Strathprints Institutional Repository

Dunsire, Gordon and , JISC (Joint Information Systems Committee) (Funder) (2005) *HaIRST: harvesting institutional resources in Scotland testbed. Final project report.* [Report]

http://strathprints.strath.ac.uk/

Project

Project Acronym	HaIRST	Project ID	
Project Title	Harvesting Institutional Resources in Scotland Testbed		
Start Date	1 Aug 2002	End Date	31 Jul 2005
Lead Institution	Centre for Digital Library Research, University of Strathclyde		
Project Director	Gordon Dunsire		
Project Manager & contact details	Gordon Dunsire (g.dunsire@strath.ac.uk)		
Partner Institutions	St Andrews University Napier University Glasgow College Group (GCG) John Wheatley College		
Project Web URL	http://hairst.cdlr.strath.ac.uk/		
Programme Name (and number)	FAIR		
Programme Manager	Neil Jacobs		

Document

Document Title	Final report		
Reporting Period	n/a		
Author(s) & project role	Gordon Dunsire, Project Director		
Date		Filename	HaIRST-FAIR-FP.pdf
URL	http://hairst.cdlr.strath.ac.uk/documents/HaIRST-FAIR-FP.pdf		
Access	☐ Project and JISC internal		☐ General dissemination

Document History

Version	Date	Comments

Harvesting Institutional Resources in Scotland Testbed (HaIRST) project

Final Report

Authors

Gordon Dunsire

Lead Institution

Centre for Digital Library Research, University of Strathclyde

Contact details

Director: Gordon Dunsire, Depute Director
Address: Centre for Digital Library Research, University of Strathclyde, Livingstone
              Tower, 26 Richmond St, Glasgow G1 1XH
Email: g.dunsire@strath.ac.uk

Table of Contents

Acknowledgements

Executive Summary

The HaIRST project conducted research into the design, implementation and deployment of a pilot service for UK-wide access of autonomously created institutional resources in Scotland, the aim being to investigate and advise on some of the technical, cultural, and organisational requirements associated with the deposit, disclosure, and discovery of institutional resources in the JISC Information Environment.

The project involved a consortium of Scottish higher and further education institutions, with significant assistance from the Scottish Library and Information Council.

The project investigated the use of technologies based on the Open Archives Initiative (OAI), including the implementation of OAI-compatible repositories for metadata which describe and link to institutional digital resources, the use of the OAI protocol for metadata harvesting (OAI-PMH) to automatically copy the metadata from multiple repositories to a central repository, and the creation of a service to search and identify resources described in the central repository. An important aim of the project was to identify issues of metadata interoperability arising from the requirements of individual institutional repositories and their impact on services based on the aggregation of metadata through harvesting. The project also sought to investigate issues in using these technologies for a wide range of resources including learning, teaching and administrative materials as well as the research and scholarly communication materials considered by many of the other projects in the JISC Focus on Access to Institutional Resources (FAIR) Programme, of which HaIRST was a part.

The project tested and implemented a number of open source software packages supporting OAI, and was successful in creating a pilot service which provides effective information retrieval of a range of resources created by the project consortium institutions.

The pilot service has been extended to cover research and scholarly communication materials produced by other Scottish universities, and administrative materials produced by a non-educational institution in Scotland. It is an effective testbed for further research and development in these areas.

The project has worked extensively with a new OAI standard for "static repositories" which offers a low-barrier, low-cost mechanism for participation in OAI-based consortia by smaller institutions with a low volume of resources.

The project identified and successfully tested tools for transforming pre-existing metadata into a format compliant with OAI standards.

The project identified and assessed OAI-related documentation in English from around the world, and has produced metadata for retrieving and accessing it.

The project created a Web-based advisory service for institutions and consortia. The OAI Scotland Information Service (OAISIS) provides links to related standards, guidance and documentation, and discusses the findings of HaIRST relating to interoperability and the pilot harvesting service.

The project found that open source packages relating to OAI can be installed and made to interoperate to create a viable method of sharing institutional resources within a consortium.

HaIRST identified issues affecting the interoperability of shared metadata and suggested ways of resolving them to improve the effectiveness and efficiency of shared information retrieval environments based on OAI.

The project demonstrated that application of OAI technologies to administrative materials is an effective way for institutions to meet obligations under Freedom of Information legislation.

Background

The JISC Information Environment (JISCIE) is intended to be an online infrastructure providing secure and convenient access to a comprehensive collection of scholarly and educational material. A significant category of such resources comprises research output, scholarly papers, teaching materials, and documents supporting pedagogic management created within further and higher educational institutions by individuals and groups of staff.

The Focus on Access to Institutional Resources (FAIR) programme was funded by JISC to develop mechanisms and services to support the creation of digital repositories, where such materials can be placed by their creators and accessed by other researchers, teachers and learners. The programme was inspired by the success of the Open Archives Initiative (OAI)[1] in providing simple mechanisms that allow metadata for resources to be added to repositories, searched locally, and aggregated into services that allow resources from multiple institutions to be retrieved in a single search.

While many of the projects in FAIR explored intra-institutional issues in implementing and sustaining such repositories, the HaIRST project focussed on inter-institutional issues in aggregating metadata using the OAI mechanism of harvesting, or copying, it from multiple repositories to create "one-stop" search facilities. Such facilities are important for users of the JISCIE because they are more efficient than searching repositories individually, and are more effective because the user does not have to be familiar with multiple search interfaces or know which repositories are likely to contain resources of interest.

HaIRST also considered a wider range of institutional resources than many of the other projects in FAIR, which tended to concentrate on theses, dissertations, peer-reviewed papers, and other materials associated with the scholarly communication process. HaIRST included teaching and learning resources from further and higher education, and the output of organisational management such as meeting minutes.

HaIRST built on previous work carried out in a number of projects: CATRIONA II[2] had identified the nature and extent of institutional resources in Scottish universities; Co-operative Academic Information Retrieval Network for Scotland (CAIRNS)[3] had investigated metadata content interoperability issues in cross-searching the catalogues of university libraries in Scotland; High Level Thesaurus (HILT)[4] continues to research issues in cross-searching subjects in a wide variety of metadata repositories; Scottish Collections Network Extension (SCONE)[5] had created and implemented an embryonic service using collection-level description to help users searching across multiple collections, their local catalogues, and aggregations of collections and catalogues.

The ability to search and retrieve information about resources created by many education institutions is important in supporting the open access movement as an alternative to traditional scholarly communication, including the creation of digital libraries of theses and dissertations and open electronic journals. It is also an important tool for encouraging the formation of research partnerships and the reuse of teaching and learning materials. A single interface for retrieving administrative information from a range of institutions also assists the general enquirer in understanding the ways in which further and higher education is managed in the public sector.

Aims and Objectives

The HaIRST consortium proposed to conduct research into the design, implementation and deployment of a pilot service for UK-wide access of autonomously created institutional resources in Scotland, the aim being to investigate and advise on some of the technical, cultural, and organisational requirements associated with the deposit, disclosure, and discovery of institutional resources in the JISC Information Environment (IE).

More specifically, the project aimed to:

• Build access to a central repository of harvested institutional metadata which fully reflects locally defined data and processing requirements.
• Provide access to the repository through a Web-based search interface capable of supporting views of increasing granularity over the varied richness of the underlying metadata.
• Ensure integration of the pilot service into the JISC IE by further disclosing the harvested metadata within and beyond the IE via a number of different routes, including other OAI harvesters and Z39.50-based remote discovery services.
• Provide a model infrastructure for the design, implementation, and deployment of institutional e-archives and associated disclosure services in Scotland.
• Encourage the creation and deposit of institutional e-resources at institutions within the model infrastructure.

In particular, key deliverables were identified as:

1) A proof-of-concept suite of layered of metadata agreements, defined as XML applications and containing at least a top, universal layer based on unqualified Dublin Core, a middle layer which reifies agreements across HaIRST partners, and a bottom, institutions-specific layer which supports the locally-defined data and processing requirements.

2) Development or re-structuring of institutional and inter-institutional e-archives (from relational databases to file-based HTML repositories) capable of serving metadata adhering to some of the layers of the stack of agreements defined in 1) through server-side OAI-PMH functionality.

3) A pilot discovery service capable of regularly harvesting remote metadata from partner institutions through client-side OAI-PMH functionality, and of storing it into an XML-based backend (initially defined directly on a file-system for prototyping and testing purposes and later on a dedicated XML database system of choice).

4) A local Web-based query interface to the service in 3) capable of converting user-defined structured and partially structured queries against the metadata back-end, quantifying the portion of harvested metadata corresponding to given query structures, and including novel structure-based discovery of collection-level metadata.

5) A number of two-way metadata mappings to support further discovery and disclosure to and from a number of existing services and archives. At this stage the list includes other national and international OAI servers, the CAIRNS Z39.50-based remote discovery service, the RDN national subject-based access service, the CORC (now Connexion) shared cataloguing service run by OCLC, and collection-level metadata databases such as SCONE.

6) Associated changes in institutional cultures, policies, strategies, and organisational structures, as appropriate to, and agreed by, participants.

7) A pilot for a regional institutional e-archives advisory service, together with an associated website offering advice and guidance.

8) The involvement of Scottish FE and HE beyond the consortium institutions.

9) Draft institutional and inter-institutional collection development policy documents, covering all institutional e-resource activity, from institutionally created learning or research materials, through digitisation priorities, to the purchase of commercial research and learning materials.

10) An associated exploration of inter-institutional activity through the SCAMP gateway to facilitate collaborative collection development work in areas such as e-learning materials and digitisation programmes.

11) Report and recommendations on requirements in respect of changes in institutional culture, policy, strategy, and organisational structures, as well as on the communication protocols, metadata standards, and software implications of a service based on the various types of partner institutions.

12) A model that can be expanded to encompass other Scottish FE and HE institutions and perhaps offer an approach applicable to other areas of the UK.

13) Increased community understanding of issues through appropriate and sustained dissemination activities.

14) A full report on all project activities, together with recommendations covering further development guidelines on all areas of project activity.

Methodology

The project recognised that success in achieving its aims required addressing issues of *interoperability* on a number of levels:

• At the cultural and organisational level, it required action to ensure the existence of institutional environments that would stimulate and sustain the creation and deposit of quality resources into the IE, and of collaborative collection development policies that would support co-operative activity in the area.
• At the technical level, it required a measured agreement on, and adherence to, inter-institutional exchange protocols and metadata semantics capable of respecting local autonomy and expertise whilst promoting interoperability in line with the architectural requirements for the IE.

Accordingly, HaIRST took a "whole environment" approach to the issues addressed, encompassing the general areas mentioned above, together with associated specifics such as policies on IPR, preservation mechanisms, and similar.

Technically, the HaIRST approach to interoperability primarily relied on *harvesting* as its paradigm for remote interaction and on the Open Archive Initiative protocol for metadata harvesting (OAI-PMH)[6] as its standard interaction protocol. Metadata on research and learning resources available at partner institutions is created or mapped from pre-existing forms and then regularly harvested into a common repository for local querying and further disclosure.

In line with this "whole environment'" approach, the start dates of the various key processes were staggered and then allowed to proceed in parallel over most of the timescale of the project so that interim results coming out of any one process could inform developments in the others. Within this overall context, the methodology was follows.

Project activity was co-ordinated by a project team of staff from the Centre for Digital Library Research (CDLR)[7] at the University of Strathclyde, the lead institution, including the project director, project officer, and staff engaged in related work in other projects. The team also carried out proxy work in creating test repositories on behalf of some of the partners.

A project Management Group met at intervals of approximately 3 months to oversee progress and provide a forum for partners and the project team to discuss issues arising from the project and prioritise project activity for the next period.

The project was divided into the conceptually distinct phases of interoperable deposit at partner institutions and interoperable discovery and disclosure based at the lead institution. The two phases were interleaved to allow rapid prototyping: a first pass would experiment with sample metadata sets and "thin" metadata agreements, while a second pass would extend harvesting and discovery to full-sized metadata resources and "thicker" agreement layers. This was designed to allow the investigators to test most of the technical assumptions in a simplified scenario whilst minimising the risk of delays associated with the establishment of inter-institutional metadata agreements.

The range and characteristics of resources and associated metadata relevant to the partner institutions was investigated using desk research, a survey of the partners' requirements, discussion with partners at project meetings, and the supply of samples from each partner.

Software and related tools suitable for the creation of test repositories, harvesting metadata, and information retrieval of the harvested metadata was identified by desk research and test

installations. It was essential that the software and tools should interoperate on a technical and functional level.

The continuing rapid evolution of tools, standards and associated guidance in these areas required constant monitoring of developments at the national and international level. This was carried out using desk research and attendance at relevant meetings.

Tools for the creation of metadata, or its transformation from existing sources, in formats required for local and harvested repositories were identified through desk research and testing on samples of metadata supplied by project partners.

An advisory service was developed to stimulate activity and offer advice and support to those running institutional or inter-institutional archiving services on issues such as standards, intellectual property right, managing security, and preservation. This was initially based on Strathclyde University's Digital Information Office[8] (DIO) and subsequently developed in its own right using content originally included in the HaIRST project website.

Investigation into institutional cultures, policies, strategies and structures was carried out using desk research and group discussion. A focus group involving a mix of senior staff from the participating types of institutions, project staff, and others informed an analysis of the various institutional environments represented within the project, and aimed to identify key organisational structures in each, possible roles for these, and any changes that might be needed if the HaIRST initiative was to succeed. Small groups were set up at each type of institution to interface between project processes and the institution in question with the aim of facilitating change in both institutional and project processes as necessary. These groups fed back to the project Management Group.

Evaluation was used as a key tool in guiding activities, with the results of an early formative evaluation and interim evaluations all feeding back into and affecting other project processes. A summative evaluation was carried out at the end of the project.

It was important that the tools investigated by the project, including software, documentation and guidance meet certain characteristics: they should be open source, in keeping with the spirit of the open archives movement and to ensure wide availability; and they should conform to existing and emerging standards for OAI. The scalability of the pilot infrastructure was also a significant factor, to assure its effectiveness for operational services at regional and national level.

The project and advisory service websites were developed to meet standards for usability and are strictly compliant with XHTML.

Implementation

The project quickly created a dedicated website[9] to support and reflect project activities, both internally and externally. The website was intentionally kept simple, with further structure and material added as the need arose. The website was overhauled mid-way through the project and made XHTML compliant. It has been brought up-to-date with project documentation, and is cross-linked with the OAISIS[32]**Error! Bookmark not defined.** service.

An early task of the Resource Assessment workplan determined that three of the project partners, Glasgow College Group (GCG), John Wheatley College, and Napier University, did not intend to implement local repositories in the near future. In order to meet the project plan, they delegated to the lead institution the design, deployment, testing, hosting, and maintenance of proxy repositories which could accommodate harvesting solutions, metadata resources, and associated data resources. Each of the partners agreed to provide at least 50 data resources as a test population for their respective proxies.

It became apparent that resource identification would be an ongoing process for at least the first year of the project, and that it would be one of the main areas of project activity for the partner institutions. Partners were at varying stages in the development of their policies for digital materials, with timescales dependent on many factors including definition of which resources to offer on open access, the role of the library, and integration with local virtual learning environments. The project schedule had been designed to accommodate this possibility by adopting a two-phase approach to prototyping the testbed. For the first phase, in particular, some sample data and metadata from each partner was thought sufficient to gather early feedback on the overall feasibility of the approach.

The FE partners experienced further, serious delay when a post for a Metadata Officer was advertised but did not result in an appointment. Although the HaIRST schedule dictated that this take place during the summer vacation when it is difficult to arrange interviews, there was a significant difficulty in matching the required balance of information management and information technology skills. After further consultation, the Scottish Library and Information Council (SLIC)[10] agreed to second one of its staff to the post from December 2003, and eventually suitable sample resources were identified. The involvement of SLIC had an additional benefit in investigating the use of the OCLC Connexion[11] cataloguing service for the preparation of metadata for local requirements and its transformation into the format and content required to meet partner agreements for harvesting, with SLIC providing access to its shared account.

It also became clear that the development of pilot services at partner institutions, while remaining one of the main project objectives, would stretch further into the duration of the project than anticipated, rather than be concentrated in the first third. Comparison with pilot services produced by other FAIR projects, which were more limited in scope, implied that the HaIRST partner services might remain relatively undeveloped for the duration of the project.

It was originally intended to disclose Napier University's resources via the DLESE OAI Data Provider[12], but the publication in October 2003 by the OAI of the beta version of the "Specification for an OAI Static Repository and an OAI Static Repository Gateway"[13] was quickly identified as a preferred solution for the implementation of proxy repositories for all three partners requiring them. Nonetheless, the DLESE OAI Data Provider remains as an option available to partners to upgrade to more functional and heavyweight OAI participation models.

The static repository approach suits the case of metadata collections which do not change too frequently, are not too big (up to 5000 records, approximately), and cannot be managed

with high implementation and maintenance costs. It was found that, in this stage of the project, the collections and resource models available at Napier University and the FE partners met this type of scenario. Two static repositories were created, one for Napier University and the other a joint facility for the FE partners. Both partners provided the project with MARC21 records which either originated from pre-existing collections, as was the case for Napier University, or were created for the project using the OCLC Connexion service, as was the case for records contributed by the FE partners. The MARC records were transformed to the OAI-specified encoding of unqualified Dublin Core using the freely available MarcEdit[14] package and a MARC21-DC conversion tool available with Connexion, and then uploaded into the two dedicated static repositories using Perl scripts specifically developed for the task. The solution developed for the first sample of records was designed to be reused for incremental repository updates and the process was streamlined in collaboration with partners.

Initially the static repositories were registered with the experimental gateway service offered by the Research Library of the Los Alamos National Laboratory (LANL)[15] to allow harvesting by the project. Subsequently, the project set up its own static repository gateway using the LANL software.

St Andrews University developed its own local repository service[16] for eprints using the open source software from eprints.org, and launched it in Summer 2003. Initially, progress was slow and not dissimilar to experiences collected and disseminated by similarly FAIR projects, namely the reluctance of faculty and other interested stakeholders to convert an initial interest in the aims of an institutional repository into a more substantial commitment. Accordingly, resource aggregation could not rely on self-archiving and the service retained traditional high costs and lack of homogeneity. In turn, this slowed the formation of a user base and further promotion of the service. St Andrews University has completed a case study report on its experiences and published it via its pilot repository[17].

It was clear that implementing a full operational repository service for Strathclyde University was beyond the resources available to HaIRST. The university had not formulated a firm strategy or commitment to such a service, and project staff concluded that it would be premature and potentially damaging to future development to go beyond a minimal pilot service focused around the core functionalities of resource discovery and resource deposit.

Extensive work was carried out towards the roll-out of the pilot service for Strathclyde University, Strathprints[18]. This work concentrated on a partial rewriting of the eprints.org (now ePrints)[19] software to meet the university's potential requirement for a service that went beyond a standard installation and customisation process. The resulting system was developed to be considerably autonomous with respect to the original installation, relying on completely re-engineered modules for, among other components, (i) the configuration layer (which is now far more flexible and fully XML-based), (ii) the Web control layer (which is based on a unifying notion of data views and offers templating and screen-flow management), and (iii) the data model layer at the foundation of the system. This offers significantly more modularity than the original eprints.org software for the search and deposit functionality identified for the pilot service, resulting in much smaller and manageable code with savings of up to 80%, and better support for the programming and non-programming skills that may be made available to a full-blown service.

At this point JISC expressed interest in developing the approach for application beyond HaIRST. The project staff agreed to allocate resources to this, as the unforeseen delays in resource identification for partner repositories had resulted in a considerable postponement of work on "thickening" the harvesting service. The development of the Strathprints pilot then branched into three conceptually distinct directions and thus into three corresponding sub-deliverables: an application framework, Strathprints as a sample application built on top of

the framework, and detailed documentation for both framework and application. The principles underlying the development of the Strathclyde pilot were successfully reported at the Eprints Round Table held in London in June 2004 to "consider the transition towards more community-led technical development and user support for Eprints software".

The set of system services necessary to support the application services targeted for the pilot took the form of a stand-alone application framework aimed at supporting the development of metadata-driven digital library services on the Web, and known as PLAF, the Perl Lightweight Application Framework. In particular, PLAF was designed to simplify development of application services which are (i) implemented in Perl, (ii) defined against metadata for information resources, and (iii) remotely invoked via the HTTP protocol. To do so, PLAF relies on the system services of (i) a mod-Perl enabled installation of the Apache HTTP server, (ii) the MySQL relational DBMS, and (iii) a distribution of the Linux operating system. Among PLAF services are an object-relational mapping for the metadata model, a simple resource publishing framework, a flexible programming paradigm for metadata processing based on the notion of *processing views*, a lightweight templating mechanism for the declarative design of the user interface and other service outputs, and session, response, screenflow, and error management to simplify interactions with remote clients through a mediating web server.

The Strathprints pilot became just *an* application built on top of it; core Strathprints services include a search service, a deposit service with associated model and policy, user and resource management services, and a dedicated OAI-based disclosure service. Both PLAF[20] and Strathprints services were individually documented to maximise the potential for their impact on the digital library community.

In the meantime, Strathclyde University continued to debate the nature and infrastructure of its institutional repository, and there was little progress on populating the Strathprints pilot. Work was undertaken to supply additional resource samples for the pilot from CDLR's staff publication databank, these meeting the definition of institutional resources.

The core work of HaIRST required the development of a cross-partner service for the testing of a strong research hypothesis concerning the feasibility of a flexible metadata and service architecture for the JISC Information Environment. The hypothesis was more clearly articulated during the initial stages of the project and successfully exposed to JISC and FAIR projects for comment[21].

A workshop and orientation event was held early on in the project to discuss issues of metadata and metadata agreements within HaIRST. The Project Manager gave a talk on the nature and scope of four relevant metadata initiatives (Dublin Core, IMS, OAMS, and MARC21), their relationships with each other, the OAI harvesting framework, and the metadata strategy planned for HaIRST[22]. This was followed by a brainstorming session among participants, who included members of the HaIRST Management Group as well as delegates from Glasgow City public library services, Strathclyde University's Computer and Information Sciences Department, and the Daedalus FAIR project.

After a first successful installation of the DLESE OAI Service Provider, the OAI-ARC software[23] became the preferred harvesting solution for the project. This is mostly due to the database-driven, out-of-the-box discovery service that ARC makes available. All of the unqualified Dublin Core metadata made locally available at partners is currently searchable via the ARC installation[24].

A poorly documented but valuable feature of the ARC system allows the exposure of the harvested metadata not only to the discovery service, which remains local to the data, but also to other remote harvesters. It is perhaps not widely known that ARC can serve as a

data provider in this way. The HaIRST installation has implemented this feature; specifically, the data harvested from the project partners can now be harvested in turn[25].

A presentation on the differences and commonalities between the sharing models adopted within the learning object and eprints communities as perceived through the HaIRST experience was delivered in March 2004 to the CETIS Metadata Special Interest Group[26].

The Scottish Z39.50 cross-searching service CAIRNS[27] was developed to accommodate servers which did not support the full range of CAIRNS indexes, which includes author keyword, title keyword, subject keyword, general keyword, ISBN, and ISSN. CAIRNS catalogues which do not offer a particular index are automatically removed from the selection list if that index is chosen by the user[28]. This development was essential for incorporating the HaIRST service as it is not likely to offer ISBN or ISSN indexes; locally-created resources do not usually have international standard numbers assigned to them.

The OCKHAM Harvest-to-Query (H2Q)[29] open source software for interoperating OAI repositories with Z39.50 was installed and successfully tested. HaIRST will be incorporated into CAIRNS when test data and temporary repository proxies are removed.

The "thin" metadata agreement was fully developed in the first year of the project in the form of compliance with the schema for Open Archives Initiative Dublin Core[30].

It became apparent after two years of the project that practical experimentation with harvesting richer metadata remained an unrealistic goal because of the unreadiness of project partners to "thicken" the metadata in institutional repositories, and the lack of available tools for deposit, disclosure, and harvesting.

While partners showed an interest in keeping abreast of those metadata-related developments which are relevant to their sphere of influence, for example the FE partners and the Learning Object Metadata (LOM) model, they were not able to justify that interest against the requirements of specific services, which institutions had typically not yet identified, nor against the resources they would allocate to support the specialisation and maintenance of freely-available tools, for example to extend eprints.org to map a custom data model onto harvesting requests. The cross-sectoral nature of the project partnership was a further obstacle to metadata refinement, for no obvious extension to Dublin Core was of simultaneous interest to all partners. Individually, each partner developed a pilot infrastructure for OAI disclosure which, according to the design principles of the harvesting model, was agnostic with respect to the particular metadata model chosen for disclosure. However, OAI-based communities built around non-DC metadata models were still limited, while generic OAI tools capable of effectively accommodating arbitrary metadata models, perhaps under the syntactic umbrella of the XML standard, were not available.

The focus in the final year of HaIRST therefore shifted to laying the basis for, rather than firmly establishing, cross-partner agreements on content semantics for the metadata which would be adhered to, at a minimum, at the point of disclosure. This work involved identifying guidelines and application profiles emerging from other OAI communities, extension of the HaIRST harvester and experimental discovery service to non-partner institutions, and discussion on ways of improving metadata content interoperability between partners.

The harvester was successfully extended to cover OAI repositories developed under JISC programmes at Edinburgh University and Glasgow University. Two permanent static repositories were created for the minutes of CILIPS business meetings and digitised photographs from the Springburn Virtual Museum, a service maintained by CDLR. An OAI-compliant repository was created for digitised documents in the Victorian Times[38]**Error!**

Bookmark not defined. service maintained by CDLR. The harvester was extended to include these three new repositories.

The project drew on experience gained in the HILT[4]**Error! Bookmark not defined.** and CAIRNS[3]**Error! Bookmark not defined.** projects to promote discussion between partners on the potential of standard authority files and other controlled terminology sets to improve interoperability at a richer level.

Partners agreed on the need for an agreed source of name and subject authority headings, with the Library of Congress authorities[31] for names (LCNA) and subject headings (LCSH) suggested as suitable sources.

An advocacy and advisory service for OAI-related developments and activities in Scotland was developed in the latter half of the project. The OAI Scotland Information Service (OAISIS)[32] is a website offering information derived from HaIRST, other projects in the FAIR programme, and related projects world-wide.

Project staff identified a number of useful online documents relating to OAI standards and services during their investigations. These include deliverables from related JISC projects in the FAIR programme, technical standards, English-language reports from OAI projects worldwide, and associated software and documentation tools. Reference to these documents were incorporated within OAISIS in two ways. Direct references and links are given for key documents and website homepages, in narrative text as well as lists. References and links to all documents were made available within the digital library of the SLAINTE service of the Chartered Institute of Library and Information Professionals in Scotland (CILIPS). This includes a catalogue[33] of MARC21 metadata records for online resources which support continuing professional development for library and information professionals working in Scotland. The catalogue supports URL-specified searches which were added to OAISIS as dynamic links for resources associated with specific keywords, subjects, and organisations. Project staff were given access by SLIC to the OCLC Connexion[11]**Error! Bookmark not defined.** service, to identify existing MARC21 or Dublin Core metadata in WorldCat or create it if necessary. The Connexion tools for automatic transformation of MARC21 to DC and vice-versa had already been investigated by HaIRST. Very little relevant metadata was found in WorldCat, so this activity resulted in the addition of some 60 new records. Copies of the WorldCat records were then downloaded to the SLAINTE catalogue. Any subsequent addition of new OAI-related resources by SLIC or CILIPS staff will be automatically available from OAISIS as a result of agreed metadata content standards, including LCSH and Dewey Decimal Classification.

HaIRST was closely involved with the Open Access Group of the Scottish Science Information Strategy Working Group of the Scottish Confederation of University and Research Libraries, subsequently renamed as the Open Access Team for Scotland (OATS)[34], from its beginning. OATS seeks to foster open access publishing in Scotland, and one of its first activities was to organise a meeting on 11 October 2004 to raise the profile of Open Access publishing amongst Scottish research funders and universities, and to launch a "Scottish Declaration of Open Access"[35]. CDLR staff associated with HaIRST are members of OATS, and were closely involved in the submission of a successful bid to JISC to develop a network of digital repositories of scholarly output of the Scottish universities; the new project is named IRIScotland[36].

Outputs and Results

Although HaIRST did not succeed in implementing operational repositories in all of the partner institutions, and as a result did not develop any substantive agreement between partners on approaches to metadata content to improve interoperability in aggregator services, it did set the groundwork and establish a testbed for future development in these areas.

It has shown that the OAI approach can be successfully applied to managing and accessing a much wider range of resources than the eprints it was originally developed for, matching the range of assets of concern to FE and HE institutions.

It has identified the static repository as a low-barrier approach to exposing metadata to OAI aggregator services.

It has demonstrated the application of static repositories to institutional administrative resources such as meeting minutes.

It has demonstrated the application of static repositories to inter-institutional digitisation projects.

It has identified and tested a number of open source tools that can be used in building repositories and aggregator services, and shown that they can successfully interoperate to produce effective services.

It has identified issues affecting the interoperability of metadata content derived from multiple sources, and explored the impact of transformation from one metadata format to another.

It has identified and tested tools for transforming and manipulating metadata in the Dublin Core and MARC21 formats in widespread use for information resources in FE and HE institutions.

It has raised awareness amongst senior FE and HE staff of the potential of OAI-based repositories for managing a range of institutional digital resources and the importance of metadata in improving the effectiveness of information retrieval from such repositories.

It has demonstrated the relevance of these technologies in meeting institutional requirements under the UK's Freedom of Information legislation.

It has shown that collaborative initiatives between FE and HE can be successful and fruitful.

In addition, HaIRST has created a number of tangible Web-based resources.

The project website[9]**Error! Bookmark not defined.** has links to partners' institutional repositories, an experimental harvested repository, tools including relevant standards and open source software, related projects, and the OAISIS service, as well as project documentation.

Institutional repositories for St Andrews University[16]**Error! Bookmark not defined.** and Strathclyde University[37] are operational.

A repository for materials digitised for the Victorian Times[38] project funded by NOF has been created using the Index+ proprietary software from Systems Simulation Ltd.

A static repository for exposing materials falling with the scope of the Freedom of Information (Scotland) Act, including minutes of meetings, has been created for the Chartered Institute of Library and Information Professionals in Scotland (CILIPS)[39].

A static repository for exposing materials digitised under the New Opportunities Fund (NOF) has been created for the Springburn Virtual Museum[40].

Temporary static repositories for institutional resources from Glasgow Colleges Group and John Wheatley College[41], and Napier University[42] have been created for test purposes.

A static repository gateway[43] has been created using open source software from the Los Alamos National Laboratory. This enables harvesting of the static repositories created by the project and other initiatives.

A pilot information retrieval service has been developed using open source software. The service harvests metadata from all of the repositories created as part of the HaIRST project. The service also harvests most other Scottish repositories including the Edinburgh Research Archive[44] of Edinburgh University and Glasgow University Eprints Service[45]. The service provides keyword searches of title, author and abstract across all harvested metadata, and for specific institutions.

The harvested repository has been made available to other OAI-PMH harvesters[25]**Error! Bookmark not defined.**.

The OAI Scotland Information Service (OAISIS)[32]**Error! Bookmark not defined.** has been implemented. This website provides advice and advocacy resources for developing services based on OAI and open access in Scotland. Much of the information is derived from the work of the HaIRST project and the FAIR programme. A section on metadata discusses interoperability issues and the use of application profiles.

Documentation about the Perl Light Application Framework (PLAF) has been created[20]**Error! Bookmark not defined.**.

60 MARC21 records for selected online documents related to OAI, institutional repositories, and open access have been created and deposited in OCLC's WorldCat union catalogue. The records are also available in the SLAINTE digital library of materials for continuing professional development, as well as OAISIS.

Collection-level descriptions for HaIRST repositories[46] have been added to the Scottish Collections Network (SCONE) and linked to OAISIS.

The HaIRST project has influenced the work of several other projects, including the COPAC/Clumps continuing technical interoperability (cc-interop)[47] project, the Scottish Portals for Education, Information and Research (SPEIR)[48] project, and the Managing Digital Assets in Tertiary Education (Mandate)[49] project.

Exit strategy

Test metadata created solely for project use will be removed from the harvested set, and test repositories will be deleted from the harvester's range.

Test static repositories will be de-registered from the static repository gateway, but will be available as XML files from the project website.

The harvester and experimental discovery service will remain operational, and will be utilised in the IRIScotland project.

A Z39.50 server for the harvested metadata set will be activated in CAIRNS.

Operational static repositories created for the project will continue to be maintained by CDLR. Relevant institutions will be advised on registering them with OAI directory services to encourage harvesting from other aggregators.

The project website will continue to be made available for at least three years.

The OAISIS service website will continue to be made available, and will be developed as the opportunity arises.

Outcomes

Aims:

a) Build access to a central repository of harvested institutional metadata which fully reflects locally defined data and processing requirements.

This was largely achieved, taking into account the uncertainty over local requirements. The pilot central repository is operational and easily accessible. Local repository implementers should aim to clearly define local requirements independently of the choice of metadata schema or service delivery platform. Aggregator services need to be aware of structure or content issues arising from the transformation of local records during the harvesting process, and their impact on the resulting cross-institutional repository.

b) Provide access to the repository through a Web-based search interface capable of supporting views of increasing granularity over the varied richness of the underlying metadata.

This was partially achieved. The Web search interface is technically capable of development to accommodate varying levels of metadata richness, but little development was undertaken in practice for reasons already stated. The methodology of establishing a baseline service using the simplest metadata format was essential for any progress in this area, as it helped identify issues in providing access between the simplest and richest levels. Aggregation services require agreed, clearly-defined community-wide policies on processing harvested metadata to provide views at levels of granularity higher than unqualified Dublin Core.

c) Ensure integration of the pilot service into the JISC IE by further disclosing the harvested metadata within and beyond the IE via a number of different routes, including other OAI harvesters and Z39.50-based remote discovery services.

This was achieved. The harvested metadata has been disclosed as an OAI repository itself. Other aggregators using OAI-PMH can harvest the metadata in turn. A Z39.50 connection to CAIRNS has been successfully tested, and details will be published when the connection is made permanent. Other aggregators using Z39.50 will be able to include the harvested metadata in a meta-search. Collection-level descriptions for the sets of harvested metadata have been made available in SCONE, and can be automatically output in a variety of standard formats including Dublin Core and JISC Information Environment Services Registry[50]. Collection-level aggregators can request descriptions in specific formats to avoid re-keying.

d) Provide a model infrastructure for the design, implementation, and deployment of institutional e-archives and associated disclosure services in Scotland.

This was achieved. The functional model has been successfully tested with a variety of open source software packages using data from further education, higher education, research and professional bodies. Operating components of the model, including standard and static repositories, a harvesting service, a discovery service, and a technical advisory service have been implemented on the Web and can be easily accessed by service developers and implementors.

e) Encourage the creation and deposit of institutional e-resources at institutions within the model infrastructure.

This was achieved. The pilot services for two universities are undergoing further development at the institutional level. Other project partners have a clearer understanding of the advantages of institutional repositories as well as the barriers to a successful implementation. The utilisation of static repositories as a low-cost solution for institutions with low volume and slow aggregation of specific types of institutional resource has been successfully demonstrated.

Key deliverables:

1) A proof-of-concept suite of layered of metadata agreements, defined as XML applications and containing at least a top, universal layer based on unqualified Dublin Core, a middle layer which reifies agreements across HaIRST partners, and a bottom, institutions-specific layer which supports the locally-defined data and processing requirements.

This was partially achieved. It was difficult to define a middle layer due to disparity in project partners' progress towards local, institutional agreements on requirements.

2) Development or re-structuring of institutional and inter-institutional e-archives (from relational databases to file-based HTML repositories) capable of serving metadata adhering to some of the layers of the stack of agreements defined in 1) through server-side OAI-PMH functionality.

This was achieved. Free or low-cost tools for automatic transformation of MARC21 records directly from the bottom layer to the top have been identified and successfully tested. A local tool for transforming metadata held in Microsoft Access to the top layer has been developed and successfully used to create a static repository. Developers of institutional repositories should be aware of such tools so they can accommodate the requirements for richer metadata at the local level and harvestable metadata at the global, extra-institutional level.

3) A pilot discovery service capable of regularly harvesting remote metadata from partner institutions through client-side OAI-PMH functionality, and of storing it into an XML-based backend (initially defined directly on a file-system for prototyping and testing purposes and later on a dedicated XML database system of choice).

This was achieved. The pilot service harvests metadata from the dynamic and static repositories created by partners during the project. The service is operational and has been extended to harvest metadata from other Scottish repositories. It is thus an embryonic aggregation service for electronic resources of all types created by Scottish institutions.

4) A local Web-based query interface to the service in 3) capable of converting user-defined structured and partially structured queries against the metadata back-end, quantifying the portion of harvested metadata corresponding to given query structures, and including novel structure-based discovery of collection-level metadata.

This was partially achieved, as outlined in b). The collection-level metadata schema used by CAIRNS has been developed to accommodate information about the degree of query-structuring supported by the metadata harvested from each repository. In principle, this information can be used to select a sub-set of the harvested metadata which supports the user-defined query. For example, if the user defines a subject query, the records for HaIRST repositories identify those which include subject elements in their metadata. This information can generate a filter to limit the search to metadata harvested from those repositories; the filter can be applied to the user's structured query. This benefits the user by shortening the time taken to search the aggregated repository, and confirming that failed searches are the results of the user's query.

5) A number of two-way metadata mappings to support further discovery and disclosure to and from a number of existing services and archives. At this stage the list includes other national and international OAI servers, the CAIRNS Z39.50-based remote discovery service, the RDN national subject-based access service, the CORC (now Connexion) shared cataloguing service run by OCLC, and collection-level metadata databases such as SCONE.

This was achieved. The harvested metadata is exposed to other OAI harvesters and can be exposed to CAIRNS. The project has identified the OCLC MARC21-DC mapping used in Connexion[11]**Error! Bookmark not defined.** and the Library of Congress MARC21-MARCXML and MARCXML-OAIDC mappings used in MarcEdit[14]**Error! Bookmark not defined.** as suitable for supporting transformations to and from the partners' dynamic and static repositories as well as the harvested metadata. SCONE has mappings[50]**Error! Bookmark not defined.** from its collection-level metadata schema to MARC21 and DC. The mappings are available at little or no cost, and allow institutions and aggregator services to interoperate between MARC21 metadata, often used in library management systems, and DC metadata used in many virtual learning environments and content management systems as well as institutional repositories..

6) Associated changes in institutional cultures, policies, strategies, and organisational structures, as appropriate to, and agreed by, participants.

This was achieved to the extent that partners were able to develop their own repositories.

7) A pilot for a regional institutional e-archives advisory service, together with an associated website offering advice and guidance.

This was achieved with the development of the OAISIS website to offer advice and guidance to individual institutions on setting-up and managing local repositories, and advice on issues affecting regional and national aggregation services. Although OAISIS focuses on OAI and open access initiatives in Scotland, it provides general advice and links to relevant documents and websites worldwide.

8) The involvement of Scottish FE and HE beyond the consortium institutions.

This was achieved through project participation in OATS[34]**Error! Bookmark not defined.**, and liaison with SLIC[10]**Error! Bookmark not defined.** and the Confederation of Scottish Mini-Cooperatives (CoSMiC)[51]. OATS is a sub-group of the Scottish Confederation of University and Research Libraries which includes all HE institutions in Scotland. SLIC has a strong representation of Scottish FE and HE institutions in its membership, and CoSMiC is an informal umbrella group for coordinating a common information environment in Scotland. HaIRST is given as an example in a toolkit[52] developed by SLIC for Scottish FE. The harvesting service has been extended to all operational OAI repositories in Scotland, currently all from HE institutions.

9) Draft institutional and inter-institutional collection development policy documents, covering all institutional e-resource activity, from institutionally created learning or research materials, through digitisation priorities, to the purchase of commercial research and learning materials.

This was partially achieved. Progress was severely hampered by delays in establishing basic institutional policies on locally-created resources. A number of external events having a significant potential impact on institutional policies occurred during HaIRST. This includes the Freedom of Information legislation, applying to public institutions like FE and HE and concerning access to institutional information, and the development of Creative Commons licenses under Scots law[53], which can be applied to intellectual property rights for resources

created by institutions. These have been monitored by the project and incorporated into the OAISIS service as appropriate.

10) An associated exploration of inter-institutional activity through the SCAMP gateway to facilitate collaborative collection development work in areas such as e-learning materials and digitisation programmes.

This was partially achieved. The SCAMP (Scottish Collections Access Management Portal)[54] gateway provides links to CAIRNS and SCONE, and to documents supporting collaborative collection development in Scotland. Several Wiki forums have been set-up to facilitate online discussion of metadata standards, digitisation programmes, and institutional resources, but have not yet been fully utilised. CoSMiC, SLIC and the National Library of Scotland intend to use SCAMP to build on previous work to develop a strategy for all Scottish collections[55] and incorporating recent developments including digitisation and open access repositories.

11) Report and recommendations on requirements in respect of changes in institutional culture, policy, strategy, and organisational structures, as well as on the communication protocols, metadata standards, and software implications of a service based on the various types of partner institutions.

This was partially achieved. Technical advisory information is included in the OAISIS service.

12) A model that can be expanded to encompass other Scottish FE and HE institutions and perhaps offer an approach applicable to other areas of the UK.

This was achieved. The HaIRST pilot services have been expanded to other Scottish institutions including one outside the education sector. The HaIRST approach uses open source software and open standards which are freely available to other areas of the UK.

13) Increased community understanding of issues through appropriate and sustained dissemination activities.

This was achieved. HaIRST was fully engaged in the e-FAIR cluster of the FAIR programme. Presentations about HaIRST were given to appropriate seminars throughout the life of the project, as detailed in the dissemination section of the project website. Reference to the HaIRST project and its work has been made in relevant standing committees of professional organisation such as SLIC, CILIPS, SCURL, the Cataloguing and Indexing Group in Scotland (CIGS) and others.

14) A full report on all project activities, together with recommendations covering further development guidelines on all areas of project activity.

This was achieved with this final report.

Conclusions

There is a range of open source software and open access documentation available to develop services based on OAI standards and protocols. The range caters for metadata repository management, metadata transformation, metadata aggregation through harvesting, and information retrieval.

The software overlaps in functionality, and the range is being rapidly expanded as a result of active research and development. The software can be difficult to implement, but components from different sources interoperate successfully

They can be used to create an aggregation service for the retrieval of a wide range of digital resources created by multiple FE and HE institutions.

They can also be used to create an aggregation service for the exposure of metadata records to other aggregators based on OAI or Z39.50 technologies.

OAI static repositories are a low-barrier method of exposing metadata for low-volume, low-volatility collections of resources to aggregator services.

It is difficult to improve retrievability of harvested metadata through agreed content standards unless consortial members have a clear understanding of their local requirements, which can be complex, for the local metadata.

Implications

OAI metadata repositories can be applied to a wide range of institutional information resources, not just research output. These include documents generated by administrative processes, so repositories can be used by institutions wishing to comply with the Freedom of Information legislation. If such documentation is digital, the repository can be used to provide unmediated search and access facilities to the public.

Institutions intending to exploit OAI repositories should ensure that local metadata requirements for library systems, content management systems and virtual learning environments are clearly understood before negotiating content standards with known harvesting services.

Groups of FE and HE institutions can successfully develop cross-institution information retrieval services, for example to allow sharing of resources or provide a consortial view of information assets, using the harvesting paradigm. This offers an alternative option to the established approaches of distributed search across multiple metadata sets and central search on submitted copies of metadata.

Cross-institution services benefit from agreeing on a standard metadata structure and guidance on creating content for metadata elements in the form of an application profile. Without this, the coherency and consistency of harvested metadata cannot be guaranteed, with a resulting deleterious impact on the efficiency and effectiveness of cross-institution information retrieval.

Groups of institutions can provide aggregated metadata repositories to other aggregation services, including those based on centralised OAI harvesting and distributed Z39.50 models. This can reduce the number of individual repositories selected and processed by higher-level aggregation services, with one harvesting process covering multiple institutions.

Institutions with small, stable collections of resources have a low-barrier method for widening and improving access in creating a simple XML file of metadata compliant with the OAI static repository specification and registering it with a static repository gateway. This may be useful where the metadata records already exist and are in a format that can be readily transformed to the unqualified Dublin Core required by the specification, but cannot be easily or widely used.

Static repositories can be created for each collection or set of collections and registered with different gateways, allowing the development of low-cost selective aggregation services based on collection characteristics such as subject topic or item format. This approach does not allow the inclusion of a collection in more than one aggregation service because a repository can only be registered with one gateway. There is no technical reason why clones of the repository cannot be associated with different gateways, although this would involve yet more duplication of the metadata.

Institutions can implement repositories and aggregation services using open source software, but should attempt to ensure technical support from staff aware of the pitfalls and problem-solving methods associated with this type of software. They also need to monitor worldwide development activity in this field.

Recommendations

Specific teaching, learning or research communities involving multiple institutions should consider using OAI technologies to share access to relevant resources created within member institutions.

The effectiveness and efficiency of such cross-institutional discovery services will be significantly improved if communities agree on an application profile for the structure and content of metadata stored in harvestable institutional repositories. The application profile should accommodate local institutional requirements as well as those of the community.

The application profile should be made freely available to inform other metadata aggregation services using OAI-PMH.

Communities should be aware that individual member institutions may wish to participate in aggregation services maintained by other communities, and that communities themselves may wish to participate in higher-level aggregation services. This needs to be considered when application profiles are being developed.

Smaller institutions with low volumes of resources should consider participating in aggregation services by using static repositories as a low-barrier, low-cost method for exposing metadata.

Institutions and communities using static repositories need to identify or implement a suitable static repository gateway to expose the metadata to harvesters.

Publicly-funded institutions should consider using OAI technologies to offer disintermediated access to administrative information within the scope of Freedom of Information legislation.

The project supports the recommendations to JISC given in the summative evaluation appended to this report.

References

1. Open Archives Initiative. Available at: http://www.openarchives.org/
2. Nicholson, D. and others (1999). Should universities manage services offering institutional or extra-institutional access to locally-created electronic teaching and learning resources?: final report of the CATRIONA II project. Available at: http://cdlr.strath.ac.uk/pubs/nicholsond/catriona2final.pdf
3. Co-operative Academic Information Retrieval Network for Scotland. Available at: http://cairns.lib.gla.ac.uk/
4. HILT. Available at: http://hilt.cdlr.strath.ac.uk/index.html
5. SCONE: Scottish Collections Network Extension. Available at: http://scone.strath.ac.uk/
6. Open Archives Initiative Protocol for Metadata Harvesting (2002). Available at: http://www.openarchives.org/OAI/openarchivesprotocol.html
7. Centre for Digital Library Research. Available at: http://cdlr.strath.ac.uk/
8. The Digital Information Office. Available at: http://dio.cdlr.strath.ac.uk/
9. HaIRST: Harvesting Institutional Resources in Scotland Testbed. Available at: http://hairst.cdlr.strath.ac.uk
10. Scottish Library and Information Council. Available at: http://www.slainte.org.uk/slic/index.htm
11. OCLC Connexion. Available at: http://connexion.oclc.org/
12. Digital Library for Earth System Education. DLESE OAI software. Available at: http://www.dlese.org/libdev/interop/oai/dlese_software/index.html
13. OAI Executive and others (2002). Specification for an OAI Static Repository and an OAI Static Repository Gateway. Protocol Version 2.0. Available at: http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm
14. Reese, T. (2004). MarcEdit. Available at: http://oregonstate.edu/%7ereeset/marcedit/html/
15. Research Library of the Los Alamos National Laboratory. Experimental OAI Static Repository Gateway. Available at: http://libtest.lanl.gov/
16. St Andrews Eprints. Available at: http://eprints.st-andrews.ac.uk/
17. Korycinski, C. (2004). HaIRST Project Report. Available at: http://eprints.st-andrews.ac.uk/archive/00000362/
18. University of Strathclyde Institutional Repository. Available at: http://eprints.cdlr.strath.ac.uk/
19. ePrints free software. Available at: http://www.eprints.org/software/
20. PLAF documentation is available from JISC.
21. Simeoni, F. (2003). Cluster coordinates: scope and assumptions in the eFair context. Presented at the eFair Cluster Meeting, 21 Mar 2003. PDF (zipped) version. Available at: http://hairst.cdlr.strath.ac.uk/documents/HaIRST-efair.zip - PPT (zipped) version. Available at: http://hairst.cdlr.strath.ac.uk/documents/ClusterCoordinatesPPT.zip
22. Simeoni, F. (2003). Metadata Issues: goals, principles, standards, and strategies. Presented at the Hairst Metadata Orientation Event, Glasgow, 12 May 2003. PDF (zipped) version. Available at: http://hairst.cdlr.strath.ac.uk/documents/HaIRST-efair.zip - PPT (zipped) version. Available at: http://hairst.cdlr.strath.ac.uk/documents/MetadataIssuesPPT.zip
23. Old Dominion University Digital Library Group (2002). OAI-ARC. Available at: http://oaiarc.sourceforge.net/
24. Experimental discovery service. Available at: http://speirserver.cdlr.strath.ac.uk:8088/arc/hairst_search.jsp
25. HaIRST harvestable repository. Available at: http://speirserver.cdlr.strath.ac.uk:8088/arc/servlet/OAI-DP
26. Simeoni, F. & Barton, J. (2004). Sharing models in the Information Environment. Presented at the meeting of the CETIS SIG Metadata, 8 Mar 2004. PDF (zipped)

version. Available at: http://hairst.cdlr.strath.ac.uk/documents/SharingModels.pdf - PPT (zipped) version. Available: http://hairst.cdlr.strath.ac.uk/documents/SharingModels.ppt

27. CAIRNS: Cooperative Information Retrieval Network for Scotland. Available at: http://cairns.lib.strath.ac.uk/

28. CAIRNS: Cooperative Information Retrieval Network for Scotland: Select catalogues. Available at: http://cairns.lib.strath.ac.uk/CAIRNSService/zcatsrch.cfm

29. OCKHAM downloads / services. Available at: http://www.ockham.org/services.php

30. Johnstone, P. (2002). XML Schema … adjusted for usage in the OAI-PMH. Available at: http://www.openarchives.org/OAI/2.0/oai_dc.xsd

31. Library of Congress Authorities. Available at: http://authorities.loc.gov/

32. OAI Scotland Information Service: Welcome to OAISIS. Available at: http://hairst.cdlr.strath.ac.uk/oaisis/index.htm

33. Slainte catalogue of online resources for LIS CPD. Available at: http://slic1.cdlr.strath.ac.uk/ipac20/ipac.jsp?/profile=slainte

34. OATS: Open Access Team for Scotland. Available at: http://scurl.ac.uk/wg/oats/

35. OATS: Open Access Team for Scotland:Declaration. Available at: http://scurl.ac.uk/wg/oats/declaration.htm

36. IRIScotland: Institutional Repository Infrastructure for Scotland. Available at: http://www.jisc.ac.uk/index.cfm?name=project_iriscotland

37. Strathprints: the Strathclyde University eprint repository. Available at: http://strathprints.cdlr.strath.ac.uk/

38. Victorian times. Available at: http://www.victoriantimes.org/

39. CILIPS' pilot static repository. Available at: http://hairst.cdlr.strath.ac.uk/repositories/CILIPSSRD.xml

40. Springburn Virtual Museum static repository. Available at: http://hairst.cdlr.strath.ac.uk/repositories/spring.xml

41. Glasgow Colleges Group/John Wheatley College's pilot static repository. Available at: http://hairst.cdlr.strath.ac.uk/repositories/GSRD.xml

42. Napier's pilot static repository. Available at: http://hairst.cdlr.strath.ac.uk/repositories/NSRD.xml

43. OAI Scotland's static repository gateway. Available at: http://oaiscotland.cdlr.strath.ac.uk/gateway/registry.html

44. Edinburgh Research Archive. Available at: http://www.era.lib.ed.ac.uk/

45. Glasgow ePrints Service. Available at: http://eprints.gla.ac.uk/

46. HaIRST collection-level description in SCONE. Available at: http://scone.strath.ac.uk/service/portal/pcolnshow.cfm?ucolnid=5965

47. COPAC/Clumps continuing technical cooperation project. Available at: http://ccinterop.cdlr.strath.ac.uk/

48. SPEIR: Scottish portals for education, information and research. Available at: http://speir.cdlr.strath.ac.uk/

49. Mandate. Available at: http://www.jwheatley.ac.uk/mandate/

50. Dunsire, G. (2004). Output formats for collection-level descriptions from the SCONE database. Available at: http://cdlr.strath.ac.uk/pubs/dunsireg/CCISCONEOutput.pdf

51. Cosmic: Confederation of Scottish Mini-Cooperatives. Available at: http://cosmic.cdlr.strath.ac.uk/

52. Resources & services supporting learning: a service development quality toolkit (2004). Available at: http://www.slainte.org.uk/files/pdf/fenet/toolkit03.pdf

53. Creative Commons worldwide. International commons: United Kingdom: Scotland. Available at: http://creativecommons.org/worldwide/scotland/

54. SCAMP: Scottish Collections Access Management Portal. Available at: http://scone.strath.ac.uk/scamp/index.html

55. Osborne, B. & Tayor, M. (1999). A National strategy for Scottish material. Available at: http://www.slainte.org.uk/files/pdf/slic/scotmatestrat.pdf

Appendix

Summative evaluation

Manchester Metropolitan University

CERLIM (the Centre for Research in Library and Information Management)

Summative Evaluation of the HaIRST Project

July 2005

The HaIRST Project Evaluation was a four-week project commissioned by the Centre for Digital Library Research (CDLR) at the University of Strathclyde and undertaken by the Centre for Research in Library and Information Management (CERLIM) at the Manchester Metropolitan University.

The HaIRST website is at http://hairst.cdlr.strath.ac.uk/

Further enquiries about this Evaluation should be addressed to
Professor Peter Brophy, Director of CERLIM p.brophy@mmu.ac.uk

This document should be cited as:

Brophy, P (2005) HaIRST Project Summative Evaluation: Report. Manchester: CERLIM (Centre for Research in Library & Information Management).

Introduction

This Report is the output from a short summative evaluation study of the JISC-funded HaIRST (Harvesting Institutional Resources in Scotland Testbed) project, part of the FAIR (Focus on Access to Institutional Resources) Programme. The project ran from August 2002 to July 2005, and involved a partnership of universities and further education colleges in Scotland.

The summative evaluation was charged with examining documentary and other evidence, and interviewing project participants and others, in order to draw conclusions about the achievements of the project and to find pointers for further development.

CERLIM (the Centre for Research in Library & Information Management) at the Manchester Metropolitan University is experienced in this kind of work as well as being an active participant in JISC Programmes for many years. It was from this informed viewpoint that the evaluation was carried out.

HaIRST

HaIRST has been a complex project which engaged with a heterogeneous academic community including three universities and a cluster of ten further education colleges, the latter all in the Glasgow area. As part of the FAIR Programme, it explored ways of enabling the deposit, disclosure and discovery of heterogeneous institutional resources. A wide range of materials, including eprints, electronic learning/teaching materials, administrative resources and collections of digitised visual objects were included within the range of the HaIRST consortium's interest and concern – indeed, anything to which institutions have some form of title and which may appear in, or be described by, a digital record.

Because HaIRST depended on inter-institutional collaboration it faced a number of critical cultural issues in addition to the challenge of developing the technical environment which enables such a broad range of institutions to participate, while adhering to the framework of JISC (and egov) standards. It thus faced many of the challenges which a UK-wide service would encounter, albeit on a smaller, and thus perhaps more manageable, scale.

The technical approach of HaIRST was based on the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH)[1], which is a technical protocol defining the process for gathering metadata from distributed sources, thus facilitating cross-institutional disclosure and querying. A layered approach to metadata was intended to enable institutions to engage in the project while retaining their own autonomy of description, providing minimum standards were agreed and adhered to. The discovery service would then be able to mirror the layered approach, providing different levels of sophistication to reflect the underlying data.

Summative evaluation: process

The evaluation was carried out by a mixture of desk research and interviews with key players in the project. The former activity involved the analysis of all available project documentation[2] as well as related materials from JISC and from the wider community (such

---

[1] http://www.openarchives.org/OAI/openarchivesprotocol.html
[2] Available at http://hairst.cdlr.strath.ac.uk/resources.htm. The documents include the project proposal, the project plan, the biannual progress reports to JISC, minutes of management group meetings and a number of PowerPoint and published presentations.

as web pages on OAI compliant software, on the broader information environment in Scotland, and so on).

The interviewees are listed in the Appendix at the end of this Report. We repeat here our thanks to them all for their openness and willingness to discuss the project and its achievements. We also repeat that none of the comments reported here should be taken to represent institutional policy.

Achievements of the project

Institutional engagement

It rapidly became apparent during our investigation that one of the most important achievements of HaIRST has been to convince key players of the value of repositories as a means of managing information assets and to confirm institutional intent to establish them. As one of the College interviewees put it, "HaIRST helped to formulate the idea".

As part of this achievement it seems that there has been considerable success in getting the message across to senior managers and policy makers that metadata matters. I other words, successful disclosure and successful retrieval depend on high quality description. At the operational level, however, there is as yet little agreement as to the standards and protocols to be applied – a huge area that will clearly need attention on a number of fronts (see section 0).

However, it is also clear that this is simply part of a long process. The interviewee quoted above also commented that "you have to keep justifying it (repository development) in terms of costs and benefits". Such benefits are seen as arising from the sharing of resources, of the costs of distribution and maintenance and of expertise.

There is a clear view that there will be a need to move on rapidly from the concept of the standalone repository to interoperability between it and other systems, including VLEs/MLEs, library catalogues, authentication systems and institutional portals. It is noticeable, however, that the repository is now seen as an essential component of the institutional landscape.

While these are all positive findings, we did also note that there remains considerable uncertainty amongst most interviewees as to the exact purpose of an institutional repository. We return to this issue later (section 0).

Cross-sectoral involvement

HaIRST was unusual in involving partners from both higher and further education. We noted that both sectors appreciated the opportunity to work with the other.

From the higher education perspective there was appreciation of the work being done in colleges on learning objects and more generally on the benefits of close interweaving of activities with community concerns – and of course recognition that further education is an important recruiting ground for higher education.

Most of the emphasis in the higher education institutions has been on eprint repositories, and library involvement is more or less taken for granted. With learning object repositories the position is much less clear; none of the institutions represented has progressed far in this regard and the role of the library remains very unclear. We could find no evidence that HaIRST itself had influenced learning and teaching within the institutions in any way. The

impression gained is that this is still very much territory to be fought, or negotiated, over. The experience of further education is seen as providing useful pointers.

From further education there was appreciation of the opportunity to work with higher education on a significant development project. A major benefit was seen as "building relationships with other institutions, especially with skilled individuals". This reflects the generally limited technical expertise (and time of those individuals in post) available in further education. A specific benefit has been the securing of the JISC-funded MANDATE (Managing Digital Assets in Tertiary Education)[3] project, which is producing "a toolkit including templates, a database structure and a training programme which will enable FE colleges to bring a coherent approach to the management and preservation of digital assets". It was noted that this activity has achieved much greater involvement of FE teaching staff than in HaIRST.

One further issue to emerge was that of the balance of costs and benefits. Although not yet apparent in the consortium, there was a perceived danger that specialist institutions in either sector may find that they are contributing much more than they receive from collaborative systems. Clearly this is something which will need to be managed.

Building a pilot HaIRST service

The team built a pilot service, using open source components, and capable of harvesting records from a wide range of targets and of providing a search service based on these records. It was thus able to provide one of the few demonstrators available of the kind of service which will be feasible once institutional repositories become well-established and populated. The service is built on the ARC harvester software from Old Dominion University Digital Library Research Group[4]. ARC is available for download[5] under the NCSA Open Source License. This achievement is to be applauded.

As part of this effort it was of course necessary to establish repositories in (or for) each of the partners. In the event this effort produced mixed results, although technically all partners did achieve it in one way or another. The best developed instance was that at St Andrews University; that at Strathclyde itself was no more than a test implementation (and that institution is continuing to consider the appropriate route to a fully service oriented repository). Both used the GNU *eprints.org* software version 2.0.

A particularly interesting approach was that taken by/with Napier University, John Wheatley College (JWC) and the Glasgow Colleges Group (GCG), which involved the use of 'static repositories'. We are not aware of other FAIR projects experimenting with this approach, and it deserves to be highlighted as one of the pieces of learning from the Programme. In essence a static repository is simply an xml file. Unlike the more familiar OAI-PMH repositories there is no capability of interactivity i.e. it cannot respond to OAI-PMH harvesting requests or other queries. A static repository gateway accesses the xml file using web protocols and then responds to OAI-PMH requests from other services on behalf of the static repositories which it is intermediating. One advantage is that the host institution has much less design and maintenance to perform since it is only uploading an xml file (i.e. much as any other web request using standard http protocols); another is that it is very easy to offer this approach on a hosted basis. It may be particularly appropriate to smaller colleges and to bodies relevant to higher/further education but outwith the sectors (such as

---

[3] http://www.jwheatley.ac.uk/mandate/
[4] http://dlib.cs.odu.edu/#arc
[5] at http://oaiarc.sourceforge.net/

museums, specialist libraries and so on). It needs to be considered alongside hosted services offering full OAI-PMH functionality[6].

The attempt by HaIRST to make *eprints.org* more user friendly to implement was a worthy one, and had active encouragement from the JISC because of its wider benefits. As a result a shell (called somewhat obscurely the PLAF - Portable Light Application Framework) was developed and is to be made widely available. However, it is apparent that the rescheduling of activity this implied may have distracted some attention from other core HaIRST goals.

It is apparent that the use of open source software absorbed a great deal of resource. Although this is not unexpected it points to the need for institutions, the JISC and its funded programmes to be realistic about the implications of the open source route. It is by no means self-evident that this is always the least resource intensive route, either for time limited projects or for operational systems.

Populating the repositories

There was mixed experience in obtaining suitable records and inputting these to the repositories, and the evaluation team spent some time exploring the reasons for this with project partners.

In the first instance it was clearly the case that HaIRST was seen as an experimental project within the institutions – one which was capable of helping develop understanding of issues surrounding institutional repositories but which by its nature had a limited lifespan. At St Andrews this appears to have been less of an issue than elsewhere – the idea of setting up an institutional repository was already accepted there and initial planning had taken place. Indeed a decision had even been made to go with the *eprints.org* software.[7] At other institutions, however, there appears to have been a marked reluctance on the part of some potential academic staff contributors to become involved in something which might not be there in the longer term.

The result was that, with active deposit by or for academic staff not yet taking place, most of the records contributed to the different repositories were placed there by project staff and were frequently chosen for test purposes (Napier for example provided records for much of the Catriona project material using existing MARC metadata records). This is not to criticise this strategy, which was appropriate for the project's primary purpose. However, it helps to point up some of the issues, as expressed to us in our interviews, which will need to be overcome if repositories are to achieve acceptance in their communities. These include:

- the issue of *permanence*; staff are unlikely to put effort into providing data for systems which are perceived as experimental and temporary;
- The *balance of responsibilities* between library and academic staff; to what extent should the inputting of records be mediated?
- Institutional (and possibly regional / national / sectoral) views on what is *appropriate content*. The view from St Andrews University, that the institutional repository should contain only high quality research outputs (peer reviewed or equivalent and perhaps not even including doctoral theses), lies at one end of the spectrum. The view of John Wheatley College, that a repository should

---

[6] We note that the IRIS project refers to the latter as "a collective hosting repository" (http://www.jisc.ac.uk/index.cfm?name=project_iriscotland).

[7] But it is interesting to note that since then the current repository has become a static resource and a decision is not yet made on a permanent service, which will be a replacement for the experimental one, and will probably be based on DSpace.

provide access to whatever it is that the local college and the local community wishes to publish in this way, lies at the other end. We refer to this issue again below (section 0)

- A related issue is whether a single repository should be used for all kinds of digital object; if not, *how should objects and repositories be characterised*. Again we refer to this issue below (section 0)
- Whether the right locus for a repository is the single institution. The argument for this is presented largely as one of ownership. However, the Glasgow Colleges Group is in fact a consortium and this may be an appropriate approach for colleges. Or there might be collaboration on a regional or national scale. (An obvious instance of this issue is posed by JISC's current strategy of encouraging institution-based repositories for eprints and a national repository for learning objects.)
- What kinds of logically separate repositories an institution might wish to establish. If we take the view that a repository is simply a method for the internal publishing of digital objects with (or as) associated descriptions then it is a quite separate issue as to the extent to which each repository should expose its content and to which internal/external services. Although some interviewees clearly see all digital objects being exposed to any service which wants to harvest them, others are of the view that this is an issue which the institution will wish to manage.

An intriguing idea which has emerged from HaIRST, and specifically from the FE members, is that the repository approach could be ideal as infrastructure for enabling institutions to meet their obligations under Freedom of Information legislation effectively and efficiently. If institutional documents (committee minutes and papers; internal reports; and so on) are placed in the repository and described with a DC metadata record, those records can be harvested by a variety of services and made widely available without requiring individual attention to every request (or they could of course be restricted to institutional harvesters if desired). So, as an example, by including an appropriate date range, the applicability of sets of regulations to a particular query could be established, again saving considerable administrative time. One cannot but wonder if such an approach might greatly help the JISC itself to make its papers and reports much more accessible.

In order to expand the amount of data available for searching, HaIRST has also harvested records from Glasgow University's eprint repository, from a static repository derived from one of the collections in the Glasgow Digital Library and from *Victorian Times*, one of the NOF-funded digitisation projects[8]. It was notable that the inclusion of the latter was seen by at least one interviewee as a good way of demonstrating the value of a HaIRST-type approach i.e. for internal advocacy.

Describing resources

The description of repository content is of course a critical issue and we explored with interviewees the question of resource description at some length[9]. It was notable that there was widespread commitment to the need for high quality metadata (and frequent reference

---

[8] http://www.victoriantimes.org/

[9] The issue emerges from most of the FAIR projects and has of course been addressed (or at least the questions have been put) in a number of fora. See, for example, the UKOLN paper in Ariadne 38 (http://www.ariadne.ac.uk/issue38/guy/). eprints UK and the RDN have published guidelines on the application of Dublin Core to repositories (see http://www.rdn.ac.uk/projects/eprints-uk/docs/simpledc-guidelines/) but it is unclear as to how widely implemented these have been.

to recent failings in this regard, for example in many of the NOF-Digi projects), so that the issue was one of how that was to be achieved – no-one in the HaIRST consortium that we spoke to was arguing for simply exposing raw content to search engines, nor for a simplistic approach to metadata creation.

OAI-PMH mandates unqualified Dublin Core, and that offers a starting point. One of the HaIRST concepts in the proposal and project plan was to used a 'layered' approach in which there would be a top layer of unqualified DC, a middle layer which represented a more detailed/standardised approach common to the consortium and a bottom layer supporting institution-specific metadata requirements. Although attempts were made to achieve this, in the event it proved over-optimistic largely because problems with achieving agreement on syntax and vocabulary control for the top layer were not resolvable in the time available (an oft-cited example being whether personal names should be normalised or not) and the consortium was not itself mature enough to have previously developed and tested their own approaches. A *de facto* acceptance of MARC21 owed much to convenience but does not offer a long-term solution to non-'library' resource description.

The idea of a layered approach retains validity, however, and it is one of our recommendations that it should be explored further once there is sufficient content in UK repositories to conduct meaningful trials.

In respect of subject description there is recognition of the problem but no evidence that as yet solutions are in sight. Library staff are of course familiar with the major classification schemes and with Library of Congress (and other) subject headings, but these were seen as inappropriate for repositories where, at least in higher education, there would need to be a strong element of self-description by academics (though some felt a measure of mediation for quality control purposes would always be required). It was recognised that one of the severe weaknesses of the <u>institutional</u> approach to eprint archives is that they co-locate a very wide range of disciplines which have entirely different (and sometimes conflicting) vocabularies – from this perspective subject-based repositories make a great deal more sense. There was some suggestion that institutions need to become much more familiar with vocabularies used by disciplines and to build solutions based on these. Once harvesting services start to mature inconsistencies between subject descriptions used by different institutions could otherwise prove fatal.

Searching HaIRST

As noted above, HaIRST has implemented the ARC harvester and has built a simple interface to enable search on keywords from the author, title and abstract fields (see illustration below). This appears to be effective, although inconsistencies in the underlying metadata and the limited number of records available limit its usefulness beyond the project.
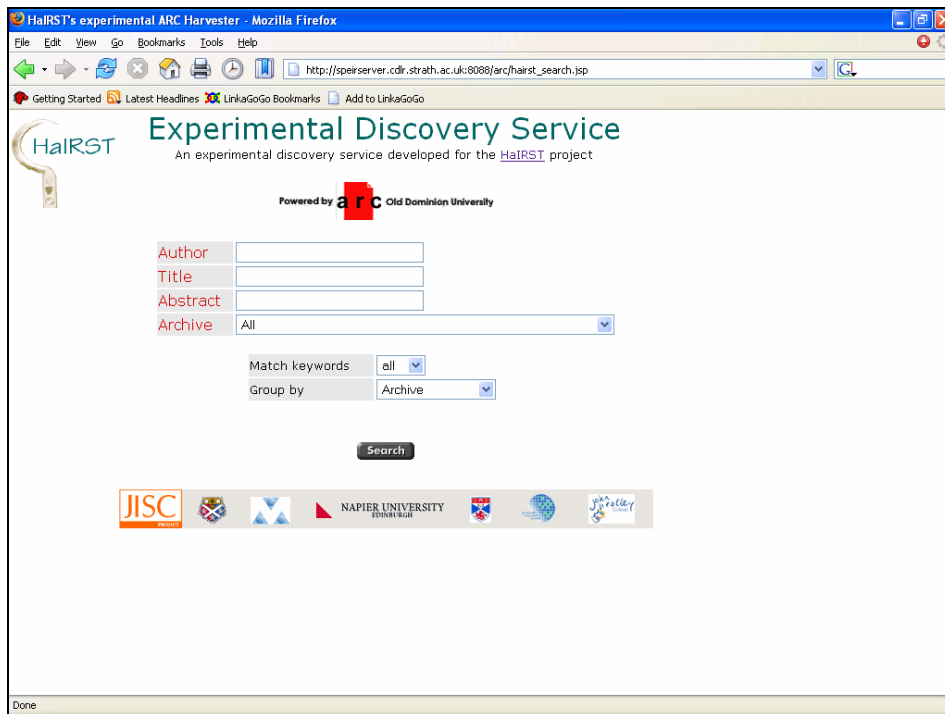
None of the project participants claimed to be using the search service in its current guise for other than project purposes.

We undertook a number of trial searches with this service and it appears to give satisfactory results given the limitations of the underlying metadata. It does not pretend, however, to be anything other than an experimental service.

Disclosure

It was the intention that the records harvested by HaIRST should themselves be exposed to other services. HaIRST itself is available through the CAIRNS (Cooperative Information Retrieval Network for Scotland) service (via its collection description). The HaIRST records can be harvested by any other OAI compliant harvester, although we are not aware that this has happened – nor would there currently be reason to do so.

One of the issues raised in discussion was the proliferation of instances of records as they are harvested by multiple services which will not themselves have the capability of providing uniform access. Although this issue takes us beyond the scope of this report, it is a matter of considerable concern and needs further investigation.

Aggregated services

We asked interviewees for their views on the types of aggregated service which would be most useful and appropriate, but found considerable uncertainty as to the correct foci. There is an obvious focus on the institution itself and this will be aggregated internally through a suitable portal / managed learning (or research) environment.

For some observers with a primary concern for research, the only other focus is international within a defined subject field – the researcher at the leading edge of her/his field would then only tighten the focus if she/he knew of a specific source in advance – say a world-renowned research centre at another university. Other than this there seems no more reason to explore, say, a Scottish aggregation than a South American one, except where access to

physical objects is involved. As one interviewee put it, "there is no point in regional aggregators for digital objects".

However, for others there is great value in a regional/national (for example Scottish) aggregator – this could be because they are working in that field or because the cultural and legal framework is of great importance – one interviewee suggested that "the Scottish framework is more suited to non-academic research". To explore this further, we asked interviewees the extent to which they, or their end-users, make use of the CAIRNS[10] service either to identify other collections or for cross-searching. It transpired that they were not aware of significant usage ("I haven't heard the academics I work with using it a lot"), although they felt it was probably useful to library staff when answering queries at the reference desk. This suggests that the purpose and intended audience for future regional aggregators will need to be determined carefully.

OAISIS

One of the HaIRST deliverables is the OAI Scotland Information Service (OAISIS – available at http://hairst.cdlr.strath.ac.uk/oaisis/). The evidence we have gathered from project participants suggests that this will be a useful development in the future and one which they themselves are most likely to use as a source for staff in their own institutions who are becoming involved in setting and using repositories for different purposes. It was further suggested that it may be useful to help project staff focus on user needs. None of our interviewees, however, had used the service themselves to a significant extent.

An issue will be the extent to which OAISIS duplicates more general open access resources on the web. It needs to achieve a blend of Scottish and international resource and to build a reputation for authoritative advice – based, no doubt, on CDLR's own reputation. In doing this, it needs to be clear exactly who the intended audience is.

OATS

The Open Access Team for Scotland (OATS)[11], which draws together SCURL, SLIC, the National Library of Scotland and CDLR, provides a forum for the discussion of open access Issues. Few interviewees had any real awareness of its work or significance. We also noted that its website appears to be rather out of date (and the 'background' section is empty). We would judge that its direct impact has been limited to those directly involved in its deliberations, but of course it forms part of the Scottish infrastructure for moving open access forward.

The Scottish Open Access Declaration

The Scottish Open Access Declaration[12] was launched at an event on 11th October 2004 attended by senior representatives from Scottish universities, the funding bodies, the Scottish Executive and other bodies. We were asked to explore with interviewees what the impact of this Statement had been to date.

The views expressed varied, as might be expected, depending on the responsibilities and concerns of the individuals we spoke to. Those at a senior level characterised it, at least in one instance, as "absolutely critical" and demonstrated its impact both with senior

---

[10] Co-operative Academic Retrieval Network for Scotland: http://cairns.lib.strath.ac.uk/
[11] http://scurl.ac.uk/WG/OATS/index.html
[12] http://scurl.ac.uk/WG/OATS/declaration.htm

institutional staff and within Scottish government. Others were more cautious but all agreed that it had been a step on the road to open access, and had value in raising the importance of the issue in the minds of senior policy makers. What is clear is that it has provided a building block for further advances.

Staffing issues

It is not unusual for the summative evaluation of a JISC-funded project to find that staffing issues have had a significant effect. In HaIRST's case the project was fortunate in being able to appoint excellent staff. It was unlucky, however, in that the Project Manager, Fabio Simeoni, left the project at a critical stage. It was also restricted by the fact that the funding for staff in partner institutions did not extend beyond the phase of setting up the repositories. It is also apparent that the course of the project was influenced by the skills and interests of the staff appointed, with most emphasis given to the technical solutions rather than the problems of embedding repositories in institutional cultures and structures. JISC's encouragement to staff to work on technical solutions to the implementation of the *eprints.org* software inevitably reduced effort available for other issues to be pursued.

Dissemination

Although there have been a number of presentations on HaIRST at a variety of meetings, and papers have been published, the project has not enjoyed the high profile achieved by some other FAIR projects. In part this has no doubt been caused by the staffing difficulties referred to earlier. In part it derives from the nature of the consortium and the internal focus of much activity by partners, apart from CDLR.

We also noted that some of the participants had experienced difficulty with the terminology and technical jargon used in the open archives community and suggested that widespread support for these initiatives would be hard to achieve until the concepts were translated into plain English. Even the term 'repository' is off-putting[13].

Exit strategy

We understand that, in formal terms, an exit strategy will be addressed in the project's final report. Here it is simply worth noting that there was never an intention that the HaIRST project itself should turn into a service in its own right, and that many of the issues it has highlighted will be investigated further in IRIS.

Conclusions and Recommendations

Types of repository

We have noted that there is a need to think through what the repository landscape in Scotland may look like in a few years time. While all repositories (at least in the context in which we are working) are simply collections of digital objects, they are usually distinguished

---

[13] One may note in passing that one of the Oxford English Dictionary's definitions of *repository* is 'a place in which a dead body is deposited; a vault or sepulchre'. So Thomas Pennant records in *A tour in Scotland and voyage to the Hebrides 1772:* "In the middle of these repositories was placed the urn filled with the ashes of the dead". We leave the interpretation of this observation to our readers!

by the types of material represented and by purpose. Some interviewees saw it as important to maintain these distinctions; others did not. Typically, the term 'repository' includes:

- repositories of eprints, whether the aim was to hold a copy of the full text (or whatever equivalent) with appropriate metadata or simply metadata[14]. If eprint repositories contain only descriptions, there is a question as to whether it should be a requirement that the metadata contains a pointer to an openly accessible copy of the resource.
- repositories of other research-related material, such as e-theses and pre-prints.
- learning object repositories, where the key question appears to be the relationship between the virtual learning environment and the repository. Incidentally we discerned no obvious support for national learning object repositories on the JORUM model.
- administrative data repositories, linked to concerns as to how to achieve compliance with FoI legislation efficiently and effectively;
- more general repositories, for example for student or members of the local community to contribute to.

We understand that these kinds of question will be explored in the new IRIS project. We **recommend** that specific future scenarios should be explored in depth with the institutions. As part of this consideration, it is important that the contributions and roles of the different players be established. So far as content is concerned, it appeared to us that the research-led universities, teaching-led universities and further education colleges had very different ideas of what repositories should be used for and the extent to which institutional assets should be exposed.

Focusing the product/service

In terms of research support, clearly the point of eprint repositories is to support the total research and other institutional business by making results available freely and economically. The question was posed during our study as to whether there was a need for anything other than three foci:

- the research team, based in its subject domain and with contacts in other related departments/centres worldwide. Such a team may wish to use a repository as a way of exposing a coherent body of research results. It will be heavily influenced by the dominant research publication paradigms in its discipline.
- the institution, which has ownership of its assets and wishes to use those assets in its economic activity, to promote its excellence and as a contribution to society.
- worldwide research information services, capable of harvesting (or otherwise acquiring) information of significant research and other activity, processing it and delivering it selectively.

In other words, why focus on Scotland? (or, more generally in the UK context, on a region?)

---

[14] It is here that confusion of purpose creeps in. For some, the research repository is a means of exposing high quality research outputs for use by the research community. For others the driving purpose is the RAE, where it might be necessary to include some outputs – perhaps from younger researchers – which did not quite meet a more strict criterion, especially if the strategy of the institution was to maximise the number of researchers submitted. Equally if an institution was focussing on the RAE it might want to exclude some material which it would otherwise be happy to expose – say an output intended as a teaching resource.

It seems a little presumptuous for an English-based team to try to answer the Scottish question (though it was posed by Scots!). However we did note that there are several possible answers, and we would **recommend** that these, and doubtless other related issues which were not raised with us, should be explored in order to reach a agreed rationale for a <u>Scottish</u> service. The relevant questions raised during this study are:

- Firstly, the question of Scottish culture and identity as a focus in itself for research and research-related (which includes teaching) activities. The question is, how can institutions best contribute their assets to underpin such a focus?
- Secondly, to what extent do the key players see a continuum across Scottish society from the leading-edge research teams working in some of the UK's leading universities to the small groups of mature learners creating their own resources in underprivileged areas of the Scottish cities or islands? Since repositories allow all such assets to be exposed, what kinds of Scottish-focused services, based on the whole spectrum of activity, are feasible and desirable?
- Thirdly, given the experience garnered through a range of projects (HaIRST, Daedalus, and now the recently-launched IRIS) what value is placed on having 'local' expertise to call on for technical and operational advice as opposed to access to 'global' sources?
- Fourthly, given the emerging concept of 'pooled research', to what extent does a Scotland focus for repository-based activity enable underpinning services to be delivered which will enhance such research effort?

Towards aggregated services

Perhaps the most important contribution of HaIRST has been the exploration of what it means to utilise data from distributed repositories in order to provide aggregated services which are meaningful to the end user. It is one of the few projects which has gone beyond the building of repositories to develop understanding of the technical, resource description, organisational and cultural issues which need to be resolved as we attempt to build meaningful services. That this has raised as many questions as it has answered comes as no surprise. It is important that work in this area is pursued with vigour and determination, and we **recommend** that it should include:

- testing of the benefits of static repositories as a solution for smaller institutions / collections.
- further exploration of the concept of building services using a 'layered' approach to metadata (see below);
- structured investigations of the quality of metadata (in terms of standards compliance, semantics, vocabulary control etc.) in UK eprint repositories.
- identification and/or development of vocabularies suitable for use by specific subject communities.
- issues relating to interoperability between institutional services i.e. the place of the repository in the broader institutional technical and operational environments.

Metadata creation

There has been little consideration in HaIRST of the locus of responsibility for metadata creation[15] although more than one interviewee commented that library staff would not be able to act as mediators for this in the long term i.e. self-description would need to be

---

[15] It may be noted that this issue has been discussed at some length in other FAIR projects.

developed in some form. We **recommend** that the different models be explored further, building on the experience of other FAIR projects.

Scalability

Although there is no reason to question the scalability of the OAI-PMH approach in itself, we **recommend** that the issue of proliferation of instances of records should be explored further. In particular, it needs to be established that services which harvest records from multiple sources (some of which are themselves made up of harvested records) are capable of identifying duplication.

Preservation

While preservation of digital objects was not a focus of HaIRST, we noted in a number of discussions with project partners that there was an expectation that some types of institutional objects would be preserved for the long-term in the kinds of repositories being built. We therefore **recommend** that this issue should be given further consideration, both in technical terms and organisationally, perhaps through the JISC Digital Repositories and Digital Preservation and Asset Management Programmes, but bearing in mind any issues specific to Scotland.

* *

Appendix

We are grateful to the following for agreeing to be interviewed for this Study.

Janet Aucock	University of St Andrews
Sara Brown	Napier University
Rachel Bruce	JISC
Mark Clark	Glasgow Colleges Group
Lynn Corrigan	Napier University
Craig Green	John Wheatley College
Andrew Jackson	Glasgow Metropolitan College
Derek Law	University of Strathclyde
Diane Lindsay	University of Strathclyde
Tom Wilson	Glasgow Metropolitan College