

# What national repository infrastructure can do for global harvesting

## 1. Introduction

The IRIScotland project ('Institutional Repository Infrastructure for Scotland') has four main strands:

- i) The creation of a pilot hosting repository for individuals and institutions who do not have a repository or access to one, with the intention that in future the whole range of published research output of Scotland might be made visible worldwide via Open Access;
- ii) The creation of an aggregation of eprints metadata plus a search and browse service for the Scottish geographical area;
- iii) A survey of the research community and university administrators in Scotland on attitudes to Open Access;
- iv) To effect a degree of cultural change in that community by the promotion of a common metadata standard across the country, and the encouragement of a culture of repository deposit.

## 3. National survey

The principle aim of the IRIScotland survey was to collaborate with university senior managers, researchers and library staff to investigate the current cultural attitudes present within academic institutions in Scotland. The outputs of the survey have helped to shape the project requirements.

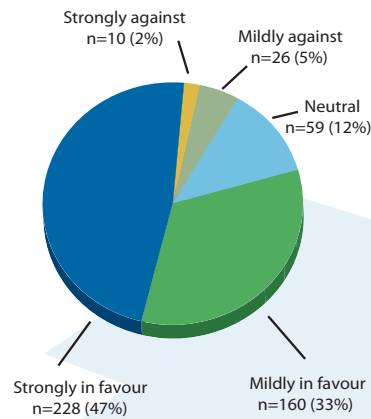


Fig.2. How Scottish academics in the IRIScotland survey feel about Open Access.

## 3. National survey (cont.)

### Survey Snippets

When asked about the principles of open access 87% of all respondents indicated that they were either strongly in favour or mildly in favour of the principles of open access. More than half of respondents were strongly in favour of the principles of open access (Fig.2.). Respondents were then asked if their institution currently had a repository. 25% said yes, but 56% were unsure. This highlights the need for even more advocacy and publicity for repositories within universities, especially given that such a high percentage of respondents are in favour of open access as a principle.

### Survey Conclusions

The response to the survey is a clear indication that there is strong support for open access amongst academics at Scottish universities. However, it is clear that there is a lot of confusion amongst academics relating to issues such as how repositories relate to traditional publication mechanisms, and what rights copyright agreements permit authors to retain. In addition, much more needs to be done to publicise the existence of institutional repositories and to encourage authors to deposit in them.

## 5. Hosted Repository Service Development

After an evaluation period we decided to use the Fedora platform. Fedora provides an application level substrate for storing, organising, disseminating and preserving digital objects, but incorporates very little end-user functionality. Thus we needed to develop the following functionality:

### Configuration management

The system is configured through a web interface which allows the user to configure the Fedora repository that will be used and also the email server for system generated email. The system is able to check the configuration while the system is used. When a configuration error is detected, the user is directed to log-in as an administrator to address the issue. Configuration

can be modified at any time.

### Account request work-flow

Repository account requests are mediated by the system. The system administrator can approve/reject requests. Email notifications are sent to the requester.

### Collection management (create, delete, etc)

A hierarchy of collections can be created by an Institutional Repository administrator. Items are deposited into specific collections.

### User account registration

A visitor must register to deposit. Once a user has an account she may then request permission to specific collections.

### Deposit permission request workflow

Once a user has an account she may then request permission to deposit in a specific collection. A repository administrator will then approve/rejects the request. Email notifications are generated at every stage of the workflow.

### Item deposit workflow

Deposits can be saved in a user's workspace to be finished in another session. An item consists of one or more files with associated metadata. Metadata is entered via a web form. Once a user is finished the item can be submitted for deposit. The item then enters the repository administrator's task queue where a repository administrator can check/modify the item before publishing it in the repository. The administrator

## 2. The model

Figure One shows how the IRIScotland tiered service architecture will interact with other OAI-service providers in the UK, and possibly further afield.

- a. By talking directly to repository administrators the IRIScotland project can ensure quality metadata records are created at source by providing recommended standards and cataloguing guidelines.
- b. Metadata from existing repositories in Scotland will be harvested. Smaller institutions without the resources to develop a repository will be offered the chance to use a centrally hosted repository based at the National Library of Scotland.
- c. In turn the entire superset of metadata from Scotland will be made available to the proposed UK national OAI-service to be run by Intute.
- d. These quality-assured records can then be reused and repurposed by other OAI-service providers; for example, the European DRIVER project or potentially OAIster.

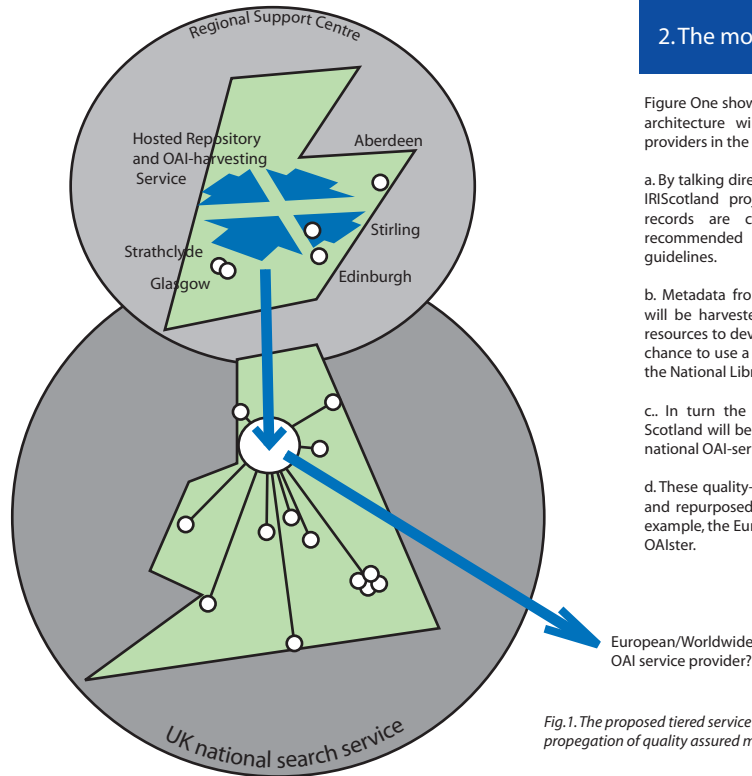


Fig.1. The proposed tiered service architecture for propagation of quality assured metadata records

## 4. Cross repository search/browse service

IRIScotland has developed an OAI harvester-based pilot cross-repository search and browse service to enhance the exposure of Scottish research output as a whole.

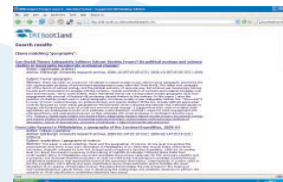


Fig.3. Screenshot of the pilot search service.

### Harvesting

A test database was created during March 2006 using the Arc harvester. Currently records are harvested from five Scottish institutions.

### Parsing

- A new parser was developed, using Visual Basic, to allow full control over the parsing process, e.g. in the handling of symbols and

non-standard characters, the processing of dates, and the handling of qualified Dublin Core records.

- The Identifier field was parsed to extract URLs and store them in a separate field. In due course this should ensure that all records link to a valid URL.

- Compound fields (Creators, Contributors and Subjects) were parsed into separate tables, thereby making it more meaningful (in theory) to browse them.

### Metadata management and research

- Metadata held in title, author, and description fields were reviewed to recorded numerous non-standard characters encountered (e.g. long dashes, smart quotes, bullets, fractions, accented characters etc).

- Procedures for metadata cleaning were successfully tested, i.e. in modifying non-standard characters, using a character translation table held in an Access database.

- Liaison with repository managers at Edinburgh, Glasgow and Strathclyde universities to ensure consistent content of subject and identifier fields etc

may return the item to the user.

### Institutional Repository Access Site

Branding can be associated with each institution. This currently consists of a CSS stylesheet and a logo image.

### OAI-PMH Static Repository Dissemination

Each institution can disseminate itself as an OAI-PMH Static Repository (SR) which provides both oai\_dc and full dublin core metadata. Dublin core metadata is generated as DC XML. Each collection in a repository is able to disseminate itself as a static repository allowing repository administrators to provide several scoped SRs.