requested from JISC: £117,893.19		
Funding Broken Down over Financial Years (April – March):		
Apr06 – Mar07	Apr07 – Mar08	Apr08 – Mar09
£ 765.18	£68791.04	£48336.97
Total Institutional Contributions: £83,859.61		
Percentage Contributions over the Life of the Project:	JISC 58.4 %	Partners 41.6 %
Outline Project Description		
<p>The project will develop an architecture and toolkit for (partially) automating preservation and ingest workflows in digital repositories, based on a set of atomic web services, each encapsulating a unit of preservation functionality, combined with web-based forms when human input is required. A workflow orchestration tool will be used to combine these services and user input functions, creating workflows that minimise the need for human intervention.</p> <p>The main objectives are:</p> <ul style="list-style-type: none"> • to develop a production quality toolkit, based on this flexible and extensible architecture, that will allow repository managers to construct preservation workflows. • to supplement this work with more speculative research, investigating the use of semantic annotation of web services to allow preservation services to be dynamically discovered and executed at run-time 		
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

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.

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

Please see <http://www.ico.gov.uk> for further information on the Freedom of Information Act and the exemptions to disclosure it contains.

Proposal: SOAPI (Service-Oriented Architecture for Preservation and Ingest of digital objects)

Submitted under: JISC Circular 04/06: Repositories and Preservation Programme (Strand H, Call Area IV(c): Preservation Enhancement Tools)

Submitted by: Arts and Humanities Data Service, King's College London

1. Introduction

1.1 Most approaches to digital preservation require the existence of a significant quantity of preservation metadata for their effectiveness. While there are a number of tools to support this process, without some form of automation the generation and management of an appropriate level of metadata is not a practical proposition. Given the variety of file formats and digital object types with which a repository may have to deal, and the swiftly changing landscape of available tools, what is required is a modular software architecture that allows individual services to be easily “pluggable”. Such a system should allow services performing various preservation actions to be combined into workflows appropriate to the requirements of an individual repository.

1.2 Our approach will be to develop an architecture based on a set of atomic web services, each encapsulating a unit of preservation functionality, and which can be executed in a distributed environment. Web-based forms will be used for circumstances where a human decision or human input is required, with access controlled by Shibboleth. A workflow orchestration tool will be used to combine these services and user input functions, creating workflows that minimise the need for human intervention. In the timescale of the proposed project, we will restrict our attention to the preservation-related processing required when digital content is ingested into a repository, including the generation of technical and other preservation metadata, and the normalisation of content to preservation formats, although the model could be extended in subsequent work.

1.3 The primary objectives of the project are as follows:

- To develop a production quality toolkit, based on this flexible and extensible architecture, that will allow repository managers to construct preservation workflows.
- To further extend this work with a more speculative element, investigating the use of semantic annotations of web services to allow preservation services to be dynamically discovered and executed at run-time, in contrast to the static configuration in the basic toolkit.

1.4 The project will provide JISC with value for money, as it develops and enhances previous JISC-funded work at the AHDS, as well as maximising re-use of the available specialised third party tools and services.

1.5 Project timescale:

- 19 months, from 1st March 2007 until 30th September 2008.

1.6 The project will make a number of contributions to the JISC Capital programme and the wider JISC community:

- a) By providing a production quality toolkit that automates the preservation actions required when ingesting digital content, the project will:
 - greatly increase the productivity of repository staff, who will be able to ingest more digital content with less chance of human error.
 - facilitate the preservation of digital objects throughout their lifecycle, by enabling the production and management of preservation metadata that would have been infeasible by manual means.
- b) As the AHDS is migrating its repository to a Fedora-based system, the project will enhance the Fedora software with preservation services and workflow.
- c) The project is highly relevant to the JISC e-Framework and e-Infrastructure:
 - The proposed architecture is based on modular, distributed, re-usable web services.

- To facilitate re-usability of services, the project will develop common interfaces for generic preservation functionality. The outcome of this will inform future work on developing common standards for such interfaces.
 - The investigation into the semantic annotation, discovery and invocation of preservation web services provides a useful case study on semantic services and registries.
- d) By aiming to automate processing for a wide variety of digital objects, the project will identify areas where specialised tools are lacking, and thus assist their development.

2. Project Description

2.1 Project Background

2.1.1 One of the core activities of the AHDS is the management of a repository for the preservation and dissemination of digital resources arising from research in arts and humanities disciplines. Most of these resources result from publicly funded projects carried out by academics at UK institutions of higher education, and represent a significant investment of time and money, which the AHDS is tasked with safeguarding. The AHDS' approach to digital preservation is based on the OAIS Reference Model (ISO 14721), which provides a broad abstract framework addressing preservation functionality. More specifically, our preservation strategy may be summarised as follows:

- a) Individual files are normalised to formats more suitable for preservation. Such formats are typically non-proprietary and based on open standards, and should be of high quality, i.e. the transformation should avoid or minimise information loss.
- b) Structured metadata is generated and stored, to enable continued access to and interpretation of the preserved objects¹. If this metadata is incomplete or inaccurate, the digital object rendered to the user may differ from the intentions of the creator.
- c) Preserved objects are periodically monitored to assess their preservation status and continued accessibility, by checking file formats and rendering software for obsolescence.
- d) Files are periodically migrated to new formats, if deemed necessary by (iii).

2.1.2 In recent years, several projects have made proposals for metadata schemas to support the preservation of digital resources. The AHDS has adopted the recommendations presented in the PREMIS Data Dictionary for Preservation Metadata, a generic approach that provides a core set of technical and management metadata elements for the description of characteristics common to all digital resources, irrespective of the file format or encoding method. PREMIS metadata is built around several inter-related entity types: digital **objects**, **events** in an object's lifecycle, **agents** (person, organisation or software) that act upon objects, **relationships** between objects, and **rights** in relation to objects. The attributes identified include administrative metadata, technical metadata, and structural metadata describing how complex objects are constructed from simple objects.

2.1.3 Although preservation metadata will be generated throughout an object's lifecycle, for example when integrity checks or format migrations occur, ingest is a key occasion for generating and recording such data. Currently, digital objects are deposited with the AHDS using portable physical media, or by electronic transfer, and after deposit they undergo a formalised but largely manual ingest process before they are lodged in the repository for preservation and dissemination to users. During this process, the objects are subject to quality review and validation at various levels, such as virus checking, file corruption, data integrity; the files making up the deposit may be converted into new formats more suitable for preservation or dissemination; preservation metadata is generated.

2.1.4 This manual approach to preservation is not scaleable, however, and consequently, as part of a JISC-funded project, the AHDS has been carrying out an investigation into the feasibility of automating this deposit and ingest process². To date, the AHDS has produced a number of outputs on which this project will build:

- ingest workflow model and scoping study.
- specification of requirements for PREMIS-based preservation metadata for the AHDS.
- survey of format migration requirements

¹ Called "Representation Information" in the OAIS model.

² http://www.jisc.ac.uk/whatwedo/programmes/programme_cminfrastructure/proj_serapis.aspx;
<http://ahds.ac.uk/about/projects/Shibboleth/index.htm>

- survey and assessment of available tools and registries for performing preservation functions, including the following: tools: JHOVE, DROID, the National Library of New Zealand Metadata Extraction Tool, XENA; registries: PRONOM, GDFR, FRED.
- proof-of-concept prototype (in progress)

2.1.5 Our investigations are not yet complete, but the work carried out to date has clarified the requirements on such a system, and provides a firm basis for further work.

2.2 Project Objectives

2.2.1 In the proposed project, we will follow two strands that build on this earlier JISC-funded work:

- a) We will **develop a production-quality web service based toolkit** that will allow repository managers to build preservation and ingest workflows in a distributed environment.
- b) We will **complement this practical tool development with more speculative research**, investigating the use of semantically annotated and semantically aware preservation web services for dynamically building workflows.

a) Toolkit Development

2.2.2 The vision is to develop preservation and ingest workflows that minimise the need for human intervention, greatly reducing the overheads of repository ingest and management, with a consequent increase in productivity. Our approach is not to develop a monolithic tool, but rather a set of modular services, each encapsulating a well-defined unit of functionality at an appropriate level of granularity, which can be executed in a distributed environment. There will be dependency relationships between some services, with the output from certain services “piped” as input into others. These atomic services provide building blocks that repository managers can configure and combine to implement workflows that meet the requirements of individual repositories.

2.2.3 Of course, it will not in general be possible to perform all required processing automatically. Instead, a semi-automated system is the most practical solution, coordinating automated web services with decision points that can prompt repository staff to supply additional information (via a web interface) before the process can continue. Some form of access control will be implemented for all user input (whether by the depositor or the staff managing ingest at the repository), both for security purposes and to identify the agent who performed the action, information that will in turn be recorded in the metadata. In line with the JISC’s access management strategy, the AHDS’ own implementation of the workflow will use Shibboleth to manage all user access.

2.2.4 It is important to note that not all repositories will have the same requirements. The modular approach will enable repositories to select and configure these services, using them as building blocks from which to construct the workflows most appropriate to their own circumstances. In particular, PREMIS metadata is inclusive and wide-ranging, so a given repository may only require a subset of the available attributes. The detailed design will pay particular attention to the configurability of the metadata set used, although it will provide a number of template metadata profiles that can be used by managers who do not want the overhead of carrying out a detailed configuration.

2.2.5 Moreover, the preservation services involve tasks such as format validation, technical metadata generation, or format migration, for which numerous specialised tools and registries have been developed³. Any preservation system constructed from the proposed toolkit would want to make use of these specialised tools (converted to web services, e.g. by using SoapLab⁴). However, given that the panoply of available services is likely to change significantly over the years to come, it would be inappropriate for the proposed toolkit to be committed to using any particular set of tools, and the modular approach will facilitate the substitution of one specialised tool for another. We propose to develop generic service interfaces for certain specialist action types (e.g. conversion of a given file from source format to target format, or validation of a file against a specified format), which can be used to integrate the system with different implementations, possibly with the help of mediator services. The software bundle produced by the project will incorporate a certain subset of these

³ e.g. JHOVE for technical metadata generation; PRONOM for format validation.

⁴ <http://www.ebi.ac.uk/soaplab/>

services off the shelf, but importantly will provide configuration tools to enable different projects to make use of their own preferred tools.

2.2.6 The proposed project represents value for money, as it is further develops work either already completed or in progress at the AHDS, using techniques whose validity and usefulness have been validated by this work. The project's service-oriented approach firmly positions it within the framework of the JISC e-Infrastructure and e-Framework. To ensure that the developed software meets the needs of the community as a whole, we will follow a user-driven, evolutionary approach to this work, involving incremental cycles of implementation and assessment in collaboration with potential users. Finally, it should be emphasised that we aim to produce production quality software that is easy to install and configure, rather than a semi-experimental toolkit, and consequently we intend to invest a significant amount of effort on rigorous testing.

b) Semantic Annotation of Preservation Services

2.2.7 As observed above, preservation services can be highly specialised, and it is expected that in future a wide variety of services will be made available in a distributed fashion, for example by service providers that specialise in a particular media types. In the above toolkit, the administrator statically configures the services and registries that are to be used in particular circumstances, and reconfigures them if circumstances change. Although such an approach is feasible in the current environment, it will fail to scale as the environment of formats, tools and services expands, particularly with the advent of complex multimedia formats that are increasingly being created and captured in digital form. What is required is a mechanism for dynamically discovering services for performing specified preservation actions on specified types of digital object, and automatically invoking them as part of a workflow.

2.2.8 By using standards such as SOAP, WSDL and UDDI, web services provide a mechanism for enabling interoperability between distributed applications and automated discovery of services. However, these standards typically enforce only weak or implicit typing of data, and do not allow the semantics of web services to be represented, which restricts the potential for dynamic discovery and invocation of such services. One approach would be to require web service definitions to be more strongly validated against a set of schemas, but this would be rather inflexible, and infeasible in an environment containing many service providers.

2.2.9 Instead, our approach will be to use an ontology that allows the preservation web services to be given machine-interpretable semantic annotations and descriptions, explicitly representing knowledge about the services in a flexible and extensible way. Given suitably extended registry functionality, these annotations will allow the dynamic discovery of appropriate services by software agents, which will be able to invoke these services, combine them into workflows, and monitor the results, with no (or minimal) user interaction. We will pay close attention to previous initiatives in semantic web service description and semantic registries, most relevantly the PANIC project⁵, which has worked on the dynamic discovery, selection and invocation of individual preservation services, but also other projects in the broader e-Science and semantic web communities (particularly in bioinformatics) such as myGrid/Feta and Grimoires⁶. In particular, we will examine existing web service description ontologies such as OWL-S⁷, WSMO⁸ and the myGrid ontology⁹, as we intend if possible to use such an ontology as a foundation, extending it for preservation-specific semantics. We will concentrate on describing *functional* properties of services, such as what a service does, its inputs and outputs, and how it is invoked. To give an artificial example, a given service may perform format migration, taking Microsoft Word documents and producing RTF 1.5, only accepting versions of Word later than Word 8, and requiring the version identifier to be supplied along with the file as input. In subsequent work, this could be extended to cover non-functional properties, such as reliability or performance.

⁵ <http://metadata.net/panic/>

⁶ <http://www.grimoires.org/>;

http://www.mygrid.org.uk/index.php?module=pagemaster&PAGE_user_op=view_page&PAGE_id=57&MMN_position=64:51:63

⁷ <http://www.daml.org/services/owl-s/>

⁸ <http://www.w3.org/Submission/WSMO/>

⁹ <http://www.mygrid.org.uk/>

2.2.10 This part of the project is more research-focused – we intend the deliverable to be a proof-of-concept demonstrator rather than a production level system, although the results will form a basis for further research. Importantly, the demonstrator will aim at *extending* the production level toolkit, to facilitate future migration from a non-semantic to a semantic system.

2.3 Interoperation with other activities

2.3.1 The AHDS is submitting a concurrent proposal under *Strand B: Tools and Innovations* of the current call, which is investigating the use of semantically annotated web services for the automated generation of *descriptive* metadata. The proposals are independent, in the sense that the funding of one does not pre-suppose the funding of the other, although with their service-oriented approaches they would combine provide a powerful framework for generation of metadata that supports both preservation and discovery of digital resources.

2.3.2 Although this proposal is being submitted under *Strand H, Call Area IV(c): Preservation Enhancement Tools* of the programme, it has links with other strands of the programme:

- the second phase investigates the use of semantically annotated and semantically aware web services for the dynamic creation of service workflows, and as such is directly related to the e-Infrastructure Programme, specifically *Call VI: Semantically Coordinating Resources and Services Across Registries* (see Paragraphs E102, E116 & E117 of the call).
- part of the project would involve investigating, developing and testing common interfaces for various classes of preservation service, for instance format migration, format validation, and metadata generation services. This is related to *Strand B: Tools and Innovations* of the Repositories and Preservation programme (see Paragraph G40, bullet 10, of the call).

2.4 Workpackage Breakdown

Workpackage 1: Project Management

2.4.1 This workpackage includes all management activities, including planning, coordination with external bodies, reporting, and assess of risks and opportunities as the project progresses. It also covers project advocacy and dissemination.

2.4.2 Deliverables will include:

- Detailed project plan.
- Progress and risk assessment reports.
- Website and dissemination activities.

Workpackage 2: Use Cases

2.4.3 To ensure that the implementation is grounded in authentic user requirements, the project will first produce a set of detailed use cases relating to preservation services, and in particular the requirements for preservation metadata generation and other preservation actions during repository ingest. The use cases will expand and enhance previous work on preservation requirements carried out at the AHDS: it will increase the depth and detail of these requirements, and will also enlarge the scope to consider the requirements of other repositories. This workpackage will involve liaison with staff at AHDS subject centres and with external digital repositories (see Workpackage 6).

2.4.4 Deliverables:

- Detailed use cases.

Workpackage 3: Preservation Metadata

2.4.5 This workpackage covers the technical and implementation work related to the metadata, including:

- Mapping PREMIS entities and semantic units to actual values, or to rules stating how these values will be generated.

- Creation of XML schemas for representing PREMIS metadata entities. We will examine proposals made so far in this area, for example by the Library of Congress¹⁰, although in some cases we may use an alternative approach. For example our prior investigations suggest that PREMIS relationship metadata is more usefully expressed using RDF.
- Specification of metadata container profiles that will be used for packaging together the various digital components and preservation metadata entities associated with a digital object. Our primary approach will be based on the Metadata Encoding and Transmission Standard (METS), and will include two options: a standard METS 1.5 profile, and a profile based on the Fedora-specific extension of METS used by versions of Fedora up to 2.1¹¹. We will also investigate the use of MPEG-21 DIDL.
- Developing XSLT stylesheets, e.g. to transform XML output by third-party tools into the required format, or to create METS packages from constituent parts.

2.4.6 Deliverables:

- Preservation metadata specification.
- XML schemas; METS profiles; MPEG-21 DIDL profiles.
- XSLT stylesheets.

Workpackage 4: Software Architecture

2.4.7 This workpackage will include the development of the overall architecture for the toolkit, together with the high-level technical design of the web services, associated user input functions, and workflow tool. This design will be evaluated to ensure that it supports the use cases in Workpackage 2.

2.4.8 Deliverables:

- Software architecture document.

Workpackage 5: Software Development

2.4.9 This workpackage will include the detailed design, coding and module testing of the software components. This iterative approach has the consequence that this workpackage overlaps with WP6 (Testing and Evaluation) to a significant degree.

2.4.10 Deliverables:

- Detailed technical specification.
- Module tested software components.
- Test descriptions and logs.
- User guides.

Workpackage 6: Testing and Evaluation

2.4.11 This workpackage includes activities related to testing and evaluation of the software. To ensure the high quality of the software, we intend to invest a significant amount of effort on rigorous testing. The primary test bed will be the AHDS repository, the variety and complexity of whose collections make it a particular challenge, furnishing us with a high degree of confidence in the applicability of the toolkit. In addition, we will work in conjunction with other digital repositories to carry out external testing and evaluation of the toolkit in different technical and administrative environments. The list of participating organisations has not been finalised, but the AHDS has close links with a number of repositories with which we have collaborated on previous projects. The project budget contains an allowance for this external testing.

2.4.12 All formal testing will be fully documented: test specifications will be written; the tests will be executed and the results logged, together with any issues (such as test failures) that arise; software will be updated and rebuilt to resolve these issues where possible. We will implement issue tracking and source version control throughout testing.

2.4.13 Testing activities will cover:

¹⁰ <http://www.loc.gov/standards/premis/schemas.html>

¹¹ <http://www.fedora.info/download/2.0/userdocs/digitalobjects/introFOXML.html#OTHERXML>

- the software components, both individually and in integrated form, to ensure that the software is reliable and robust.
- the installation and configuration package, to ensure that set-up is simple even for smaller repositories that may not have in-house technical expertise.

2.4.14 Deliverables:

- Test specifications, test logs, and issue logs.
- Modified source code.

Workpackage 7: User Guides

2.4.15 This workpackage covers the production of two user guides:

- a guide for the AHDS installation of the system, aimed at resource creators who want to deposit their digital objects, and AHDS staff who manage the ingest process.
- a guide for management staff at repositories who want to install and use the software.

2.4.16 Deliverables:

- User guides.

Workpackage 8: Semantic Annotation of Services

2.4.17 This workpackage covers all activities related to the research into extending the toolkit to incorporate semantic annotation and discovery of preservation services, including:

- ontology development
- semantically annotating a selected subset of the available preservation services
- extending the AHDS test bed to incorporate a semantic service registry
- specification and execution of test scenarios, and production of a report assessing the results.

The project will liaise with external bodies in the e-Science community, for example the National e-Science Centre (NeSC) and OMII-UK, regarding this workpackage.

2.4.18 Deliverables:

- A prototype ontology for semantically annotating preservation services.
- A set of semantically annotated preservation services, extending the base toolkit.
- A demonstrator semantic registry including these services.
- A report on the investigation, including test results.

Workpackage 9: Final reports

2.4.19 Produce final versions of all documents. Many documents will be produced in the first instance as drafts, and will be subject to update during the project as a result of feedback from user evaluation and other sources. In addition, we will produce a Final Report, incorporating case studies addressing:

- the application of the toolkit in diverse repositories
- the automation of preservation and ingest functionality
- the use of workflow tools for embedding workflow in digital repositories

2.4.20 Deliverables:

- Final versions of reports and other documentation.
- Final release of software components.
- Final report, incorporating case studies.

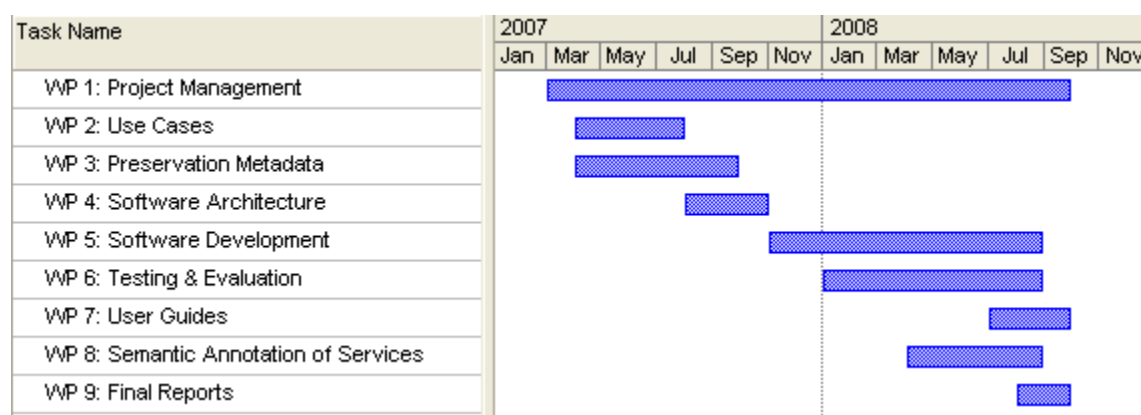
2.5 Summary of Main Deliverables

- a) A set of web services performing modular units of preservation and ingest functionality.
- b) A set of web-based functions for user input of metadata and other information that cannot be generated automatically.
- c) A tool (probably based on jBPM¹²) for combining these services and UI functions to construct semi-automated workflows.

¹² <http://www.jboss.com/products/jbpm>. A previous AHDS study included a comparative assessment of a number of workflow tools.

- d) A semi-automated preservation and ingest system for the AHDS repository based on (a)-(c).
- e) A software bundle, consisting of (a)-(c), that can be downloaded, installed and configured for use by other repositories. As the AHDS is migrating its repository to a Fedora-based system¹³, there will be a generic, repository-independent version, and a Fedora-specific extension.
- f) Report on the use of semantic annotations for discovery & invocation of preservation services, together with a prototype registry.
- g) Comprehensive documentation for the software, including:
 - Detailed technical documentation to facilitate the development of enhancements and patches for the software, both by AHDS staff and the user community.
 - Test specifications and logs.
 - User guide to enable repository management staff at other institutions to install and configure an ingest system for their own repository, using (e).
 - User guide to enable depositors and AHDS preservation staff to use the AHDS system.
- h) Final project report, incorporating case studies.

2.6 Outline Project Timetable



2.7 Risks

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing Problems (inability to attract and retain staff with appropriate skills and experience)	1 ¹⁴	4	4	Spread expertise throughout the project, create a clear project plan and document current work to ensure knowledge is not lost.
A complete solution cannot be implemented within the project time constraints	1 ¹⁵	4	4	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 projects.	1	2	2	The ingest study and prototype are on track and are very likely to be complete before the start of the project. In any case, WP1-WP3 could take place in parallel with the tail end of the previous study.

¹³ <http://fedora.info/>

¹⁴ The nominated personnel are already employed at the AHDS.

¹⁵ Given that the proposed project builds on previous work at the AHDS that validates the approach, it is considered unlikely that a solution will not be forthcoming.

2.8 IPR

2.8.1 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.

2.8.2 All software created during the project will be made available to the community on an 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.

2.9 Sustainability, dissemination and take-up

- a) All software developed during the project will be made available to the community on an open-source basis, in accordance with the *Policy on Open Source Software for JISC Projects and Services*¹⁶.
- b) The AHDS will maintain a preservation and ingest system for its repository based on the services developed during this project, and, consequently, the software is likely to be subject to enhancements and bug-fixes by AHDS staff. All changes to code will be preserved in a source repository system (using CVS or Subversion) maintained by the AHDS, and made available on an open-source basis. This will be supported by the AHDS's core funding.
- c) We will liaise with OSS Watch with a view to mirroring the AHDS source repository with an organisation such as SourceForge. The Fedora-specific extension will be offered to the Fedora development group for incorporation in their download library. The software will be made open to enhancements & patches contributed by the user community.
- d) We will produce a set of services that can be used for creating ingest systems for other digital repositories, both within UK HE/FE institutions and in the wider community. To encourage take-up, we will create an installation package (and associated user guide) that will simplify the installation and configuration process as much as possible, and we will undertake publicity activities to promote it. Widespread use of the software will encourage the creation of a self-sustaining user and developer community, ensuring longer-term sustainability.
- e) In particular, we will liaise with the Digital Curation Centre to promote the availability and use of the tools, and with OMII-UK to publicise the results of the semantic services investigation.
- f) Papers, presentations and posters at conferences and workshops.

3. Budget

Directly Incurred Staff	March 07	April 07– March 08	April 08– March 09	TOTAL £
Total Directly Incurred Staff (A)	£ 453.60	£ 55514.62	£ 29419.58	£ 85387.80
Non-Staff				
	March 07	April 07– March 08	April 08– March 09	TOTAL £
Travel and expenses	£	£ 2000.00	£1000.00	£3000.00
Hardware/software	£	£	£	£
Dissemination	£	£	£3000.00	£3000.00
Evaluation	£	£	£9000.00	£9000.00
Other	£	£	£	£

¹⁶ http://www.jisc.ac.uk/about_opensourcepolicy.html

Total Directly Incurred Non-Staff (B)	£	£2000.00	£13000.00	£15000.00
Directly Incurred Total (A+B=C) (C)	£453.60	£57514.62	£42419.58	£100387.80
Directly Allocated	March 07	April 07– March 08	April 08– March 09	TOTAL £
Staff	£	£	£	£
Estates	£53.75 £	£647.96 £ 6,489.00 £ 1,298.00	£340.29 £3,404.00 £ 681.00	£1042.00 £9,893.00 £ 1979.00
Other	£	£	£	£
Directly Allocated Total (D)	£ 53.75	£ 8434.96	£4425.29	£ 12914.00
Indirect Costs (E)	£ 368.33	£ 4440.24 £ 44,444.00 £ 8,889.00	£ 2331.43 £ 23,315.00 £4,663.00	£ 7140.00 £ 67,759.00 £ 13,552.00
Total Project Cost (C+D+E)	£875.68	£123722.82	£77154.30	£ 201752.80
Amount Requested from JISC	£ 765.18	£68791.04	£48336.97	£117893.19
Institutional Contributions	£110.50	£54931.78	£28817.33	£83859.61
Percentage Contributions over the life of the project		JISC 58.4 %	Partners 41.6 %	Total 100%

4. Key Personnel

4.1 *Project Director, 0.2 FTE, 19 months, based at the AHDS Executive Project management; acting as advocate for the project.*

Dr Mark Hedges is the Technical Manager of the AHDS. He has extensive experience of technical and project management, gained from 17 years work in the software industry. For the last 2 years, he has been manager of AHDS technical services, in particular managing a number of projects.

4.2 *Technical Officer, 1 FTE, 18 months, based at the AHDS Executive*

The Technical Officer will be responsible for all software design, development and testing, including the production of associated reports and other documentation.

Andreas Mavrides has a BS in Computer Science from the University of Arizona, and an MSc in Computing & Internet Systems from King's College London. Since September 2006, he has been working at the AHDS Executive, on JISC-funded work investigating the automation of the deposit and ingest process at the AHDS, developing a workflow model, use cases, and prototype.

4.3 *Information Officer, 0.25 FTE, 18 months, based at the AHDS Executive*

The Information Officer will be responsible for work related to the specification of preservation metadata, including the creation of relevant XML schemas and other files.

Dr Malcolm Polfreman is responsible for metadata strategy across the AHDS, and has developed significant in-house expertise in relation to metadata issues and processes. He has been a regular contributor to national and international metadata initiatives (e.g. the JISC Information Service Registry) and has written on metadata within recent JISC-funded reports, such as the Metadata Generation for Resource Discovery study (2006).

**Office of the Head
of Administration &
College Secretary**

**Head of Administration
& College Secretary**
Ian Creagh

James Clerk Maxwell Building
57 Waterloo Road
London
SE1 8WA
Tel +44 (0)20 7848 4601
Fax +44 (0)20 7848 4604
Email ian.creagh@kcl.ac.uk

KING'S
College
LONDON


University of London

To Whom It May Concern

I write to confirm, on behalf of King's College London, our full support for the proposal submitted by the Arts and Humanities Data Service entitled '**SOAPI (Service-Oriented Architecture for Preservation and Ingest)**', submitted under the JISC Circular 04/06: Repositories and Preservation Programme (Strand H, Call Area IV(c): Preservation Enhancement Tools).

This project is a significant initiative in digital preservation and we recognise its importance to the HE community.

Yours sincerely



Ian Creagh
Head of Administration and College Secretary