Search

Collection landscaping and institutional repositories in Scotland

Gordon Dunsire discusses the use of collection-level descriptions in SCONE for the IRIScotland local repositories and union catalogue

Introduction | SCONE and IRIScotland | Scaling | General retrieval | Interoperability between SCONE and IRIScotland | Context

Introduction

The Scottish Collections Network (SCONE) service is a component of the Scottish Common Information Environment. It is used to find and display collection-level descriptions (that is, metadata about collections as a whole) of archive, library and museum collections in Scotland.

An important function of SCONE is to "landscape" collections when searching for items by location, subject, or educational level. Landscaping involves identifying collections and their online finding-aids (including catalogues) which are likely to contain items of interest, before carrying out more specific searches using local retrieval interfaces or relevant "union" catalogues. Further information can be found in Collection landscaping in the common information environment: a case study using the Scottish Collections Network (SCONE) : report for Work Package B of the JISC CC-interop project (PDF file).

The report identifies the use of collection-level description to improve access to union catalogues, or aggregations of metadata from multiple databases. The Co-operative Information Retrieval Network for Scotland (CAIRNS) is one such union catalogue, where the metadata from some 50 Scottish library services can be searched in one go, as if it was a single gigantic online catalogue. In fact, the aggregation of metadata is virtual, and CAIRNS does not actually exist as a single physical database. The CURL Opac (COPAC), on the other hand, is an example of a physical union catalogue, where the metadata is copied from contributing libraries into a central database. Several Scottish university libraries contribute to COPAC, but all are also included in CAIRNS. The result is three separate places where the same item might be searched: the local university online catalogue, COPAC, and CAIRNS. The report showed that a service like SCONE could present such duplicate sources of metadata in a coherent way that reduced potential confusion for users. The SCONE approach also uses the union catalogue as a back-up for the local catalogue if it should become unavailable for any length of time, for example during system upgrades.

The report also noted that the development of institutional repositories and metadata aggregation using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) would aggravate the situation by allowing unfettered duplication of metadata into "harvested"

physical union catalogues. An institutional repository is usually a databank of electronic publications created by an organisation and its members, along with metadata describing the publications. OAI-PMH allows any third-party to copy the metadata automatically, without the host repository necessarily knowing. One reason for such third-party aggregation is to develop services for one-stop searching across multiple repositories, rather than users having to search each local repository in turn. In fact, some institutional repositories may be based on "static repositories" which do not have local retrieval interfaces; see Harvesting institutional resources in Scotland testbed project : final report (PDF file) and Stargate final report (PDF file) for more information about static repositories in Scotland.

SCONE and IRIScotland

The Institutional Repository Infrastructure for Scotland (IRIScotland) project has allowed SCONE to explore these issues further. The Centre for Digital Library Research (CDLR) is leading a work package to develop a pilot harvested union catalogue and retrieval interface for the research output of Scottish institutions; further information can be found on the IRIScotland website and the IRIScotland wiki.

In particular, SCONE has created collection-level descriptions for the IRIScotland local repositories and union catalogue. The features of SCONE allowing this to be effective include:

- Catalogues and other aggregations of metadata are treated as collections (of metadata) in their own right.
- "Functional granularity" allows a collection of resources to be inferred if there is a corresponding collection of metadata.
- A collection of resources is related to corresponding collections of metadata (i.e. collection-level descriptions of type "catalogue") using an "Is-Described-By" relationship.
- Collections can be related to each other in a hierarchical way, to accommodate super- and sub-collections, using "Contains" or "Is-Part-Of" relationships.
- SCONE does not apply hierarchical relationships to collections of type "catalogue", only to the corresponding collections of resources.

For further information about the underpinnings of SCONE, see An Analytical model of collections and their catalogues.

The methodology for applying these features to the harvested metadata for IRIScotland is:

1. Add a record to SCONE for the "IRIScotland catalogue", the collection of all harvested metadata from IRIScotland partners' research repositories.
2. Add a corresponding record for the "IRIScotland collection".
3. Add the "IRIScotland collection" Is-Described-By "IRIScotland catalogue" relationship.
4. Add similar records and relationships for each of the harvested repositories.
5. If the harvested repository itself organises its metadata into sub-sets, add records and relationships for each sub-set and add an [Institutional repository] Contains [Institutional repository sub-set] relationship.
6. Add an "IRIScotland collection" Contains [Institutional repository collection] relationship.

This results in the following collection hierarchy:

- IRIScotland collection

- Edinburgh research archive
- Glasgow eprints service
- St Andrews eprints
- Strathprints
- University of Stirling digital research repository
  - Dept. of Applied Social Science eTheses
  - Dept. of Computing Science and Mathematics eTheses
  - Dept. of English Studies eTheses
  - Dept. of Environmental Science eTheses
  - Dept. of Film and Media eTheses
  - Dept. of History eTheses
  - Dept. of Management and Organization eTheses
  - Dept. of Philosophy eTheses
  - Institute of Aquaculture eTheses
  - Institute of Education eTheses
  - School of Biological and Environmental Sciences eTheses

At each entry in the hierarchy, there is a link to the online information retrieval interface for the collection's metadata at the same level of granularity as the collection. For example, the electronic theses of the Institute of Aquaculture have a specific retrieval service, but they can also be retrieved along with other University of Stirling theses by using the university-level interface, and they can be retrieved along with the research publications of Scotland as a whole via the pilot IRIScotland service being developed.

Scaling

This approach can be scaled to any level and number of sub-collections.

A possible scenario is that the University of Stirling repository expands to include research papers by members of staff, and these are to be grouped by department alongside the eTheses. The resulting SCONE hierarchy would insert a departmental digital repository collection above the existing eTheses collection, and add a research papers collection at the same level:

- University of Stirling digital research repository
  - Dept. of Applied Social Science research repository
    - Dept. of Applied Social Science eTheses
    - Dept. of Applied Social Science research papers
    - Etc.
  - Etc.

General retrieval

The SCONE service allows retrieval of collection-level descriptions by several methods. Those that are relevant to digital collections and IRIScotland include:

Collection title: Search for words in the title

This option allows the user to input a single text string, including embedded spaces to allow a phrase. SCONE lists the titles of collections containing the string anywhere in the title.

For example: "prints" retrieves "Glasgow eprints service", "St Andrews eprints", "Strathprints".

- **Try it out**: Search SCONE for collection titles containing "prints" (using a URL parameter as described below).
- **Try it out**: Search for words in collection titles (using the option from the full SCONE service).

People and organizations: List names A-Z

This option allows the user to browse a list which can be limited to organization names and/or specific relationships. SCONE displays the titles of all collections related to a selected name. Relationships include owners and subjects; both are created for IRIScotland institution names.

For example: "University of Edinburgh" retrieves "Edinburgh research archive".

- **Try it out**: Browse names associated with collections (using the option from the full SCONE service).

People and organizations: Search for words in the name

This option allows the user to input a single text string, including embedded spaces to allow a phrase. SCONE lists names containing the string anywhere in the name, and allows the user to browse the list which can be limited to organization names and/or specific relationships. SCONE displays the titles of all collections related to a selected name.

For example: "Stirling" retrieves "University of Stirling digital research repository".

- **Try it out**: Search for words in names associated with collections (using the option from the full SCONE service).

Educational level

This option displays a list of standard education levels, including "Masters - SVQ 5" and "Doctorates". When a user chooses a level, SCONE displays the titles of all collections associated with it.

For example: "Masters - SVQ 5" retrieves "University of Stirling digital research repository", etc.

- **Try it out**: Education levels of collections (using the option from the full SCONE service).

Subject retrieval

Subject retrieval methods in SCONE include:

- List Library of Congress Subject Headings A-Z.
- List names of people and organizations as subjects A-Z.
- Search for words in subjects.
- List Dewey Decimal Classification numbers.

The second method, "People and organizations as subjects", is a pre-limited version of the general option for retrieval by names.

SCONE collection-level descriptions are assigned at least one Dewey Decimal Classification (DDC) number and at least one Library of Congress Subject Heading (LCSH), where appropriate. The IRIScotland collections, at the institutional level, are not appropriate for assigning DDC or LCSH because there is no apparent specific subject focus. However, the departmental sub-collections for the University of Stirling digital research repository are appropriate, assuming that most, if not all, theses reflect the subject of the department.

For example: "302.23" (DDC) and "Mass media" (LCSH) both retrieve "Dept. of Film and Media eTheses".

- **Try it out**: Search SCONE for collections classified with DDC number 302.23 (using a URL parameter as described below).
- **Try it out**: Dewey classification numbers of collections (using the option from the full SCONE service).
- **Try it out**: Search SCONE for collections with LCSH containing "mass media" (using a URL parameter as described below).
- **Try it out**: Browse collection subjects (using the option from the full SCONE service).
- **Try it out**: Search for words in collection subjects (using the option from the full SCONE service).

For collections which have no specific subject focus, SCONE can offer retrieval by subject strength by using the Scottish Conspectus scheme used in the Research Collections Online (RCO) service. Conspectus uses the set of subject topics represented in the top levels of the Library of Congress Classification and assigns a number to each to represent the collection strength in that topic. CDLR is investigating the feasibility of applying the scheme to institutional repositories. Tools for adding new Conspectus collections to RCO and SCONE are already in place, having been developed as part of the Scottish Collections Network Extension project. For further information, see Extending the Research Collections Online (RCO) service : final report of the RSLP SCONE project, annexe C, and Technical and functional description of the Scottish Collections Access Management Portal (SCAMP) demonstrator service.

Interoperability between SCONE and IRIScotland

A record in SCONE includes a link to the most co-extensive finding-aid or catalogue appropriate to the collection. For individual IRIScotland repositories, this is the repository's local retrieval service; for IRIScotland as a whole, this is the pilot search/browse interface.
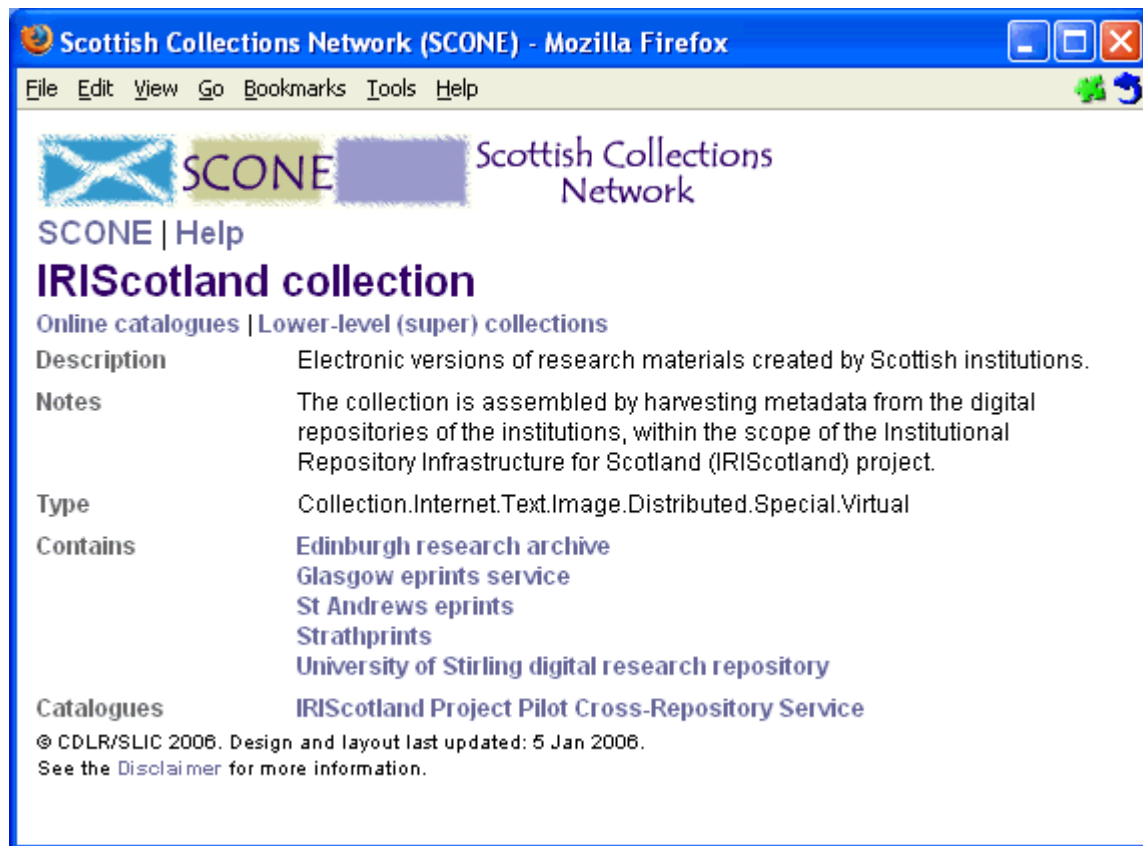
If, for any reason, a collection catalogue is notified as being unavailable, SCONE automatically

substitutes the next higher-level finding-aid. In the case of IRIScotland, if the local repository retrieval service is taken offline for maintenance or as a result of local networking problems, the IRIScotland search/browse interface is linked instead. The IRIScotland service will thus act as a back-up to local services (at least, for searching; depending on local conditions, the resources themselves may be offline).

SCONE provides a link-back facility which allows a single collection-level description to be displayed by specifying its ID number in a URL parameter name/value pair. The record is displayed in a "simple" version of the SCONE interface which restricts navigation to the collection and its super- and sub-collections. An option to link to the full SCONE service is also provided.

The URL for the top-level IRIScotland collection is
http://scone.strath.ac.uk/service/portal/pColnShow.cfm?uColnID=7972.

Screenshot: The top-level IRIScotland collection.

This Screenshot shows the IRIScotland "collection" with:

- A link to the pilot IRIScotland search/browse service.

- Links to the specific repositories which form sub-collections of IRIScotland.

Screenshot: An IRIScotland repository collection.

This Screenshot shows a specific IRIScotland repository collection-level description with:

- A link to the local repository retrieval interface, rather than the IRIScotland search/browse service. Only items in this repository can be found in the cited catalogue.
- Links to sub-repository collections of the local repository.
- A link to contact information about the owner of the items in the collection.
- A link to the aggregated IRIScotland super-collection.
- Links to other aggregation services which harvest this repository, treated as super-collections. Each of these has its own finding-aid which can be used as a back-up if necessary.
- A link to the University Library as a super-collection. The University repository is administered by the Library and can be considered as part of the Library's aggregated collection. The Library may contain University research material which has not been registered in the repository for legacy or format reasons, and it may contain non-research

publications by the University, its staff, and students.

Screenshot: An IRIScotland sub-repository collection.

This Screenshot shows a specific IRIScotland sub-repository collection-level description with:

- A link to the local sub-repository retrieval interface, rather than the IRIScotland search/browse service. Only items in this specific sub-repository can be found in the cited catalogue.
- LCSH and DDC subjects specific to this sub-repository.
- A link to contact information about the owner of the items in the repository.
- A link to a super-collection which is the full local repository.

Context

SCONE places IRIScotland and its constituent repositories in the context of their geopolitical focus. Users can readily identify the relationship of a specific research repository to:

- the local institution.
- Scotland's research community.
- Scotland's information environment as a whole.

Integrating access to the institutional research repositories within SCONE allows:

- Identification of Scottish institutions, staff and students conducting research on similar

topics (supporting collaboration, administration and planning of research).

- Identification of and access to non-institutional research collections located in Scottish libraries, archives and museums (supporting research methodologies such as literature searching and referencing).
- Identification of and access to learning and teaching resources derived from previous research on a specific topic (supporting research advocacy, "best value", and social inclusion).
- Improvement of collaborative collection management by information professionals in Scotland.
- Future integration with facilities and services under development, such as the High-Level Thesaurus (HILT) project.

SCONE will add new IRIScotland repositories as they become available online, as well as other Scottish repositories for non-research materials such as learning and teaching resources or administrative documents made available under Freedom of Information. If you have, or know of, an active repository managed by any Scottish organisation or individual, and cannot find it in SCONE, please contact the author and ask for it to be registered.

Gordon Dunsire
Depute Director
Centre for Digital Library Research
University of Strathclyde

Related event: Metadata issues for Scottish institutional repositories

Please note this forthcoming event, of interest to cataloguers and metadata practitioners working with digital resources and institutional repositories.

Title: Metadata issues for Scottish institutional repositories
Date: Monday 4 Dec 2006
Time: 14.00-16.30
Venue: Conference Room, National Library of Scotland, Causewayside Building, Edinburgh.
Organisers: CIGS, CDLR, NLS.
Registration: Email scotearl@slainte.org.uk with your name and institution.
More information about the event and venue is available.

Programme

1. Presentation of results of harvesting metadata from Scottish institutional repositories during the HaIRST and IRIScotland projects / Alan Dawson. This will include a demo of the pilot search/browse service for IRIScotland.
2. Presentation on the DC ePrints application profile / Julie Allinson.
3. Presentation on the potential use of the DC ePrints application profile in Scotland / Gordon Dunsire.
4. General discussion and plenary.

Aim

1. To inform the IRIScotland Work Package 5 deliverable of a metadata agreement.
2. To inform the IRIS Work Package 4 deliverable of a pilot integrated hosted repository

service.

3. To inform the IRIS Work Package 5 deliverable of a pilot search/browse service.

4. To inform the Scottish and wider cataloguing/metadata community of issues arising from the development of institutional repositories in general, and progress on related issues in IRIS.

5. To inform the National Library of Scotland of issues in relation to the Trusted Digital Repository initiative and the mixed metadata schema environment of the John Murray Archive and other developments.

Background

The Cataloguing and Indexing Group in Scotland (CIGS) has collaborated with the Centre for Digital Library Research (CDLR) on similar exercises in the past.

CIGS co-ordinated and chaired a series of meetings of SCURL cataloguers to discuss local practice and its potential effect on Z39.50-mediated distributed union catalogues, as part of the CAIRNS project. The discussions at the meetings were used to draw up a checklist of short- and long-term changes to local practice which would improve interoperability in the CAIRNS service. The checklist was approved by the Scottish Confederation of University and Research Libraries (SCURL), and was effectively a metadata agreement. This subsequently formed the basis of a work package in the cc-interop project which investigated the development of a similar agreement on a UK-wide basis. CIGS hosted a seminar for Scottish cataloguers, and liaised with a parallel seminar for England and Wales. The outcome was a report saying that the original CAIRNS agreement, although a suitable starting-point, could not be extended without the support of a high-level UK agency to co-ordinate metadata schema, formats and content guidelines.

The current CDLR approach to delivering a metadata agreement for Scottish institutional repositories is similar: identify issues directly from examination of the content of aggregated metadata; expose those issues to a wide professional audience for feedback and identification of additional background issues; develop an initial proposal for discussion by SCURL cataloguers; develop a draft final proposal for consideration by SCURL itself. The deliverable is scheduled for March 2007.

Audience

1. Cataloguing/metadata practitioners in Scottish universities and colleges.

2. Cataloguing/metadata practitioners and managers in the NLS (including those working on the John Murray Archive and the Trusted Digital Repository).

3. CIGS members.

4. Anyone else who is interested in metadata standards and workflows applied to institutional repositories and self-archiving services. You do not have to be a member of CILIP/CILIPS or CIG/CIGS to attend.

Preparation in advance of the meeting

There are several public documents on the IRIS project wiki.

These give the background to the metadata issues, and attendees may find it useful to read them in advance.

Registration fee: Free. Please register in advance to avoid any problems with security and refreshments.

Light refreshments will be provided at 13.45.

Collection landscaping and institutional repositories in Scotland |
Related event: Metadata issues for Scottish institutional repositories

Home | Contents by issue | SAPIENS homepage