

**IRIScotland Project**  
**Work Package 5**  
**Report, evaluation and recommendations for the  
pilot service**

Gordon Dunsire and Alan Dawson

Centre for Digital Library Research, University of Strathclyde, Glasgow

September 2007

# Contents

Contents .....	2
Summary of recommendations.....	6
Introduction.....	7
Background.....	7
Design of pilot IRI-Scotland search/browse service .....	8
Methodology .....	8
Initial harvesting of extant repositories.....	8
Development of scenarios and use case .....	8
Analysis of stakeholder questionnaire.....	8
Design requirements .....	8
Initial pilot .....	8
Integration with Scottish Common Information Environment.....	9
Final pilot.....	9
Interface enhancements.....	10
Subject browsing .....	11
Evaluation of the pilot service.....	12
Pilot service usage statistics.....	12
Pilot service harvesting statistics .....	12
Stakeholder feedback .....	12
Functionality of other cross-repository information retrieval services.....	13
Comparison with functional requirements.....	13
Notes .....	14
Improvement of pilot service functionality.....	15
Maintenance of the pilot service .....	15
Skills requirements.....	15
Harvesting issues and problems.....	15
Harvester configuration .....	15
Harvesting schedules .....	16
Harvester target maintenance .....	16
Running multiple harvester configurations .....	16
Other harvesters.....	17

Re-harvesting .....	17
Metadata issues and problems .....	17
Metadata agreement .....	17
Metadata processing .....	17
Conclusions .....	18
Appendix A: Scenarios and use cases .....	19
Appendix B: Analysis of the search service retrieval question (q.12) of the IRI-Scotland project academic author survey .....	30
Retrieval on specific subject scheme .....	30
Retrieval on full text of the document .....	31
Appendix C: Functional specification of IRIScotland search service .....	32
Basic search options .....	32
Input search value .....	32
Intermediate results page .....	33
Intermediate results filter .....	33
Intermediate results sort .....	33
Short description content .....	33
Full description content .....	34
Advanced search page .....	34
Basic interface flow .....	34
Appendix D: Some notes from the EThOS workshop, Glasgow University, 20 July 2006.....	40
Appendix E: Collection-level description and IRIScotland information retrieval.....	42
SCONE .....	42
Application to IRIScotland collections .....	42
Scaling .....	43
General retrieval .....	44
Subject retrieval .....	45
Appendix F: Subject browsing issues.....	46
Appendix G: Comparison of IRIScotland and DRIVER Dublin Core metadata guidelines .....	57
Brief .....	57
Metadata element analysis .....	57
DC Element Variable: Creator .....	57
Analysis.....	57

DC Element Variable: Title .....	57
Analysis.....	57
DC Element Variable: Subject.....	58
Analysis.....	58
DC Element Variable: Description.....	58
Analysis.....	58
DC Element Variable: Publisher.....	58
Analysis.....	58
DC Element Variable: Contributor .....	59
Analysis.....	59
DC Element Variable: Date .....	59
Analysis.....	59
DC Element Variable: Type.....	59
Analysis.....	60
DC Element Variable: Format .....	60
Analysis.....	60
DC Element Variable: Identifier .....	60
Analysis.....	60
DC Element Variable: Source .....	60
Analysis.....	61
DC Element Variable: Language.....	61
Analysis.....	61
DC Element Variable: Relation.....	61
Analysis.....	61
DC Element Variable: Coverage .....	61
Analysis.....	62
DC Element Variable: Rights.....	62
Analysis.....	62
DC Element Variable: Audience.....	62
Analysis.....	62
Summative Analysis.....	62
Appendix H: Technical configuration.....	63

Hardware.....	63
Software.....	63
ARC harvester.....	63
ARC parser.....	63
MySQL.....	63
MySQL Control Center.....	63
ColdFusion.....	63
Microsoft Access.....	64

## Summary of recommendations

- ✓ The pilot service for searching and browsing cross-repository metadata for Scottish research output should be upgraded to an operational service.
- ✓ The operational service requires the technical and administrative skills required in a local repository, together with programming and data processing expertise.
- ✓ An operational service should consider making its metadata aggregation available for re-harvesting via OAI-PMH to support higher-level aggregations such as DRIVER.
- ✓ An infrastructure to support regular monitoring, assessment, review and upgrading of technical components of the operational service is required.
- ✓ An infrastructure to support the maintenance of technical information used by the harvester to interact with repositories is required. The infrastructure must
  - support regular and ad hoc contact between service and institutional repository administrators;
  - allow regular monitoring, trouble-shooting and resolution of technical information about each institutional repository in the service.
- ✓ It should be recognised that a necessary requirement of a successful operational service is the availability of appropriate technical expertise to the local institutional repository.
- ✓ An infrastructure to support regular monitoring, assessment, review and amendment of technical components of any metadata agreement used by partners in the service is required.
- ✓ An infrastructure to develop, maintain and apply processes for manipulating harvested metadata from each repository to ensure service functionality is required.
- ✓ Adoption of the draft metadata agreement by partners in the service should be a priority.
- ✓ The operational service should develop additional functionality as and when application of the metadata agreement by partners in the service permits.
- ✓ The National Library of Scotland and Scottish Library and Information Council should consider further integration of the operational service with the Scottish Common Information Environment to provide enhanced functionality.

## Introduction

This report constitutes the final deliverable of Work Package 5 of the IRIScotland project<sup>1</sup>.

Work Package 5 was carried out by the Centre for Digital Library Research (CDLR)<sup>2</sup>.

## Background

Deliverable D.5.6 of the IRIScotland project plan<sup>3</sup> is an "Evaluative report of the service with recommendations for moving the service from pilot to operation after completion of the project."

The project plan states that the aim of the pilot service is "... to establish ... a harvester-based pilot cross-repository search and browse service to enhance exposure of the Scottish research output as a whole ..."

The project plan schedule sets the context for this deliverable:

Due	Reference	Description
March 2006	D.5.3	Creation of test database of harvested records, following testing.
July 2006	D.5.1	Determination, agreement, and initial implementation of the service design requirements.
December 2006	D.5.4	Develop user interface facilities.
March 2007	D.5.2	Agreement and implementation of metadata standards, and production of best practice guidelines.
May 2007	D.5.5	Implementation and testing of the pilot service, in conjunction with staff and user evaluations in partner institutions.
July 2007	D.5.6	Evaluative report of the service with recommendations for moving the service from pilot to operation after completion of the project.

---

<sup>1</sup> IRIScotland. Available at: <http://www.iriscotland.lib.ed.ac.uk/>

<sup>2</sup> Centre for Digital Library Research. Available at: <http://cdlr.strath.ac.uk/>

<sup>3</sup> IRI-Scotland project plan, version 1.5, 15 February 2007

# Design of pilot IRI-Scotland search/browse service

## Methodology

### Initial harvesting of extant repositories

The institutional repositories of Edinburgh, Glasgow, St Andrews and Strathclyde universities were being harvested as part of the maintenance of the Experimental discovery service<sup>4</sup> implemented during the Harvesting Institutional Resources in Scotland Testbed (HaIRST)<sup>5</sup> project.

The HaIRST database of harvested metadata records was configured to allow the development of a basic retrieval interface based on the ARC package. The interface was used to make initial analyses of the content of the metadata and its impact on recall and precision in the pilot service.

### Development of scenarios and use case

A set of six scenarios<sup>6</sup> for the pilot service was developed using a standard JISC template. The content of the scenarios was based on the aims of the project, informal discussion with the IRI-Scotland Technical Committee and the Scottish Library and Information Council.

The scenarios were used to develop a use case<sup>7</sup> for the pilot service using a standard JISC template.

### Analysis of stakeholder questionnaire

A question on the kinds of retrieval option required by service stakeholders was included in the IRI-Scotland project academic author survey. The responses to the question were analysed<sup>8</sup> to identify the options most often chosen.

### Design requirements

A functional specification<sup>9</sup> for the pilot service was developed, based on the use case and questionnaire analysis, and using the expertise of the CDLR staff in designing and implementing simple information retrieval systems.

### Initial pilot

A first version of the pilot service was made available for demonstration, testing and comment in June 2006.

---

<sup>4</sup> Experimental discovery service. Available at:  
[http://speirserver.cdlr.strath.ac.uk:8088/arc/hairst\\_search.jsp](http://speirserver.cdlr.strath.ac.uk:8088/arc/hairst_search.jsp)

<sup>5</sup> HaIRST: Harvesting Institutional Resources in Scotland Testbed. Available at:  
<http://hairst.cdlr.strath.ac.uk/>

<sup>6</sup> See Appendix A.

<sup>7</sup> See Appendix A.

<sup>8</sup> See Appendix B.

<sup>9</sup> See Appendix C.

## Integration with Scottish Common Information Environment

Proposals for further development of the EThOS service<sup>10</sup> and its potential impact on the pilot service<sup>11</sup> showed that the simple harvesting of metadata from Scottish research repositories would not, of itself, guarantee that the service would cover all repositories containing Scottish research output. For example, The DEPOT service<sup>12</sup> is also scoped as a UK "national" repository, and may be used by individual researchers associated with Scottish universities which do not have their own institutional repository.

The CC-interop project<sup>13</sup> had investigated the application of collection-level description to landscaping distributed and overlapping metadata aggregations, and providing a source of information about aggregations at multiple levels of granularity, using the Scottish Collection Network (SCONE) as a case study<sup>14</sup>. This approach was adopted to include the institutional repositories and harvested metadata aggregation of the pilot service<sup>15</sup>.

Integration with SCONE has also enhanced the pilot service by providing:

- The institutional and departmental browse functionality identified in the functional specification.
- Discovery of collections related to institutional research repositories through ownership, administration and subject.

This should inform the ongoing discussions between the National Library of Scotland and Scottish Library and Information Council about approaches to the Scottish Government and the development of Digital Access Scotland.

- ✓ **Recommendation:** The National Library of Scotland and Scottish Library and Information Council should consider further integration of the operational service with the Scottish Common Information Environment to provide enhanced functionality.

## Final pilot

The final version of the pilot service<sup>16</sup> was made available for demonstration, testing and comment in May 2007.

The service currently includes seven Scottish research repositories, for Aberdeen, Edinburgh, Glasgow, Queen Margaret, St Andrews, Stirling and Strathclyde universities.

---

<sup>10</sup> Electronic theses in the UK: the open access future. Available at: [http://www.jisc.ac.uk/whatwedo/programmes/programme\\_rep\\_pres/ethosnet\\_announcement\\_mar07](http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/ethosnet_announcement_mar07)

<sup>11</sup> See Appendix D.

<sup>12</sup> The DEPOT. Available at: <http://depot.edina.ac.uk/>

<sup>13</sup> CC-interop: COPAC/Clumps Continuing Technical Cooperation Project. Available at: <http://ccinterop.cdrl.strath.ac.uk/>

<sup>14</sup> Collection landscaping in the common information environment: a case study using the Scottish Collections Network (SCONE). Available at: <http://ccinterop.cdrl.strath.ac.uk/documents/ccicldlandscaping.pdf>

<sup>15</sup> See Appendix E.

<sup>16</sup> IRISScotland Project Pilot Cross-Repository Service. Available at: <http://cdlr.strath.ac.uk/irisotland/>

A technical configuration of the service is available<sup>17</sup>. Technical components supplied by third-parties are in continuous development, and the service does not currently use the latest versions of such components.

- ✓ **Recommendation:** An infrastructure to support regular monitoring, assessment, review and upgrading of technical components of the operational service is required.

## Interface enhancements

Various enhancements were made to the user interface during the course of the project, including:

- Providing a basic keyword search (searching all main fields) alongside an advanced search of specific fields.
- Providing options to search one or more repositories together.
- Paging of results, to improve response speed.
- Parsing of user queries, to remove stop words, intercept odd characters, and validate term length.
- Providing various options for ordering of search results.

---

<sup>17</sup> See Appendix H.



## IRIScotland Project Pilot Cross-Repository Service

Search records harvested from participating institutions, including Aberdeen, Edinburgh, Glasgow, Queen Margaret, St Andrews, Stirling and Strathclyde universities.

### Basic Search:

Search for:

This searches all repositories for matches in title, author, abstract or subject.

### Advanced Search:

Search for:

Search in:  Title and Abstract  Title  Author  Subject

Select from:

- Aberdeen University
- Edinburgh University
- Glasgow University
- Queen Margaret University
- St Andrews University
- Stirling University
- Strathclyde University

Select one or more institutions, or select an [i] link to display information about that repository.

Sort results by:

We welcome feedback on this IRIScotland pilot service. Please send comments to [alan.dawson@strath.ac.uk](mailto:alan.dawson@strath.ac.uk)

### Screen shot of the final version of the pilot service.

### Subject browsing

An option to provide subject browsing was set up and evaluated, but the service was so poor and unhelpful that it was not offered to users via the pilot service. This was due to the absence of any agreed standard for subject cataloguing, and to several well-documented problems in subject interoperability<sup>18</sup>. An analysis of the content of subject fields in metadata harvested from the repositories extant at the start of project illustrates the difficulties<sup>19</sup>.

---

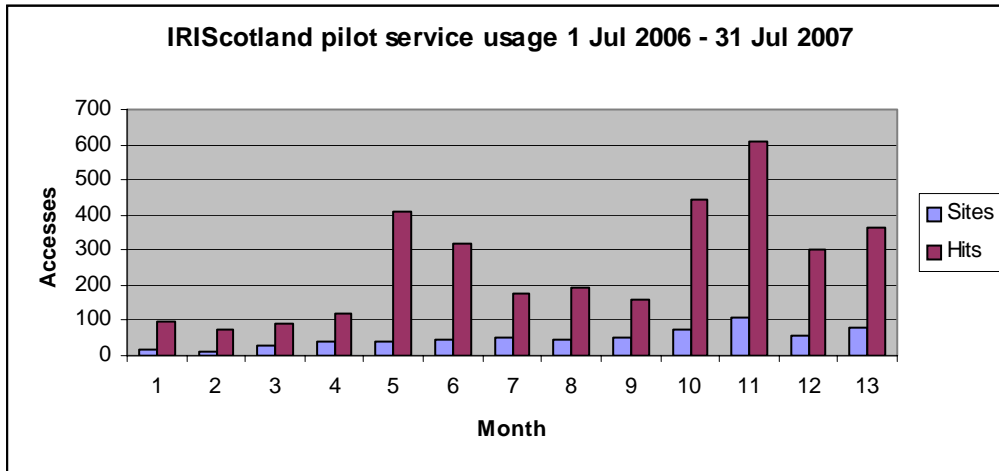
<sup>18</sup> Thirty problems for subject interoperability (and a few possible solutions). Available at: <http://cdlr.strath.ac.uk/pubs/dawsona/ad200602.htm>

<sup>19</sup> See Appendix F.

# Evaluation of the pilot service

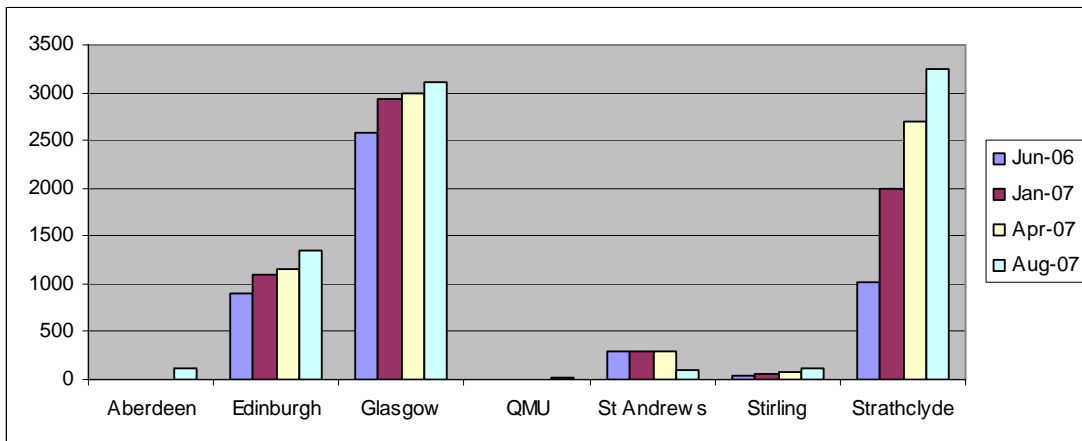
## Pilot service usage statistics

Usage statistics for the first year of the pilot service were compiled and analysed. Overall accesses were relatively small, as expected for a pilot service, but show a substantial increase over the year:



## Pilot service harvesting statistics

Snapshot totals for the number of harvested metadata records from each Scottish research repository were taken at ad hoc intervals. Overall, they show that the number of harvested records has been increasing at an approximately even rate for each repository. The downturn in the figures for St Andrews University is a result of the re-scoping, reconfiguration and relocation of its repository.



## Stakeholder feedback

The first and final versions of the pilot service have been demonstrated to several groups of stakeholders as part of the general dissemination activities of the IRIScotland Project Manager.

CDLR, in collaboration with the Cataloguing and Indexing Group in Scotland and the National Library of Scotland, organized a general seminar to discuss metadata issues for Scottish repositories. The seminar was held on 4 December 2006 at the National Library of Scotland, and was attended by approximately 50 metadata professionals, most from the library sector in

Scotland. Feedback from discussions was used to inform the pilot service as well as the draft metadata agreement.

Feedback via email from casual users of the pilot service has been sought on the service homepage since June 2007.

Feedback from casual visitors to CDLR has been sought, and has been generally favourable once the limitations of the harvested metadata have been demonstrated.

## Functionality of other cross-repository information retrieval services

There are a number of operational cross-repository information retrieval services which overlap with the current geographical and resource-type scope of the pilot service. These include The Depot<sup>20</sup>, DRIVER<sup>21</sup>, Intute repository search<sup>22</sup>, and OAlster<sup>23</sup>.

None of these services appears to offer, or plan to offer, an information retrieval focus on Scottish institutions and research output.

- ✓ **Recommendation:** The pilot service for searching and browsing cross-repository metadata for Scottish research output should be upgraded to an operational service.

## Comparison with functional requirements

The following table compares the facilities of the pilot service with the functional requirements identified during its design phase:

Function	Available?
<b><i>Basic search functions</i></b>	
General keyword search	Y (title and abstract)
Title keyword search	Y
Subject browse	N <sup>(a)</sup>
Subject keyword search	Y
Author browse	N <sup>(b)</sup>
Author search	Y

---

<sup>20</sup> The DEPOT. Available at: <http://depot.edina.ac.uk/>

<sup>21</sup> DRIVER: Digital Repository Infrastructure Vision for European Research. Available at: <http://www.driver-repository.eu/>

<sup>22</sup> Intute: repository search. Available at: <http://irs.ukoln.ac.uk/>

<sup>23</sup> OAlster – find the pearls. Available at: <http://www.oaister.org/>

Institution browse	N <sup>(c)</sup>
Institution search	Y (as a filter)
Department browse	N <sup>(c)</sup>
Department search	N <sup>(c)</sup>
RAE Unit of Assessment browse	N <sup>(d)</sup>
RAE Unit of Assessment search	N <sup>(d)</sup>
<b><i>Input search value</i></b>	
Case-independent matching	Y
Phrase matching	Y
<b><i>Search results</i></b>	
Short intermediate descriptions	Y
Results filter	Y (institution filter is pre-search)
Results sort	Y
Full description	N <sup>(e)</sup>

## Notes

(a) The content of the harvested metadata is too divergent<sup>24</sup> to make this an effective option. Content varies in:

- Consistency of spelling and syntax
- Structure with multiple terms in a single field
- Terminology of different classification and subject heading schemes

(b) The content of the harvested metadata is too divergent to make this an effective option. Content varies in:

- Consistency of varying forms of the same name and similar forms of different names
- Structure with multiple names in a single field

---

<sup>24</sup> See Appendix F.

- (c) This facility is provided via the linked SCONE collection description which provides access to the local repository retrieval service.
- (d) The content of the harvested metadata does not include this information.
- (e) The identifier field in most of the harvested metadata links to the full metadata description in the local repository, rather than the resource itself. Although this is not best practice, it duplicates the functional requirement for the pilot service, which is currently omitted to avoid confusing end-users.

## Improvement of pilot service functionality

Addition of functional requirements missing from the pilot service is dependent on the quality of the harvested metadata.

The draft IRIScotland metadata agreement<sup>25</sup> addresses these issues and makes recommendations for ensuring that local repository metadata made available for harvesting is suitable for meeting the functional requirements of the pilot service.

- ✓ **Recommendation:** The operational service should develop additional functionality as and when application of the metadata agreement by partners in the service permits.

## Maintenance of the pilot service

### Skills requirements

The technical and administrative skills required to maintain the service include many of those identified for a local institutional repository<sup>26</sup>. The aggregation database of metadata records in the service has similar qualities to the databases used in local repositories, and the technical standards infrastructure is identical. The service administrators have to be able to interact with local repository administrators rather than the local resource creators.

The service administrators also require technical skills for maintaining the harvester and processing harvested metadata as outlined below.

- ✓ **Recommendation:** The operational service requires the technical and administrative skills required in a local repository, together with programming and data processing expertise.

## Harvesting issues and problems

### Harvester configuration

The ARC software is supplied with a web user interface which works well for adding new repositories for harvesting. This simply provides a convenient interface to the MySQL database.

However, for testing of harvesting, reconfiguring repository details, switching harvesting on and off for specific sites, etc, it is necessary to edit the tables in the MySQL database. The MySQL Control Center is useful for this purpose.

---

<sup>25</sup> Draft IRIScotland metadata agreement: standards and guidelines for institutional repositories. Available at: <http://cdlr.strath.ac.uk/pubs/dawsona/irismetadatadraft.pdf>

<sup>26</sup> Institutional repositories: staff and skills requirements. Available at: <http://www.sherpa.ac.uk/documents/sherpaplusdocs/notts-Repository%20Staff%20and%20Skills.pdf>

## Harvesting schedules

Harvesting can be scheduled to run automatically at regular intervals, e.g. weekly. In practice this only worked well sporadically, so is supplemented by manually controlled harvesting (issuing a specific command to the harvester from a batch file).

## Harvester target maintenance

The number of repositories registered as targets for the harvester increased from four to seven during development of the pilot service.

Registration of a single target often required several email communications with the institutional repository administrators.

One target registration required complete reconfiguration when the institutional repository (St Andrews University) changed location and infrastructure.

One new repository (Queen Margaret University) was identified during a casual discussion. The repository administrators did not know of the pilot service; the repository was subsequently added to the service.

One repository (Edinburgh University) required complete re-harvesting on two separate occasions as a result of local configuration changes.

The implication is that monitoring of changes in the availability of harvestable repositories is a significant task if the service is to be kept up-to-date.

- ✓ **Recommendation:** An infrastructure to support the maintenance of technical information used by the harvester to interact with repositories is required. The infrastructure must
  - support regular and ad hoc contact between service and institutional repository administrators;
  - allow regular monitoring, trouble-shooting and resolution of technical information about each institutional repository in the service.
- ✓ **Recommendation:** It should be recognised that a necessary requirement of a successful operational service is the availability of appropriate technical expertise to the local institutional repository.

## Running multiple harvester configurations

For the first year of the IRIScotland project the same database of harvested records was used for the HAIRST Experimental discovery service<sup>27</sup> and the pilot service.

ColdFusion scripts were used to filter out repositories in the database that were not relevant to IRIScotland. This worked successfully with a small number of repositories, but was not a viable solution in the longer term.

The original database was therefore cloned, with one being configured for HAIRST and one for IRIScotland.

The same single installation of the ARC harvester is currently used to populate both databases at different times. Two separate batch files, one for HAIRST and one for IRIScotland, are used to control configuration prior to harvesting and to specify which database is to be updated.

---

<sup>27</sup> Experimental discovery service. Available at:  
[http://speirserver.cdrl.strath.ac.uk:8088/arc/hairst\\_search.jsp](http://speirserver.cdrl.strath.ac.uk:8088/arc/hairst_search.jsp)

## Other harvesters

There are various options for harvesting software.

The PKP harvester was installed and tested and worked well within the institutional domain, but it was not possible to configure it with details of a proxy server, so it could not be used for harvesting from distributed repositories.

## Re-harvesting

The aggregation of metadata harvested by the pilot service is not specifically exposed for re-harvesting by other aggregation-based services.

It is technically possible to do this, and there are several potential advantages in doing so:

- Metadata in the pilot service aggregation is in a better standards-compliant state than the metadata in the local institutional repository, as a result of post-harvest processing by the service.
- The aggregation constitutes a single repository from the point of view of a third-party aggregation service, with lower resource implications than if all of the local repositories were to be harvested separately.
- The aggregation has a well-defined geographical scope, which is a good fit with third-party services based on national/regional granularity such as DRIVER.
- Local repositories would not have to deal with queries or make notification of changes of repository configuration to third-party services.

The draft metadata agreement has been compared with the Specific guidelines for DRIVER metadata<sup>28</sup>, and no significant interoperability issues have been identified.

- ✓ **Recommendation:** An operational service should consider making its metadata aggregation available for re-harvesting via OAI-PMH to support higher-level aggregations such as DRIVER.

## Metadata issues and problems

### Metadata agreement

The draft metadata agreement has been distributed to project partners, but it has not been ratified by any consortium representing the partners.

Implementation of the agreement is currently not apparent in metadata harvested from repositories, notably those using DSpace.

- ✓ **Recommendation:** Adoption of the draft metadata agreement used by partners in the service should be a priority.

### Metadata processing

In order to run the pilot service it has been necessary to manipulate metadata after harvesting, by controlled editing of the MySQL database. Some examples of metadata modification are given below:

---

<sup>28</sup> Exposing textual resources with OAI OAI-PMH: Test-bed application. Version 1.1. Annexes. 1: Specific guidelines for DRIVER metadata. Available at: [http://www.driver-support.eu/documents/DRIVER\\_guidelines\\_annexes\\_1%200.pdf](http://www.driver-support.eu/documents/DRIVER_guidelines_annexes_1%200.pdf)

- To replace non-standard characters, such as smart quotes and long dashes, with standard ASCII and XML-compliant characters.
- To parse citations stored in the Identifier field, extract URLs from the citation, and replace the citation with the URL.
- To delete obsolete records from test harvesting or from superseded repositories.
- To change the repository name or identifier, e.g. after server or service renaming, so that all records harvested from the same institution are recognised as being from the same repository.

Some metadata manipulation has been carried out by manually editing the database, and some by running a metadata modification program to retrieve, modify and update the metadata in the MySQL database. Although this has improved the pilot service significantly, numerous metadata interoperability problems remain.

Adoption of a common metadata agreement by service partners will resolve many of these problems. As the draft IRIScotland agreement indicates, the technical components of an agreement are subject to rapid development of underlying standards and interoperability architectures, and will require regular review and amendment.

- ✓ **Recommendation:** An infrastructure to support regular monitoring, assessment, review and amendment of technical components of any metadata agreement by partners in the service is required.

Although some simple metadata processes can be applied to metadata from several repositories, other processes are specific to a single repository. It is also likely that repository-specific processes will be required if the need arises to partially-harvest metadata from institutional repositories with a wider scope than the service and to de-duplicate records from overlapping repositories such as third-party aggregations. This has implications for the scaling of the service.

- ✓ **Recommendation:** An infrastructure to develop, maintain and apply processes for manipulating harvested metadata from each repository to ensure service functionality is required.

## Conclusions

The pilot service has met many of its design and functional requirements, and usage and content statistics show a steady increase. The pilot service is the only current means of cross-searching the research output of Scotland, and only Scotland.

The technical aspects of the harvesting process are fairly straightforward. The ARC harvesting software is robust, reliable and easy to configure, and no problems were found with running it. However, managing and running a service based on harvesting requires frequent monitoring and periodic manual intervention, even with the small number of repositories participating in the IRIScotland project.

Although some significant changes were made to the harvested records, metadata interoperability remains a big issue. The value of any service based on harvested metadata will be critically dependent on the quality and consistency of the harvested metadata.

## **Appendix A: Scenarios and use cases**

The following scenarios and use case are presented in standard JISC forms.

They were published on the IRI-Scotland wiki for comment in July 2006. They were also submitted and added to the wiki of JISC's Research Repositories Team at [http://www.ukoln.ac.uk/repositories/digirep/index/All\\_the\\_Scenarios\\_and\\_Use\\_Cases\\_Submitted](http://www.ukoln.ac.uk/repositories/digirep/index/All_the_Scenarios_and_Use_Cases_Submitted)

## JISC Digital Repositories Programme

### Scenario Collection Form

<b><i>Title</i></b>	<b>Identifying research on a topic with a focus related to Scotland</b>
<b><i>Author</i></b>	<b>Gordon Dunsire</b>
<b><i>Project</i></b>	<b>IRI-Scotland</b>
<b><i>ID</i></b>	<b>IRIScotland_Scenario_01</b>
<b><i>Date</i></b>	<b>6 Jul 2006</b>
<b><i>Narrative</i></b>	<p>Angela is a member of a research unit in a Scottish university. Her main research topic is the use of controlled vocabularies to improve information retrieval, and she is currently working on a project which is part of a programme to develop access to local government information online, funded by the Forum of Local Authorities in Scotland.</p> <p>She wishes to identify and access relevant research conducted elsewhere, and is particularly interested in the impact of the use of general, non-local vocabularies.</p> <p>She thinks that it is likely that most of the relevant research involving local Scottish vocabularies has been conducted by institutions located in Scotland.</p> <p>She tries to use a retrieval service scoped for the whole of the UK, but it does not provide a means of specifying this aspect of her search criteria. The service does allow the specification of a named institution, but not "all institutions located in Scotland". She does not want to have to specify every Scottish university and college in her search, and she is not sure that she even knows the names of all of them.</p>
<b><i>Other information</i></b>	

## JISC Digital Repositories Programme

### Scenario Collection Form

<b><i>Title</i></b>	<b>Identifying Scottish institutions active in a specific research field</b>
<b><i>Author</i></b>	<b>Gordon Dunsire</b>
<b><i>Project</i></b>	<b>IRI-Scotland</b>
<b><i>ID</i></b>	<b>IRIScotland_Scenario_02</b>
<b><i>Date</i></b>	<b>6 Jul 2006</b>
<b><i>Narrative</i></b>	<p>The Institute of Holistic Logic is a teaching and research department in an English university. The Institute wishes to tender for a major research project funded by the Research Funding Council of Scotland, but the Council requires the lead partner to be a Scottish university.</p> <p>The Institute wishes to identify Scottish institutions with similar research interests, with a view to collaborating on the tender.</p>
<b><i>Other information</i></b>	

## JISC Digital Repositories Programme

### Scenario Collection Form

<b>Title</b>	<b>Identifying the level of activity of in a specific research field within Scotland</b>
<b>Author</b>	<b>Gordon Dunsire</b>
<b>Project</b>	<b>IRI-Scotland</b>
<b>ID</b>	<b>IRIScotland_Scenario_03</b>
<b>Date</b>	<b>6 Jul 2006</b>
<b>Narrative</b>	<p>The University of Fife intends to develop a centre of academic excellence in the emerging field of coal-based textile science. Prior analysis has shown that the centre could attract viable numbers of students for postgraduate courses in the topic, and has identified a small but significant number of commercial research opportunities within Scotland.</p> <p>The University wants to assess the potential competition within Scotland for attracting students and commercial collaborators.</p> <p>The University also wishes to identify individuals located within a reasonable travelling distance who may be able to advise on the development of the centre, act as external assessors or supervisors, etc.</p> <p>The University would like to identify the current research interests of other Scottish institutions by examining their research output overall.</p> <p>The University would also like to identify Scottish-based authors of research output in subjects related to coal-based textile science.</p>
<b>Other information</b>	

## JISC Digital Repositories Programme

### Scenario Collection Form

<b><i>Title</i></b>	<b>Measuring and assessing the research infrastructure of Scotland</b>
<b><i>Author</i></b>	<b>Gordon Dunsire</b>
<b><i>Project</i></b>	<b>IRI-Scotland</b>
<b><i>ID</i></b>	<b>IRIScotland_Scenario_04</b>
<b><i>Date</i></b>	<b>6 Jul 2006</b>
<b><i>Narrative</i></b>	<p>The Research Funding Council of Scotland has been asked by the Scottish Administration to review the publicly-funded research infrastructure of Scotland.</p> <p>The Council wishes to gather information in a range of parameters, including:</p> <p>Yearly research outputs by individuals, departments and institutions.</p> <p>Concentrations and gaps in coverage of research topics.</p> <p>Accessibility of outputs, and proportion of open-access to restricted access.</p>
<b><i>Other information</i></b>	

## JISC Digital Repositories Programme

### Scenario Collection Form

<b><i>Title</i></b>	<b>Accessing up-to-date information about a specific topic of interest to people living in Scotland</b>
<b><i>Author</i></b>	<b>Gordon Dunsire</b>
<b><i>Project</i></b>	<b>IRI-Scotland</b>
<b><i>ID</i></b>	<b>IRIScotland_Scenario_05</b>
<b><i>Date</i></b>	<b>6 Jul 2006</b>
<b><i>Narrative</i></b>	<p>The Hebrides Health Promotion Service is preparing an information pack on the banning of smoking in public places in Scotland.</p> <p>It wishes to include the results of recent research into passive smoking in public houses and restaurants, and is particularly interested in giving general statistics for the previous year.</p> <p>The Service identifies a paper published in a recent research journal, written by a postgraduate student at a Scottish university, which is based on a statistical survey covering the previous three years. The Service would like access to the survey results so that it can extract figures pertaining to the last year only.</p>
<b><i>Other information</i></b>	

## JISC Digital Repositories Programme

### Scenario Collection Form

<b>Title</b>	<b>Identifying and accessing a comprehensive set of information about a specific topic pertaining to Scotland</b>
<b>Author</b>	<b>Gordon Dunsire</b>
<b>Project</b>	<b>IRI-Scotland</b>
<b>ID</b>	<b>IRIScotland_Scenario_06</b>
<b>Date</b>	<b>6 Jul 2006</b>
<b>Narrative</b>	<p>Andrew is a Canadian whose great-grandfather emigrated from Scotland. He is now retired, but was formerly the Head of the Department of Scottish Studies at a Canadian university. He has a keen interest in the life and works of Robert Burns.</p> <p>Andrew and his wife have decided to take an extended holiday in Scotland. They plan to visit the birthplace of Andrew's great-grandfather as well as the major tourist attractions. Andrew would also like to visit places associated with Burns, including museum exhibitions. Andrew's wife has agreed to allow him side-trip of a week to do this (she does not share his passion for poetry), so he wants to plan ahead to maximise the opportunity.</p> <p>He would also love to be able to speak to anyone who has done recent research into Burns' poetry.</p> <p>He wants to be able to access as much information as possible on the Web to identify the locations of Burns-related sites, researchers and the locations of their institutions. He wants to know details of transport to these locations, local accommodation, and contact information for organisations and individuals.</p>
<b>Other information</b>	

## JISC Digital Repositories Programme

### **Use Case Name**

Web based service for information retrieval of academic research output from a country or region.

### **Use Case ID (unique identifier, used for file name also)**

IRIScotland\_Use\_Case\_01

### **Authors**

Author 1	Alan Dawson
Author 2	Gordon Dunsire

### **Project**

IRIScotland

### **Date Created**

7 Jul 2006

### **Use Case Summary**

User conducts search of repository of metadata for a variety of research outputs (including eprints, conference presentations, books, and raw data) harvested from academic institutions located in Scotland. End user utilises a public web-based interface and accesses the identified research outputs online.

### **Primary Actor (and goal)**

End user	To identify and access research-level information matching a range of criteria.
----------	---

### **Other Actors (and goals)**

Interface provider	To retrieve and display metadata matching user-specified values for one or more search parameters.
--------------------	--

Metadata repository	To obtain, store, and organise metadata for research outputs in a variety of forms and formats from all academic institutions located in a specified area (Scotland).
Institutional repositories	To provide metadata for local research outputs to the aggregated metadata repository, and to provide online access to locally-deposited open-access research outputs.

### ***Stakeholders and Interests***

Academic researchers	Identify external research activity and results of interest; expose record of own research beyond the institution.
Academic departments	Identify opportunities for research pools and other forms of collaboration; identify external research activity and results of interest; expose record of departmental research beyond the institution.
Academic institutions	Inform planning of research strategies.
Public administration organisations	Inform planning and funding of national (Scottish) research strategies.
Commercial organisations	Identify opportunities for investment and product development.
Research students	Identify individuals, departments and institutions active in subjects of interest; identify research activities of institutions of interest.
Adult learners, lifelong learners, learners engaged in continuing professional development	Identify most up-to-date information on topics of interest.
Teachers	Identify individuals, departments and institutions active in subjects of interest; identify research activities of institutions of interest; identify most up-to-date information on topics of interest.
Other information retrieval services (including the Scottish Common Information Environment)	Extend service scope to include resources at research-level, in various research forms, and/or produced in a specified geographical area (Scotland).

### ***Main Success Scenario***

1	End user accesses search page.
2	End user enters search criteria for metadata.
3	Search submitted to repository for processing.
4	Metadata records returned to end user.
5	Metadata records meet end users needs.

## ***Extensions***

1a	End user is unable to access search page.
1a1	Interface provider complies with standard guidelines on special needs requirements.
1a2	Interface provider publishes information about the standard web browsers which can be used to properly view the search page, and information for end users with special needs.
1b	End user is unable to find search page.
1b1	Interface provider designates the search page as its homepage of the published URL of the service.
1b2	Interface provider ensures a direct link to the search page from every other page of the service.
2a	End user is unable to enter values for desired search parameters.
2a1	Interface provider includes the most common search parameters on the search page.
2a1a	End user wishes to use a less common search parameter. (Fails)
2b	End user is unable to combine parameters to specify desired search criteria.
2b1	Interface provider includes an “advanced search” page which caters for Boolean combinations of parameters and values.
3a	End user receives error messages from the interface’s operating system.
3a1	Interface provider reviews and tests the interface when the interface software is amended.
3b	End user receives error messages from the metadata repository’s database system.
3b1	Interface provider reviews and tests the interface when the database management system is amended.
4a	End user cannot view metadata retrieved in the search.
4a1	Interface provider reviews and tests the interface when the interface software or database management system is amended.
5a	End user cannot use retrieved metadata to adequately identify records relevant to their needs.
5a1	Interface provider ensures metadata include information necessary to distinguish similar records.
5a2	Interface provider ensures metadata include information necessary and sufficient to determine relevance to the end user’s search.
5a3	Interface provider ensures metadata include adequate descriptive information.
6a	End user retrieves metadata which are not relevant to their search criteria.
6a1	Institutional repositories and metadata repository ensure consistency and coherency in metadata content.

7a	End user cannot access a resource described by the metadata because the institutional repository restricts access.
7a1	Interface provider ensures metadata includes information about legal and administrative restrictions on access.
8a	End user cannot link to a resource described as online and openly available.
8a1	Metadata repository harvests institutional repositories sufficiently frequently to ensure information about resource location and availability status is current.
9a	End user fails to access a resource retrieved from an institutional repository.
9a1	Interface provider ensures metadata includes information about technical requirements for access, including format and helper applications.

### ***Scenarios (id and name)***

IRIScotland_Scenario_01	Identifying research on a topic with a focus related to Scotland
IRIScotland_Scenario_02	Identifying Scottish institutions active in a specific research field
IRIScotland_Scenario_03	Identifying the level of activity of in a specific research field within Scotland
IRIScotland_Scenario_04	Measuring and assessing the research infrastructure of Scotland
IRIScotland_Scenario_05	Accessing up-to-date information about a specific topic of interest to people living in Scotland
IRIScotland_Scenario_06	Identifying and accessing a comprehensive set of information about a specific topic pertaining to Scotland

### ***UML Diagrams***

None.

## Appendix B: Analysis of the search service retrieval question (q.12) of the IRI-Scotland project academic author survey

This analysis is based on the results of the survey dated 25 Apr 2006.

Q.12: In a national or international cross-repository search service, would you wish your institutions [sic] publications to be retrieved by:

Author – Date – Department – Institution – RAE Unit of Assessment – Subject – Title – Other  
(please specify)

Total no. of responses to survey	488
No. of test, nonsense or no responses to q.12*	9
No. of responses analysed	479

\* 1 response said "None"; 1 response "did not understand the question".

The following table gives the results in ranked order.

Author	440
Subject	422
Title	391
Date	235
Institution	154
Department	132
RAE Unit of Assessment	73
Keyword**	49
Format***	2

\*\* Taken from "Other". Included are responses indicating retrieval by abstract and other general descriptive elements, and "tagging". Excluded are responses indicating retrieval of the text of the document.

\*\*\*Taken from "Other". The term "Format" refers to retrieval by type of document (refereed paper, thesis, etc.)

### Retrieval on specific subject scheme

3 responses specified a particular vocabulary for subjects. 2 specified Medical Subject Headings (MESH); 1 specified "classification code(s)".

## **Retrieval on full text of the document**

3 responses included retrieval based on the full text of the document; all 3 referred to "Google".

# Appendix C: Functional specification of IRIScotland search service

## Basic search options

Based on the analysis of the survey question:

- General keyword search
- Title keyword search
- Subject browse
- Subject keyword search
- Author browse
- Author search
- Institution browse
- Institution search {\*1}
- Department browse {\*2}
- Department search {\*2}
- RAE Unit of Assessment browse {\*3}
- RAE Unit of Assessment search {\*3}

\*1: Is this necessary? It would be required a) if the user did not know the proper institution name (e.g. X University, or University of X), and b) there were too many institutions to browse effectively. Condition b) is unlikely in the context of IRIScotland.

\*2: This data is not currently supplied by any of the harvested repositories, so Department would need to be included in the metadata agreement.

\*3: This data is not currently supplied by any of the harvested repositories, so RAE Unit of Assessment would need to be included in the metadata agreement.

The order of options is based on general-to-specific, and grouping similar entities (e.g. author, institution, department). This principle is often used in other information retrieval services, and is likely to be familiar to users.

The other options specified in the survey question, Date and Format, are best used as sort and filter options. A broad interpretation of Format is probably required, to cover file format (e.g. PDF) and availability (e.g. open-access format) as well as resource type (e.g. conference presentation).

An additional sort/filter option, Type, should be introduced to avoid confusion with Format.

An additional sort/filter option, Status, might be introduced to meet requirements for version-tracking.

## Input search value

All matching is case-independent.

Multiple word input is treated a) as a phrase (e.g. “quantum mechanics”) or b) with implicit Boolean AND between each word (e.g. “quantum” AND “mechanics”). Both cases can be allowed if the user encloses phrase input in quotes.

## Intermediate results page

Scaling issues require the use of intermediate short descriptions. A search may hit hundreds of records. If full records are displayed directly, the results page will be very large with increased loading time and inconvenience for the user.

Consistency is achieved by using results paging: Display first 10 or 20 results as short descriptions, with an option to view more.

(A possible refinement is to display a list of full descriptions if there are fewer than X hits, but this is inconsistent and may confuse users.)

## Intermediate results filter

The scale of even the intermediate results page may be too great for some users. The number of search results can be reduced by filtering by:

- Date
- Institution
- Format
- Type
- Status

## Intermediate results sort

Some users will find it convenient to re-order the presentation of search results:

- Date (both ways)
- Institution (and Department if included in the metadata agreement)

## Short description content

The function of the short description is to display enough detail to allow the user to identify which of the search results merits a full display.

The minimum field set includes:

- Title
- Date
- Resource type (e.g. conference presentation, pre-print)

It may be useful to add First author and/or the first lines of Description, emulating Google.

It may be useful to add an active URL for the resource, so the user can get direct access without going through the full description. This can be labelled with the Title (so clicking on the Title retrieves the resource).

The full description can be displayed using a “View full record” option. This follows Perx and RDN practice.

## Full description content

The function of the full description is to give as much information about the resource as possible.

For some users, the full description will be the required result of their search, giving enough information for reference or citation purposes.

## Advanced search page

The need for an advanced search, where the user can input multiple search values for multiple data elements, will be investigated after the initial pilot is made available for testing by IRIS staff and end-users.

This might take the form of query-by-example (a “blank” description in which the user enters search terms against data elements), or allowing the user to select multiple data elements and Boolean operators from specific lists (common in library catalogue interfaces).

User feedback, the quantity of records in the service, and the metadata agreement will all contribute to decisions on this functionality.

## Basic interface flow

Agent	Action	Notes
<b>General keyword search</b>		
User	Input a single search string.	
Service	Match input <i>anywhere-in-field</i> in Title or Summary/abstract.	Include other fields: Author, Institution, Department?
Service	Display <u>short description list</u> of matching records.	Alphabetic by title.
User	Select an entry from the list.	
Service	Display <u>full description</u> of selected record. Or Display resource linked to selected record.	

## Author browse

Service	Display alphabetic list of Author names.	Display A-Z pick?
User	Browse and select an entry from the list.	
Service	Display <u>short description list</u> of matching records.	Alphabetic by title?
User	Select an entry from the list.	
Service	Display <u>full description</u> of selected record. Or Display resource linked to selected record.	

### Author search

User	Input a single search string.	
Service	Match input <i>beginning-of-field</i> in Author name.	Post-truncation match.
Service	Display <u>short description list</u> of matching records.	Alphabetic by author, then title.
User	Select an entry from the list.	
Service	Display <u>full description</u> of selected record. Or Display resource linked to selected record.	

### Subject browse

Service	Display alphabetic list of subjects.	Display A-Z pick?
User	Browse and select an entry from the list.	
Service	Display <u>short description list</u> of matching records.	Alphabetic by title?
User	Select an entry from the list.	

Service      Display full description of selected record.  
Or  
Display resource linked to selected record.

### **Subject keyword search**

User          Input a single search string.

Service      Match input *anywhere-in-field* in Subject.      Full-truncation match.

Service      Display short description list of matching records.      Alphabetic by author, then title.

User          Select an entry from the list.

Service      Display full description of selected record.  
Or  
Display resource linked to selected record.

### **Title keyword search**

User          Input a single search string.

Service      Match input *anywhere-in-field* in Title.      Full-truncation match.

Service      Display short description list of matching records.      Alphabetic by author, then title.

User          Select an entry from the list.

Service      Display full description of selected record.  
Or  
Display resource linked to selected record.

### **Institution browse**

Service	Display alphabetic list of Institution names.	
User	Browse and select an entry from the list.	
Service	Display <u>short description list</u> of matching records.	Alphabetic by title?
User	Select an entry from the list.	
Service	Display <u>full description</u> of selected record.	
	Or	
	Display resource linked to selected record.	

**Institution search {see \*1 above}**

User	Input a single search string.	
Service	Match input <i>beginning-of-field</i> in Institution name.	Post-truncation match.
Service	Display <u>short description list</u> of matching records.	Alphabetic by institution, then title.
User	Select an entry from the list.	
Service	Display <u>full description</u> of selected record.	
	Or	
	Display resource linked to selected record.	

**Department browse {see \*2 above}**

Service	Display alphabetic list of Institution/Department names.	
User	Browse and select an entry from the list.	
Service	Display <u>short description list</u> of matching records.	Alphabetic by title?
User	Select an entry from the list.	

Service      Display full description of selected record.  
Or  
Display resource linked to selected record.

### **Department search {see \*2 above}**

User          Input a single search string.

Service      Match input *anywhere-in-field* in Department name.      Full-truncation match.

Service      Display short description list of matching records.      Alphabetic by institution, then title.

User          Select an entry from the list.

Service      Display full description of selected record.  
Or  
Display resource linked to selected record.

### **RAE Unit of Assessment browse {see \*3 above}**

Service      Display alphabetic list of RAE Unit names.

User          Browse and select an entry from the list.

Service      Display short description list of matching records.      Alphabetic by title?

User          Select an entry from the list.

Service      Display full description of selected record.  
Or  
Display resource linked to selected record.

### **RAE Unit of Assessment search {see \*3 above}**

User	Input a single search string.	
Service	Match input <i>anywhere-in-field</i> in RAE Unit name.	Full-truncation match.
Service	Display <u>short description list</u> of matching records.	Alphabetic by unit, then title.
User	Select an entry from the list.	
Service	Display <u>full description</u> of selected record. Or Display resource linked to selected record.	

## Appendix D: Some notes from the EThOS workshop, Glasgow University, 20 July 2006

A post-EThOS development, if approved by JISC and others, would be a theses digitisation and hosted resource repository service managed by the British Library. There would also be a metadata repository for hosted and remote e-theses. Both would utilise planned developments in the BL Digital Object Management system.

The resource repository would be populated by submission of copies of born-digital (or locally-digitised) theses by "member" HE institutions (including consortia), and by on-demand digitisation of paper and microfilm copies of theses already deposited in the BL's existing service.

The metadata repository would be populated by harvesting from local repositories, and by transforming existing records created and held by the BL. The BL has MARC21 catalogue records for the existing physical theses in its collection, and would integrate these with harvested or submitted metadata; whether by on-the-fly methods, or creating a mixed-format repository, or converting every record from MARC21 to DC is, I think, dependent on the DOMS module for integrating metadata. The BL will manage the metadata according to professional practice, including authority control.

The metadata repository will be harvestable; the resource repository will not.

The metadata repository proposal raises some implications for IRI-Scotland.

(1) It would be useful to be able to harvest metadata for e-theses awarded by Scottish institutions from the EThOS repository.

This is assuming that not all Scottish institutions will maintain a copy of the metadata in their local repositories (included any hosted or consortia repositories) which would subsequently be harvested by IRIScotland.

This might be the case where the theses has been digitised by the EThOS service rather than locally, or the scope of a local repository is limited by date or current research structure.

This would require the EThOS repository to facilitate selective harvesting on a institution basis. The best approach for the IRIScotland harvester would be to be able to select all Scottish institutions in one go (but see (3)). Failing this, the harvester could be set to harvest specific Scottish institutions one by one; this is not a generally-scalable approach, but is feasible for Scotland.

(2) The EThOS service might find it useful to harvest metadata from IRI-Scotland.

(a) This is assuming that not all Scottish institutions will participate in the EThOS resource repository.

This would require the IRI-Scotland metadata repository to facilitate selective harvesting on a format-of-content basis (specifically, thesis format) and by specific institution.

(b) All Scottish institutions decide to form a consortium to participate in EThOS.

This would be a mirror-image of the requirement identified in (1).

(3) Scottish institutions might find it useful to harvest metadata from EThOS.

This is assuming that an institution wants to maintain a full metadata repository for its research output, and wishes to avoid unnecessary duplication of effort in creating metadata which already exists in the BL.

This would require the EThOS repository to facilitate selective harvesting on a specific institution basis.

If the IRI-Scotland harvester is aware of which institutions meet case (3) and it harvests from EThOS by specific institution rather than an all-Scotland option, then it can avoid duplicate harvesting of metadata for the same e-thesis.

In all cases, the harvesting service would have to be able to de-duplicate records for the same resource, as overlap between local institution, IRI-Scotland, and EThOS repositories is inevitable.

For all cases, the issues of consistency, completeness and coherency of metadata harvested from multiple repositories identified by the HalRST and other projects apply.

E-theses do not currently fall within the scope of legal deposit, as they are considered unpublished. Thus the scenario of the NLS populating the IRI-Scotland hosted repository via this route does not seem to apply.

The National Library of Wales is a partner in EThOS, and will probably participate in any subsequent EThOS service based at the BL.

The EThOS project has also developed a metadata application profile for e-theses.

This may be subsumed by agreement on the ePrints application profile currently in development, as e-thesis falls with the definition of e-print.

If not, the effectiveness of the IRI-Scotland search/browse service may be impaired, assuming that the ePrints AP is integrated into any IRI-Scotland metadata agreement.

The ePrints AP itself may be too narrow in scope to meet IRI-Scotland requirements.

## Appendix E: Collection-level description and IRIScotland information retrieval

The Scottish Collections Network (SCONE) service is used to store collection-level descriptions (clds) for the IRIScotland information retrieval interface.

### SCONE

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. clds 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.

### Application to IRIScotland collections

Applying these to the harvested metadata for IRIScotland:

- 1) Add a cld to SCONE for the “IRIScotland catalogue”, the collection of all harvested metadata from IRIScotland partners’ research repositories.
- 2) Add a corresponding cld for the “IRIScotland collection”.
- 3) Add the “IRIScotland collection” Is-Described-By “IRIScotland catalogue” relationship.
- 4) Add similar clds and relationships for each of the harvested repositories.
- 5) If the harvested repository itself organises its metadata into sub-sets, add clds 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** (Electronic versions of research materials created by Scottish institutions.)
  - **Edinburgh research archive** (Research output from the University of Edinburgh, including full-text digital theses and dissertations, book chapters, journal pre-prints and peer-reviewed journal reprints.)
  - **Glasgow eprints service** (Published and peer-reviewed papers by members of the University of Glasgow.)
  - **St Andrews eprints** (An institutional archive for the academic community of St Andrews University.)

- **Strathprints** (Strathclyde University eprint repository.)
- **University of Stirling digital research repository** (Digital versions of research theses in full text (PhD and Masters).)
  - **Dept. of Applied Social Science eTheses** (Electronic copies of theses produced by students from the Department of Applied Social Science of the University of Stirling.)
  - **Dept. of Computing Science and Mathematics eTheses** (Electronic copies of theses produced by students from the Department of Computing Science and Mathematics of the University of Stirling.)
  - **Dept. of English Studies eTheses** (Electronic copies of theses produced by students from the Department of English Studies of the University of Stirling.)
  - **Dept. of Environmental Science eTheses** (Electronic copies of theses produced by students from the Department of Environmental Science of the University of Stirling.)
  - **Dept. of Film and Media eTheses** (Electronic copies of theses produced by students from the Department of Film and Media of the University of Stirling.)
  - **Dept. of History eTheses** (Electronic copies of theses produced by students from the Department of History of the University of Stirling.)
  - **Dept. of Management and Organization eTheses** (Electronic copies of theses produced by students from the Department of Management and Organization of the University of Stirling.)
  - **Dept. of Philosophy eTheses** (Electronic copies of theses produced by students from the Department of Philosophy of the University of Stirling.)
  - **Institute of Aquaculture eTheses** (Electronic copies of theses produced by students from the Institute of Aquaculture of the University of Stirling.)
  - **Institute of Education eTheses** (Electronic copies of theses produced by students from the Institute of Education of the University of Stirling.)
  - **School of Biological and Environmental Sciences eTheses** (Electronic copies of theses produced by students from the School of Biological and Environmental Sciences of the University of Stirling.)

The collection title is in bold, followed by the collection description (not the full cld).

At each bullet, there is an online information retrieval interface for the collection's metadata at the same level of granularity as the collection, with a link from the cld.

## 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** (Digital versions of research theses in full text (PhD and Masters).)

- **Dept. of Applied Social Science research repository** (Electronic copies of research papers produced by staff from the Department ...)
- **Dept. of Applied Social Science eTheses** (Electronic copies of theses produced by students from the Department of Applied Social Science of the University of Stirling.)
- **Dept. of Applied Social Science research papers** (Electronic copies of research papers produced by staff from the Department ...)
- **Etc.**
- **Etc.**

## General retrieval

The SCONE service (<http://scone.strath.ac.uk/Service/index.cfm>) allows retrieval of clds by several methods. Those that are relevant to digital collections and IRIScotland include:

- Collection title: Search for words in the title.
  - Allows the user to input a single text string, including embedded spaces to allow a phrase.
  - Lists the titles of collections containing the string anywhere in the title.
  - Example: "prints" retrieves **Glasgow eprints service, St Andrews eprints, Strathprints**.
- People and organizations: List names A-Z.
  - Allows the user to browse a list which can be limited to organization names and/or specific relationships.
  - Displays the titles of all collections related to a selected name.
  - Relationships include owners and subjects; both are created for IRIScotland institution names.
  - Example: "University of Edinburgh" retrieves Edinburgh research archive.
- People and organizations: Search for words in the name.
  - Allows the user to input a single text string, including embedded spaces to allow a phrase.
  - Lists names containing the string anywhere in the name.
  - Allows the user to browse the list which can be limited to organization names and/or specific relationships.
  - Displays the titles of all collections related to a selected name.
  - Example: "Stirling" retrieves University of Stirling digital research repository.
- Educational level.
  - Displays a list of standard education levels, including "Masters (SVQ 5)" and "Doctorates".
  - Displays the titles of all collections associated with a selected level.

- Example: "Masters (SVQ 5)" retrieves **University of Stirling digital research repository**, etc.

## 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

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

SCONE clds 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.

Example: "302.23" (DDC) and "Mass media" (LCHS) both retrieve **Dept. of Film and Media eTheses**

## Appendix F: Subject browsing issues

Searching 3267 records on 14 February 2006, harvested from four institutions (Edinburgh, Glasgow, St Andrews, Strathclyde), produced the following subject terms. Terms are listed in the order and layout that a user would see in a subject browse facility. The numbers indicate the frequency of occurrence of the subject term.

282 {no subject terms}  
1 2-Nitrophenol | Gas-to-liquid partitioning | Bubble column  
1 3-dehydroquinase | catalyze | enzymes | metabolic | DHQase  
1 74.25.Fy | 74.72.Jt | 74.25.Ha  
1 74.72.Jt | 74.62.Bf | 74.62.Dh  
1 Absence | Genome-wide | changes | DNA Methylation | development | zebrafish  
1 acoustic emission | deformation rate | permeability  
1 Air mass trajectories | Synoptic weather types | Black smoke | PM10  
1 air monitoring | chemical model | nitrogen oxides | passive sampler  
1 air pollution epidemiology | back trajectory analysis | metals  
1 Air | Vertical | Water | Flow | Tubes  
1 AMPA/kainate glutamate receptors | neural | hippocampus | memory  
1 anisotropic parameters | subsurface data | seismic data | P-wave | vertical transverse isotropy | S-wave | C-wave  
1 Antigenic variation, | genetic polymorphism, | nematodes | filariasis  
1 Antirrhinum majus | snapdragon | angiosperms | variation | inheritance | genetics | plant development | Arabidopsis thaliana  
1 Arabian Sea Summer Monsoon | Somali margin | Indian margin | Dansgaard-Oeschger cycles | Heinrich Events | climatic change  
1 Arf invariant | L-theory | Q-groups | UNil-groups  
1 Artificial Intelligence | Lerdaahl | Jackendoff | musical performance  
1 Asymmetric | leaves | mediates | leaf | patterning | stem cell | function | Arabidopsis  
1 Asynchronous Processor Architectures  
1 Atmospheric concentrations | deposition | trichloroacetic acid | Scotland | sampling  
1 ATOFMS | Mace Head | marine aerosols | Saharan dust | aerosol mass spectrometry  
1 Atomic Structure | Mycobacterium tuberculosis | CYP121 | Novel Features | Cytochrome P450  
1 Aurora kinases | cytokinesis | cell division | Cancer | cellular | geography  
1 automated reasoning | temporal logic | formal methods | proof theory  
1 AZ History of Scholarship The Humanities | DF Greece  
16 B Philosophy (General)  
2 B Philosophy (General) | BF Psychology  
1 B Philosophy (General) | BJ Ethics | BF Psychology  
1 B Philosophy (General) | JA Political science (General) | L Education (General)  
10 B Philosophy (General) | L Education (General)  
6 B Philosophy (General) | LB Theory and practice of education  
1 B Philosophy (General) | LB2300 Higher Education  
1 B Philosophy (General) | QA Mathematics  
1 B Philosophy (General) | RT Nursing  
1 B Philosophy (General) | ZA Information resources  
1 back trajectory analyses | pollution episode.  
1 Bacterial Transferrin | Dityrosyl motif | iron-binding protein | iron transport | oxo-iron cluster | x-ray crystallography  
1 Barthes | Travel | French | Literature | Tokyo | Japan  
1 basal ganglia | temporal processing  
4 BD Speculative Philosophy  
3 BF Psychology  
1 biofilm | coevolution | competition | iron | kin selection | symbiosis  
1 biogeochemical cycling | dissolved organic matter | particulate organic matter | dissolved inorganic nitrogen | estuarine mixing | Tyne Estuary | Tweed 2  
1 Bird- Meertens Formalism | Parallel computing  
1 BJ Ethics  
1 BJ Ethics | LB Theory and practice of education  
1 BL Religion | BD Speculative Philosophy  
1 borocarbides  
1 BR Christianity | PR English literature  
1 brain arteriovenous malformations | Lothian health board | Scotland | neuroscience

1 British Empire  
 1 Broccoli head rot | *Pseudomonas fluorescens*  
 1 Browsing intensity | Sheep | Cattle | Natural regeneration | *Betula*  
 1 C-waves | Multicomponent  
 1 CA1 | hippocampus | Marrī¿ çĩ3/4€ĩ3/4(tm)s model | perforant path | neocortex | subiculum  
 1 Calculus of Communicating Systems | verification | explicit proof plans  
 1 Caledonian pinewood ecosystems | Glen Affric | shade tolerance | tree seedlings  
 1 capability brokering | problem solving | agents  
 1 capillary electrochromatography | time-of-flight mass spectrometry | two-step laser mass spectrometry  
 1 Capillary | Electrochromatography | Ion | Trap | Accumulation | Time-of-flight | Mass | Spectrometry  
 11 CC Archaeology  
 1 CC Archaeology | DS Asia  
 1 CC Archaeology | GE Environmental Sciences  
 1 CD921 Archives  
 2 CD921 Archives | ZA4050 Electronic information resources  
 1 Cell | Signalling | Calcium | plant | cell nuclei  
 1 centromeres | heterochromatin | chromosome | vertebrate cells  
 1 Changing | Potency | spontaneous | fusion | stem cells  
 1 Channel Theory  
 1 Chromium Remediation or Release? | Iron(II) Sulfate Addition | Chromium (VI) Leaching | Columns | Chromite Ore | Processing Residue  
 1 circle valued Morse theory, | Novikov complex, | Novikov homology  
 1 Cognitive | apolipoprotein E | APOEε4 | allele | ageing | dementia | demented | Alzheimerī¿ çĩ3/4€ĩ3/4(tm)s disease  
 1 Coherence relations | Linguistics  
 1 Coinduction | proof rule | induction  
 1 colloidal lass  
 1 Combinatory Categorical Grammar  
 1 Computer Science | Observational Equivalence | Logical Relation  
 1 Conditional Reasoning  
 1 Controlled surgery | homology manifolds.  
 1 Cooperation | Competition | Pathogenic Bacteria  
 1 corporate social responsibility | financial performance | Cochrane systematic review | FTSE  
 1 cosmogenic | radionuclide | San Bernardino Mountains | denudation | cosmogenic 10Be | San Andreas Fault  
 1 counselling | feminisms | policy | UK | women  
 1 countable non-determinism | denotational semantics | domain equations | lambda-calculus  
 1 Cretaceous | Eratosthenes seamount | carbonates | Neotethys  
 1 cryptography | cipher | side-channel information  
 1 cryptography | cryptographic protocols  
 1 Culler | Postmodern | French | Derrida | Handelman | Levinas | Literary Theory | Mansfield  
 1 Cultural Geography  
 2 D051 Ancient History | CC Archaeology  
 1 D051 Ancient History | DA Great Britain | CC Archaeology  
 1 D204 Modern History | ZA Information resources | DA Great Britain  
 3 DA Great Britain | D839 Post-war History, 1945 on  
 1 Data Protection Act | United Kingdom s Medical Research Council  
 1 day students  
 1 deficit | b-amyloid | plaques | Alzheimer s  
 1 definite description  
 1 Deposition velocity | Flux divergence | Boundary layer | RO2 | HO2  
 1 diagrammatic proofs | mathematical reasoning  
 1 Diagrammatic Reasoning | Theorem Proving | Continuous Domain  
 1 direct | observation | DNA | translocation | cleavage | EcoKI endonuclease | atomic force | microscopy  
 1 discourse analysis | EFL | conversation | film | media | South Korea  
 1 discourse connectives | lexical acquisition | machine learning  
 1 DNA interactions | monofunctional organometallic ruthenium (II) | antitumour complexes | cell-free media  
 1 DNA | base pairs | EcoR124I | restriction | modification | enzyme | binding  
 1 DSpace | E-Theses | Tapir | EUL | Software | Digital Library  
 1 DSpace | ETD-db | Theses | Software | Evaluation  
 2 DU Oceania (South Seas)  
 1 e-Learning | Art Teaching | French Studies  
 1 e-theses | Edinburgh University  
 1 Eastern Mediterranean | Eratosthenes Seamount  
 1 Eastern Mediterranean, | Neotethys, | Late Cretaceous, | Early Tertiary, | basin analysis.  
 1 Ecological modelling | object-oriented programming | Ecology  
 1 EFL | CALL | CMC | Arab | Syria | learning styles | learning strategies

1 EFL | TESOL | motivation | Greece | teaching | teacher development  
1 Electron Transfer Protein | Multihaem | NMR | Paramagnetic shift  
1 electronic theses | intellectual property | copyright | retrospective digitization | patent law | database rights  
1 embryonic stem (ES) cells | Wnt | cDNA | Sfrp2  
1 Emotion | psychotherapy | relationality | symbolisation  
1 Enzyme activity | Exposure | Forest damage | Needles | Protein | soil  
1 Eratosthenes Seamount | Cyprus active margin  
1 Eratosthenes Seamount | shallow-marine sedimentary rocks  
1 Eratosthenes Seamount | Shallow-water limestones  
1 Eratosthenes Seamount | synthesis  
1 Eritrea | Globalisation  
1 Eritrea | Zimbabwe | Nationalism | Liberation | Guerrillas | Mobilization | State | post-colonial | Post-liberation | Politics  
1 Eritrea | Zimbabwe | Youth | Kalashnikov | National Service | Post-Liberation | Students | Conscription | University of Asmara  
1 Ethiopia | Ethiopian State | Ethnic identity | Ethnicity  
1 Exposure | Sitka spruce seedlings | trichloroacetic acid  
1 Fault currents | load flow analysis | optimization methods | power generation planning  
1 felspar | diagenesis | sandstone | secondary porosity | compaction  
1 filariasis, | antigens, | inhibitors, | cytokine | homologies, | vaccine  
1 Finance | hydroelectric power generation | hydrology | meteorology | risk analysis  
1 Flooding | South Downs  
1 fluoroalkyltin | F-doped tin oxide | CVD | thin film  
1 Forest | tropical montane | Chiapas | Mexico  
1 formal argumentation | artificial intelligence | knowledge engineering  
1 French | Baudelaire | Literature | Poetry | Internet | Hypertext  
1 French | Marker | Identity | Narrative | Film | Mansfield | Internet | Cold War  
1 Function | Synthesis | Ribosomes  
1 functional heads  
1 G Geography (General)  
2 GB Physical geography  
1 GB Physical geography | GC Oceanography  
1 GB Physical geography | HC Economic History and Conditions  
2 GC Oceanography  
12 GE Environmental Sciences  
1 genome sequence | filamentous fungus | Neurospora crassa  
1 Geomagnetism | Geomagnetically Induced Currents | Magnetotellurics | Magnetovariation | Electromagnetic Induction | Electromagnetic Modelling  
1 GF Human ecology. Anthropogeography  
1 GF Human ecology. Anthropogeography | B Philosophy (General)  
1 GF Human ecology. Anthropogeography | HC Economic History and Conditions  
1 Glutamate-receptor-mediated | encoding | retrieval | paired-associate learning  
1 GN Anthropology | NB Sculpture  
1 gradience | grammaticality | linguistic judgements | linguistic data | optimality theory | magnitude estimation | computational modeling  
1 Greece | language transfer | grammar | EFL | interlanguage | verb tense  
2 GV Recreation Leisure  
1 GV Recreation Leisure | QP Physiology  
1 GV Recreation Leisure | TJ Mechanical engineering and machinery  
1 H Social Sciences (General)  
1 HA Statistics  
1 HA Statistics | QA Mathematics  
1 HB Economic Theory  
1 HB Economic Theory | GE Environmental Sciences  
2 HC Economic History and Conditions  
2 HD Industries. Land use. Labor  
1 HD Industries. Land use. Labor | HB Economic Theory  
2 HD Industries. Land use. Labor | HC Economic History and Conditions  
2 HD Industries. Land use. Labor | HD28 Management. Industrial Management  
1 HD Industries. Land use. Labor | HN Social history and conditions. Social problems. Social reform | RA0421 Public health. Hygiene. Preventive Medicine  
1 HD Industries. Land use. Labor | HV Social pathology. Social and public welfare  
1 HD Industries. Land use. Labor | LC Special aspects of education  
1 HD Industries. Land use. Labor | RA0421 Public health. Hygiene. Preventive Medicine  
1 Helicase | Messenger RNA | Oskar s message | Translation | Drosophila melanogaster  
1 Helium | volcanic rocks | isotope | olivine | phenocrysts | Baffin Island | proto-Iceland plume | 3He/4He ratios | geochemistry

1 Helminth | Cytokine | Cystatin | Serpin | Vaccine  
1 herd | genetically | uniform | cattle | Deleterious | alleles | feral strain | inbred | cows  
3 HF Commerce  
30 HF5601 Accounting  
3 HG Finance  
1 HG Finance | HB Economic Theory  
1 Highrise | urban development  
1 HJ Public Finance  
2 HJ Public Finance | HB Economic Theory  
2 HM Sociology  
1 HM Sociology | G Geography (General)  
11 HQ The family. Marriage. Woman  
1 HQ The family. Marriage. Woman | HC Economic History and Conditions  
1 HQ The family. Marriage. Woman | LC Special aspects of education  
5 HT Communities. Classes. Races  
1 HT Communities. Classes. Races | DA Great Britain | HC Economic History and Conditions  
1 HT Communities. Classes. Races | HC Economic History and Conditions  
1 HT Communities. Classes. Races | HD28 Management. Industrial Management  
1 HT Communities. Classes. Races | LC5201 Education extension. Adult education. Continuing education  
1 HT Communities. Classes. Races | RA0421 Public health. Hygiene. Preventive Medicine  
1 HT Communities. Classes. Races | RA0421 Public health. Hygiene. Preventive Medicine | RJ101 Child Health. Child health services  
1 human cognitive architecture | skill acquisition | meta-cognition  
1 Hungarian | syntax | semantics | pragmatics | focus  
2 HV Social pathology. Social and public welfare  
1 HV Social pathology. Social and public welfare | HF5601 Accounting  
1 HV Social pathology. Social and public welfare | HQ The family. Marriage. Woman  
1 HV Social pathology. Social and public welfare | LB Theory and practice of education  
1 HV Social pathology. Social and public welfare | LC Special aspects of education  
1 HV Social pathology. Social and public welfare | RA0421 Public health. Hygiene. Preventive Medicine | LC Special aspects of education  
1 HV Social pathology. Social and public welfare | RC0321 Neuroscience. Biological psychiatry. Neuropsychiatry  
1 HV Social pathology. Social and public welfare | RJ101 Child Health. Child health services  
1 HX Socialism. Communism. Anarchism | LB Theory and practice of education  
1 ice sheet advance | Late Pleistocene | continental margin | Late Devensian  
1 immunology | Antigen | Maintenance | Immunological | Memory  
1 India | Scotland | TESOL | education | multiculturalism  
1 Induction Rules | Proof Planning  
1 inflammation  
1 information and communication technologies | micro and small enterprises | Tanzania | Africa  
1 Information services | Skills training | Tutorials | Feedback  
1 Information storage | Supramolecular surface patterns  
1 information theory | expected partition entropy | Koller-Sahami  
2 Institutional Repositories | Digital Library | Librarianship  
1 Institutional Repository | Software  
1 intelligent control | autonomous mobile robots  
1 Internet  
1 internet resources | neurology  
1 Internet | Medical resources  
1 Isoprene | volatile organic compound | tropospheric chemistry | photosynthesis | Quercus virginiana  
1 Iterative Compilation | Performance Prediction | Numerical Applications  
1 JA Political science (General) | L Education (General)  
1 JN1187 Scotland  
1 JQ Political institutions Asia  
4 JZ International relations  
1 JZ International relations | LB Theory and practice of education  
1 Kinetics | micelle | Aqueous lecithin-bile salt mixtures | vesicle  
1 Knowledge Acquisition | Databases  
1 Korea | motivation | attitude | culture | monolingual | learning goals  
1 Kyoto | Vladimir Putin | Russia | Climate change  
6 L Education (General)  
1 L Education (General) | LB Theory and practice of education  
1 L Education (General) | LB2300 Higher Education  
1 language production | syntactic structure  
1 Latrunculin  
37 LB Theory and practice of education

1 LB Theory and practice of education | LC Special aspects of education  
1 LB1501 Primary Education  
2 LB1603 Secondary Education. High schools  
10 LB2300 Higher Education  
1 LB2300 Higher Education | LB Theory and practice of education  
3 LB2361 Curriculum  
1 LB2361 Curriculum | HV Social pathology. Social and public welfare  
1 LB2361 Curriculum | LC5201 Education extension. Adult education. Continuing education  
25 LC Special aspects of education  
5 LC5201 Education extension. Adult education. Continuing education  
2 LC5201 Education extension. Adult education. Continuing education | LB2300 Higher Education  
1 LEGO | type system  
1 Lepidoptera | phenylacetaldehyde | a-pinene | wind tunnel | volatiles | insect | insect | preference | feeding  
1 Leukemia inhibitory factor | cytokine receptor | signaling | ES cells | tetracycline | episome  
1 Lexical Generalisations | Inheritance Hierarchy Construction  
12 LF Individual institutions (Europe)  
2 LF Individual institutions (Europe) | Z719 Libraries (General)  
1 logic programming  
1 logic | control | Prolog  
1 Lord Jesus Christ | Devotion | Early Christianity  
1 Low-temperature geochemistry | Geochemistry  
1 Machine humour | puns  
1 macroinvertebrate | litter decomposition | tropical | montane | cloud forest | Mexico  
1 Magnetic cluster  
1 malaria | gametocyte | plasmodium | apicomplexa | local mate competition  
1 Malaria | transmission-blocking | vaccines  
1 Management | Construction | Systems | Industry | Expert | Training  
1 Marine energy | wave energy | wind energy | wind climate | wave climate | climate change  
1 MBD2 | transcriptional | repressor | MeCP1 histone | deacetylase complex | Mammalian DNA | methylation  
1 Mechanics | vibration | beams  
1 memory | maintenance | protein kinase C | Drosophila  
1 Mendelian disorders | oligogenic disorders | gene  
1 Messinian | Eratosthenes Seamount | Mediterranean  
1 methylated component | Neurospora crassa | genome  
1 MHC | body odors | pleasantness  
1 MHC | body odours  
1 Microtubules | Schizosaccharomyces pombe | tea1p | mod5p  
1 Middle French | Christine de Pizan | XML | XSLT | Image Files  
1 Mineral | Protein | transferrin  
1 ML Literature of music  
1 ML Literature of music | LB Theory and practice of education  
1 Monolithic | Frequency | Filter Designs | Wave  
1 Montreal | Canada | Travel | Literature | Quebec  
1 Mouse | Mecp2-null | Mutation | neurological | symptoms | Rett syndrome  
1 MT Musical instruction and study | LB Theory and practice of education  
1 Mud volcanism | Mediterranean Ridge | accretionary prism  
1 mud volcanoes | Mediterranean Ridge  
1 Multiple Processors | Compilation Techniques | High-Performance Embedded Systems  
16 N Visual arts (General) For photography, see TR  
1 N Visual arts (General) For photography, see TR | BH Aesthetics | NX Arts in general | BV1460 Religious Education | BR Christianity  
1 N Visual arts (General) For photography, see TR | NX Arts in general | HF Commerce  
1 NA Architecture | NK Decorative arts Applied arts Decoration and ornament  
1 NA Architecture | Z719 Libraries (General)  
1 natural language | punctuation  
1 Natural Language Processing | thesauri  
1 natural language | speech | dialogue initiative problem  
1 NB Sculpture  
1 network atmospheric data | primary aerosol | secondary aerosol  
1 Neurology | publication  
1 noncommutative localization, | K-theory, | L-theory, | triangulated category.  
1 nursing studies | stroke | young  
1 NX Arts in general | TR Photography | LC Special aspects of education  
1 Oaxaca | Mexico | Nitrogen Cycling | tree - grass inter-cropping | Carbon Cycling | silvopastoral system  
1 Oil Shale bings | West Lothian | Scotland | biodiversity  
1 Open-Learning | ODL | eLearning | JavaScript | English | Poetry  
1 P52 | nomina sacra

1 Pakistan | refugee | NGO  
 1 Palaeoceanographic | genetic variation | North Atlantic | Neogloboquadrina | pachyderma | Foraminifera  
 1 parallel computers | Bird-Meertens Formalism | topology  
 1 parallel computing | Bulk Synchronous Parallel model | Skeletal Parallel Programs  
 1 Part-based Grouping | Recognition  
 1 Partitioning | Temperature dependence | Aqueous solubility  
 1 Pathogens | bovine tuberculosis | badger culling | tuberculosis | British cattle | Meles mele  
 2 PB1501 Gaelic (Scottish Gaelic, Erse)  
 1 PB1501 Gaelic (Scottish Gaelic, Erse) | PB Modern European Languages  
 1 PE English  
 1 Perception | Motor control | Golf  
 1 perception | sensory mechanisms  
 1 PETROGRAPHY | VOLCANICLASTIC SAND | SANDSTONE | WOODLARK RIFT BASIN  
 1 phase diagrams | hard spheres | van der waals | Gibbs-Duhem  
 1 Phase Mapping | Phase Switch | Lattice Switch | Simulated Tempering | Weighted Histogram Analysis Method | Fast Growth | Jarzynski method | Umbrella | Multi-stage | Multicanonical | Path Integral Monte Carlo | Path Sampling | fluctuation theorem | equilibrium | non-equilibrium | statistical mechanics | condensed matter | computational | crystal | feynman  
 1 Phenotypic Complementation | Oct-3/4 | Embryonic Stem Cells | POU  
 1 PI | postinfection | I-line | immune-selected line | naive-selected line | N-line  
 1 PIOSPHERES | PLANT-HERBIVORE INTERACTIONS | SEMI-ARID RANGELAND  
 1 Pliocene | sediments | tectonics | Turkey | Pleistocene  
 1 pluripotency | teratocarcinoma | Nanog | Oct4 | STAT3 | GATA6  
 1 PN0080 Criticism | PR English literature  
 1 Poetry  
 1 Polarised light | butterfly | mating signal  
 1 Polymorphism | recursive types | full abstraction | ideals | metric models  
 1 PQ Romance literatures  
 6 PR English literature  
 1 principal | Plasmodium falciparum | human | immunity | malaria | molecular | population | genetic  
 1 Programming paradigms | Information types | graphical representation | program comprehension  
 1 proof planning  
 1 Proteomic | receptor-adhesion | protein | complexes | NDMA | N-methyl-D-aspartate  
 1 PS American literature  
 1 psychological plausibility | proof theory | semantics  
 1 PUBLICATION BIAS, EVOLUTION, BEHAVIOR, ECOLOGY, KINSHIP, METAANALYSIS, SELECTION, TRIM, FILL  
 1 Q Science (General)  
 5 Q Science (General) | GE Environmental Sciences  
 6 QA Mathematics  
 17 QA75 Electronic computers. Computer science  
 2 QA75 Electronic computers. Computer science | P Philology. Linguistics | KD England and Wales  
 3 QA75 Electronic computers. Computer science | TJ Mechanical engineering and machinery  
 1 QA75 Electronic computers. Computer science | TP Chemical technology  
 2 QA75 Electronic computers. Computer science | Z665 Library Science. Information Science  
 179 QA76 Computer software  
 1 QA76 Computer software | KDC Scotland  
 1 QA76 Computer software | TA Engineering (General). Civil engineering (General) | ZA4050 Electronic information resources  
 1 QA76 Computer software | TK Electrical engineering. Electronics Nuclear engineering  
 17 QB Astronomy  
 55 QB Astronomy | QC Physics  
 1 QB Astronomy | QE Geology  
 766 QC Physics  
 1 QC Physics | CC Archaeology  
 2 QC Physics | GE Environmental Sciences  
 2 QC Physics | LB Theory and practice of education  
 1 QC Physics | Q Science (General)  
 4 QC Physics | QD Chemistry  
 1 QC Physics | QD Chemistry | QM Human anatomy  
 2 QC Physics | QE Geology  
 1 QC Physics | QH Natural history  
 3 QC Physics | QP Physiology  
 5 QC Physics | RC0254 Neoplasms. Tumors. Oncology (including Cancer)  
 1 QC Physics | RJ Pediatrics | QP Physiology  
 1 QC Physics | TA Engineering (General). Civil engineering (General)  
 3 QC Physics | TJ Mechanical engineering and machinery

1 QC Physics | TK Electrical engineering. Electronics Nuclear engineering  
58 QD Chemistry  
3 QD Chemistry | CC Archaeology  
1 QD Chemistry | GC Oceanography  
3 QD Chemistry | GE Environmental Sciences  
5 QD Chemistry | QE Geology  
1 QD Chemistry | QH Natural history  
6 QD Chemistry | QH345 Biochemistry  
1 QD Chemistry | RS Pharmacy and materia medica  
1 QD Chemistry | S Agriculture (General)  
1 QD Chemistry | TP Chemical technology  
175 QE Geology  
8 QE Geology | CC Archaeology  
1 QE Geology | GB Physical geography  
2 QE Geology | GC Oceanography  
3 QE Geology | GE Environmental Sciences  
11 QH Natural history  
1 QH Natural history | GE Environmental Sciences  
8 QH301 Biology  
1 QH301 Biology | G Geography (General)  
2 QH301 Biology | GC Oceanography  
1 QH301 Biology | GF Human ecology. Anthropogeography  
2 QH301 Biology | QA Mathematics  
2 QH301 Biology | QD Chemistry  
1 QH301 Biology | QH345 Biochemistry  
3 QH301 Biology | QL Zoology  
1 QH301 Biology | QL Zoology | QH Natural history  
1 QH301 Biology | QP Physiology  
1 QH301 Biology | RA Public aspects of medicine  
19 QH345 Biochemistry  
1 QH345 Biochemistry | RM Therapeutics. Pharmacology  
10 QH426 Genetics  
1 QH426 Genetics | QH Natural history  
2 QH426 Genetics | QP Physiology  
19 QK Botany  
3 QK Botany | QD Chemistry  
1 QK Botany | TK Electrical engineering. Electronics Nuclear engineering | GE Environmental Sciences  
25 QL Zoology  
2 QL Zoology | QD Chemistry  
1 QL Zoology | QH Natural history | QA Mathematics  
1 QL Zoology | QH345 Biochemistry  
6 QL Zoology | QH426 Genetics  
1 QL Zoology | TJ Mechanical engineering and machinery  
2 QM Human anatomy | TJ Mechanical engineering and machinery  
7 QP Physiology  
1 QR Microbiology  
1 QR Microbiology | G Geography (General)  
1 QR Microbiology | GE Environmental Sciences  
1 QR Microbiology | RC Internal medicine  
2 QR Microbiology | RK Dentistry  
1 QR Microbiology | RK Dentistry | RA0421 Public health. Hygiene. Preventive Medicine  
4 QR180 Immunology  
8 QR180 Immunology | RC Internal medicine  
1 QR180 Immunology | RC Internal medicine | RM Therapeutics. Pharmacology  
1 QR180 Immunology | RG Gynecology and obstetrics  
1 QR180 Immunology | RG Gynecology and obstetrics | QP Physiology  
1 QR355 Virology  
1 Quantum | Physics | Altered | States  
28 R Medicine (General)  
1 R Medicine (General) | QP Physiology  
2 R Medicine (General) | RC Internal medicine  
1 R Medicine (General) | RC Internal medicine | QH426 Genetics  
1 R Medicine (General) | RC0254 Neoplasms. Tumors. Oncology (including Cancer)  
1 R Medicine (General) | RC0321 Neuroscience. Biological psychiatry. Neuropsychiatry  
1 R Medicine (General) | TA Engineering (General). Civil engineering (General)  
3 R Medicine (General) | TJ Mechanical engineering and machinery  
10 RA Public aspects of medicine

1 RA Public aspects of medicine | HN Social history and conditions. Social problems. Social reform  
1 RA Public aspects of medicine | QP Physiology  
1 RA Public aspects of medicine | R Medicine (General)  
1 RA Public aspects of medicine | RC Internal medicine  
1 RA Public aspects of medicine | RJ Pediatrics | QP Physiology  
12 RA0421 Public health. Hygiene. Preventive Medicine  
1 RA0421 Public health. Hygiene. Preventive Medicine | HQ The family. Marriage. Woman  
1 RA0421 Public health. Hygiene. Preventive Medicine | HQ The family. Marriage. Woman | RG Gynecology and obstetrics  
3 RA0421 Public health. Hygiene. Preventive Medicine | LC Special aspects of education  
1 RA0421 Public health. Hygiene. Preventive Medicine | QR355 Virology  
1 RA1001 Forensic Medicine. Medical jurisprudence. Legal medicine | QC Physics  
3 RB Pathology  
8 RB Pathology | QH426 Genetics  
2 RB Pathology | QP Physiology  
1 RB Pathology | RC Internal medicine  
1 RB Pathology | RC Internal medicine | QH426 Genetics  
69 RC Internal medicine  
3 RC Internal medicine | QH426 Genetics  
1 RC Internal medicine | QM Human anatomy  
22 RC Internal medicine | QP Physiology  
7 RC Internal medicine | RA0421 Public health. Hygiene. Preventive Medicine  
1 RC Internal medicine | RA0421 Public health. Hygiene. Preventive Medicine | QR355 Virology  
1 RC Internal medicine | RA0421 Public health. Hygiene. Preventive Medicine | RM Therapeutics. Pharmacology  
3 RC Internal medicine | RD Surgery  
2 RC Internal medicine | RG Gynecology and obstetrics | QP Physiology  
8 RC Internal medicine | RM Therapeutics. Pharmacology  
1 RC Internal medicine | RM Therapeutics. Pharmacology | QP Physiology  
1 RC Internal medicine | RM Therapeutics. Pharmacology | RG Gynecology and obstetrics  
1 RC Internal medicine | RT Nursing  
28 RC0254 Neoplasms. Tumors. Oncology (including Cancer)  
24 RC0254 Neoplasms. Tumors. Oncology (including Cancer) | QH426 Genetics  
1 RC0254 Neoplasms. Tumors. Oncology (including Cancer) | QH426 Genetics | RG Gynecology and obstetrics  
2 RC0254 Neoplasms. Tumors. Oncology (including Cancer) | QP Physiology  
1 RC0254 Neoplasms. Tumors. Oncology (including Cancer) | RA0421 Public health. Hygiene. Preventive Medicine  
26 RC0254 Neoplasms. Tumors. Oncology (including Cancer) | RB Pathology  
3 RC0254 Neoplasms. Tumors. Oncology (including Cancer) | RB Pathology | QH426 Genetics  
1 RC0254 Neoplasms. Tumors. Oncology (including Cancer) | RD Surgery  
1 RC0254 Neoplasms. Tumors. Oncology (including Cancer) | RK Dentistry | RA0421 Public health. Hygiene. Preventive Medicine  
3 RC0321 Neuroscience. Biological psychiatry. Neuropsychiatry  
1 RC0321 Neuroscience. Biological psychiatry. Neuropsychiatry | LB Theory and practice of education  
2 RC0321 Neuroscience. Biological psychiatry. Neuropsychiatry | QH426 Genetics  
1 RC0321 Neuroscience. Biological psychiatry. Neuropsychiatry | RC Internal medicine | QH426 Genetics  
1 RC1200 Sports Medicine | QP Physiology  
2 RD Surgery  
1 Reduced | Adaptation | non-recombining | neo-Y | chromosome | Drosophila miranda  
1 Reference resolution | natural language | anaphora  
1 Reinforcement Learning | Continuous Action Spaces | Markov Decision Process  
1 representation preservation | re-analysis  
1 representations of algebras, | noncommutative localization.  
1 Restricted | Growth | Schwann | Cells | Cajal | Bands | Slows | Conduction | Myelinated | Nerves  
1 RETT-SYNDROME | DNA  
1 RG Gynecology and obstetrics  
4 RG Gynecology and obstetrics | QP Physiology  
1 Risk Management | Projects  
4 RJ Pediatrics  
3 RJ Pediatrics | RC Internal medicine  
2 RJ Pediatrics | RJ101 Child Health. Child health services  
1 RJ Pediatrics | RK Dentistry | RD Surgery  
3 RJ101 Child Health. Child health services  
224 RK Dentistry  
6 RK Dentistry | RA0421 Public health. Hygiene. Preventive Medicine  
3 RK Dentistry | RC Internal medicine

3 RK Dentistry | RD Surgery  
 1 RK Dentistry | RM Therapeutics. Pharmacology  
 4 RM Therapeutics. Pharmacology  
 1 RNA | eukaryotes | mRNA metabolism | mRNA decapping | Lsm protein | enzyme  
 1 Roleplay | Story Writing  
 1 RT Nursing  
 1 Ruminant | Physiology | resource use  
 2 ruthenocuprates | ferromagnetic | superconductor  
 1 RX Homeopathy  
 5 S Agriculture (General)  
 1 S Agriculture (General) | QH Natural history  
 1 Scott domain theory | Real PCF  
 1 Search Optimisation | Evolutionary | Neighbourhood  
 1 Self-sufficient Cytochrome P450 | Primary Structural Organization | Flavin-Domain  
 1 semantic theory | syntax | natural language  
 1 Sex | Chlamydomonas reinhardtii | speed limit | evolution  
 2 SF Animal culture  
 1 SF Animal culture | QD Chemistry  
 1 SF Animal culture | TJ Mechanical engineering and machinery  
 4 SF600 Veterinary Medicine  
 4 SF600 Veterinary Medicine | QL Zoology  
 1 SF600 Veterinary Medicine | QR355 Virology  
 2 SF600 Veterinary Medicine | RC Internal medicine  
 1 SH Aquaculture. Fisheries. Angling  
 1 SH Aquaculture. Fisheries. Angling | QL Zoology  
 1 shale bings | vegetation | West Lothian | floristic data | biodiversity  
 1 Signal processing | transduction | plant | cells  
 1 Simultaneous Abstraction Calculus | lamda-calculus | Dynamic Montague Grammar | Discourse Representation Theory  
 1 Situation Semantics | computational linguistics  
 1 Situational Context | Language Production | Natural language  
 1 Sloan Digital Sky Survey | star formation | Galaxy spectra  
 1 SML.NET compiler | MIL | Moggi's  $\lambda$ -calculus | Girard-Tait reducibility | Typed Functional Programming Languages  
 1 Soay sheep | Ovis aries | St Kilda | sperm | reproduction  
 1 social scientific research | questionnaire | street interviews  
 1 software design | state machines | UML | games  
 1 soil carbon fluxes | forestry  
 1 South Africa | Cuba | health care | South-South development cooperation | aid | rural health care  
 1 South Africa | Cuba | NEPAD | Health Care  
 1 spectral interpretation  
 1 Speech | dialogue  
 1 Stable oxygen isotope | Eratosthenes Seamount | surface-water conditions | Eastern Mediterranean  
 1 strike-slip | fault  
 1 STRONTIUM ISOTOPIC AGE | LIMESTONES | WOODLARK RIFT BASIN | OCEAN DRILLING PROGRAM LEG 180  
 1 Structural | mechanistic | mapping | unique | fumarate | reductase  
 1 structural | organisation | exosome | complex | RNA | degradation  
 1 Sub-Basalt Imaging | Boundary Element Method  
 1 successional trees | proximity signals | photosynthesis  
 1 sugars | proteins | biological processes | Microarray techniques  
 1 Supercoiled DNA | liquid crystal  
 1 Superscalar processors | complex processors  
 1 Supervised Learning | Genetic Programming  
 1 surgery exact sequence | structure set | total surgery obstruction  
 1 surgery, | quadratic form, | chain complex, | Poincare duality  
 1 synapses | dendrites | central nervous system | axon | hypothalamic oxytocin neurons | soma  
 1 synapses | synaptic | terminals | skeletal | muscle  
 1 Synaptic vesicle endocytosis | cyclindependent kinase 5  
 1 Synergistic action | Metal Binding | Ferric Ion-Binding Protein | Neisseria Gonorrhoeae  
 1 syntax | semantics | pragmatics | semantic knowledge  
 1 Syria | personality | learning styles  
 3 T Technology (General)  
 30 TA Engineering (General). Civil engineering (General)  
 1 TA Engineering (General). Civil engineering (General) | QH Natural history  
 1 TA Engineering (General). Civil engineering (General) | T201 Patents. Trademarks  
 3 TA Engineering (General). Civil engineering (General) | TJ Mechanical engineering and machinery

1 TCA | Uptake | Elimination | Needles | Throughfall | Stemflow | Litterfall  
15 TD Environmental technology. Sanitary engineering  
1 TD Environmental technology. Sanitary engineering | GC Oceanography  
1 TD Environmental technology. Sanitary engineering | QD Chemistry  
1 TD Environmental technology. Sanitary engineering | QD Chemistry | QE Geology  
1 TD Environmental technology. Sanitary engineering | QR Microbiology | QE Geology  
1 TD Environmental technology. Sanitary engineering | RA Public aspects of medicine | QC Physics | QD Chemistry  
1 TD Environmental technology. Sanitary engineering | S Agriculture (General)  
2 TD Environmental technology. Sanitary engineering | TK Electrical engineering. Electronics Nuclear engineering  
1 Tectonics | Turkey | Neotethys | Orogenesis | Eastern Mediterranean | Tectonostratigraphy  
1 TESOL | EFL | literature | poetry | Taiwan  
1 Tethys Sea | ophiolite | Mesogaea | Syria  
2 TH Building construction | TJ Mechanical engineering and machinery  
1 the temporal coherence | narrative discourse.  
1 Theses Alive! | Sherpa  
1 thymic epithelium | thymus | organogenesis | embryonic | origin | cortex | medulla | epithelium  
1 tin oxide | CVD | fluorocarboxylate | X-ray structure | electron diffraction  
42 TJ Mechanical engineering and machinery  
5 TJ Mechanical engineering and machinery | QA Mathematics  
8 TJ Mechanical engineering and machinery | QP Physiology  
1 TJ Mechanical engineering and machinery | TK Electrical engineering. Electronics Nuclear engineering  
1 TJ Mechanical engineering and machinery | TL Motor vehicles. Aeronautics. Astronautics  
10 TK Electrical engineering. Electronics Nuclear engineering  
1 TK Electrical engineering. Electronics Nuclear engineering | DS Asia  
2 TK Electrical engineering. Electronics Nuclear engineering | QP Physiology  
26 TL Motor vehicles. Aeronautics. Astronautics  
1 TL Motor vehicles. Aeronautics. Astronautics | RC1200 Sports Medicine  
2 TL Motor vehicles. Aeronautics. Astronautics | TJ Mechanical engineering and machinery  
2 TN Mining engineering. Metallurgy  
3 TN Mining engineering. Metallurgy | QE Geology  
1 Toxocara; | Nematode; | Mucin | Glycoprotein  
6 TP Chemical technology  
1 TP Chemical technology | TK Electrical engineering. Electronics Nuclear engineering  
1 transversality | chain complex | CW complex  
8 TS Manufactures  
1 TS Manufactures | QH Natural history  
1 TS Manufactures | TA Engineering (General). Civil engineering (General) | LB2300 Higher Education  
2 TS Manufactures | TP Chemical technology  
1 turbulence | Renormalization Group | Navier-Stokes equations | fluid flow  
1 Turkey | Lycian Taurus | thrust sheets | extension | orogeny | Eastern mediterranean  
1 Turkey | Neotethys | tectonics | thrust sheets, | orogeny  
1 Turkish grammar | Head-driven Phrase Structure Grammar  
1 type inference | Bounded Quantification  
1 type theory | Extended Calculus of Constructions | deliverables  
1 typed lambda calculi | polynomial time algorithms  
1 typed lambda calculus | formal systems | proof checking | interactive theorem proving  
1 Uneven-Aged | Forestry | Forest Operations  
1 Unified Transformation Framework | heuristic random search algorithm  
1 Urban | gender | emotion  
1 Vaccines | pathogen | virulence | malaria  
1 vegetation dynamics | Hill Sheep and Native Woodland project | Grazing  
1 Vegetation | Logging Disturbance | Gabon | Rain Forest | Mammals | Elephants | Apes  
1 Vestiges | DNA methylation | System | Drosophila melanogaster  
1 Vigorous Exchange | Indian Ocean | Atlantic Ocean | Oceanography | Glacial Periods | Agulhas  
1 virtual environment  
1 VM Naval architecture. Shipbuilding. Marine engineering  
1 VOLCANOGENIC | TERRIGENOUS | HEMIPELAGIC SEDIMENTS | WOODLARK BACKARC RIFT BASIN | PROVENANCE | GEOCHEMISTRY  
1 volumetric models | 3-D objects | Radical Basis Function | Kalman Filter technique | Interpretation Tree algorithm | geons  
1 Voluntary sector counselling | Scotland  
1 Wallerian | degeneration | axons | synapses | Ube4b/Nmnat | chimeric gene  
1 web | poetry | French | internet | hypertext | Barthes  
1 Welding fumes | Transition metals | Inflammation | Oxidative stress  
1 worm-like micelle | stopped flow | mixed micelle | structure

3 Z004 Books. Writing. Paleography | HF5601 Accounting  
4 Z665 Library Science. Information Science  
1 Z665 Library Science. Information Science | ZA4050 Electronic information resources  
8 ZA Information resources  
1 ZA Information resources | DA Great Britain  
1 ZA Information resources | GA Mathematical geography. Cartography  
9 ZA4050 Electronic information resources  
1 ZA4450 Databases | RK Dentistry  
1 ZA4450 Databases | ZA4050 Electronic information resources  
1 Zimbabwe | elections  
1 Zimbabwe | NGOs | Africa  
1 Zimbabwe | NGOs | civil society | Africa | structural adjustment | PVO Act  
1 Zola | Baudelaire | Balzac | French | 19C | 19th | Mansfield

## Appendix G: Comparison of IRIScotland and DRIVER Dublin Core metadata guidelines

### Brief

The following report investigates the key differences and similarities between the DRIVER (Digital Repository Infrastructure Vision for European Research) guidelines for the creation of Dublin Core metadata elements and those of IRIScotland, in order to identify any key issues which might complicate the OAI metadata harvesting of IRIS\_DC records by DRIVER.

### Metadata element analysis

#### DC Element Variable: Creator

Driver	Iris	Both
<p>Gives further instructions for construction of <i>inverted name</i> format</p> <p>Advises inclusion of names (personal or corporate) from local/national thesauri is optional</p>	<p>dc.contributor.author to be avoided as duplicates</p> <p>dc.creator</p>	<p>dc.creator should be used unqualified for the main author(s) of a resource</p> <p>name to be given in browse (inverted name) format</p> <p>dc.creator is repeatable, and a separate field must be used for each author</p>

#### Analysis

While there are minor differences in the specificity of usage instructions/recommendations, these are not contradictory and so the mapping of IRIS\_DC metadata records to DRIVER\_DC would not be adversely affected.

#### DC Element Variable: Title

Driver	Iris	Both
<p>"If necessary, repeat this element for multiple titles"</p>	<p>Not repeatable, if necessary use dc.title.alternate (sparingly)</p> <p>Advice to retain leading definite and indefinite articles as they appear in the Title (The, A, An)</p> <p>Advises punctuation not to be used</p> <p>Advises not to use non-standard characters</p>	<p>Subtitles to be appended to main title, separated by space-colon-space</p> <p>Capitalise only first word and proper nouns</p>

#### Analysis

In addition to minor differences in the specificity of usage instructions/recommendations, a major difference seemingly exists in the instructions for the repeatability of **dc.title**. While DRIVER advises that it can be repeated "if necessary", IRIS advises it is "not repeatable", but

codifies this by advising that "if necessary" the qualifier **dc.title.alternate** can be used "sparingly". However, since the **alternative** qualifier should be stripped at the time of harvesting by DRIVER, the IRIS **dc.title** field should still be wholly compatible with the DRIVER schema. Therefore, this apparent difference should not cause any conflict.

### DC Element Variable: Subject

Driver	Iris	Both
Mandatory when applicable  Advises no requirement for capitalisation of keywords, but that collections should be uniform	Should be used unqualified if terms are not controlled or a local vocabulary is used	Subject element can be either keyword or classification (using standard qualifiers to encode where applicable)

#### Analysis

While there are minor differences in the specificity of usage instructions/recommendations, these are not contradictory and so the mapping of IRIS\_DC metadata records to DRIVER\_DC would not be adversely affected.

### DC Element Variable: Description

Driver	Iris	Both
Do not use for description of location (i.e. URLs)	Makes clear problems related to cutting and pasting of text  Advises that is not repeatable (DRIVER makes no such recommendation)	Mandatory where applicable

#### Analysis

While other minor differences exist in the specificity of usage instructions/recommendations, the only potential conflict is in the repeatability of this element. While IRIS\_DC makes clear that this element is not *repeatable*, DRIVER does not specify whether it either is or is not repeatable. However, in any case, should DRIVER\_DC allow the repetition of this element, it would not be of issue for the harvesting of metadata from IRIS to DRIVER. Conflict would arise, should DRIVER allow such repetition, only in the event that an OAI\_PMH project sought to aggregate an IRIS\_DC metadata through the harvesting of DRIVER\_DC records, the reverse of the instance being investigated here.

### DC Element Variable: Publisher

Driver	Iris	Both
"Mandatory where applicable"  Gives recommendations for the citation of University publications (i.e. hierarchical, from largest to smallest)	"May be omitted"	

#### Analysis

The instructions given for the usage of this element differ slightly between IRIS\_DC and DRIVER\_DC; however the instructions that, in DRIVER it is "mandatory where applicable"

and in IRIS "may be omitted" amount to essentially the same meaning, but with a differing degree of emphasis of importance. Nonetheless, this slight difference would not negatively affect the harvesting of IRIS\_DC metadata to DRIVER\_DC. No other differences or issues have been identified.

### DC Element Variable: Contributor

Driver	Iris	Both
Gives no instructions for qualification		May be used unqualified

#### Analysis

No differences or issues have been identified.

### DC Element Variable: Date

Driver	Iris	Both
	Advises on use of qualifiers	Mandatory Use of ISO 8601 – Year is mandatory : month, day are optional

#### Analysis

Here the only identifiable difference is that while IRIS\_DC advises on the use of qualifiers, DRIVER\_DC does not.

### DC Element Variable: Type

Driver	Iris	Both
Gives instructions for use of the following exact text: Article Book Conference lecture Conference report Contribution for newspaper or weekly magazine Doctoral thesis Master thesis Bachelor thesis External research report Lecture Internal report	"Recommends" use of <i>Eprints type vocabulary scheme</i> : Book Book Item Book Review Conference Item Conference Paper Conference Poster Journal Article Journal Item News Item Patent Report	Mandatory To be used unqualified

Newsletter	Scholarly Text	
Part of book or chapter of book	Submitted Journal Article	
Research paper	Thesis or Dissertation	
	Working or Discussion Paper	

### Analysis

The instructions for use of this element between the two schemas, while both defining it as both mandatory and only to be used unqualified, differ in the vocabulary they identify to be used. While IRIS\_DC "recommends" the Eprints type scheme, DRIVER identifies an exact vocabulary which must be followed exactly. This is, potentially, a major difference: although it is wholly possible DRIVER might be able to map non-standard descriptors into its own schema, it might be recommended that IRIS\_DC should adopt the DRIVER vocabulary to avoid such potential conflict.

There might also exist an issue over the repeatability of this field. While IRIS explicitly states "**dc.type** is repeatable", DRIVER\_DC usage instructions state: "In every metadata record one DC element 'type' should be used", which seems to infer that no repetition is possible. Such a difference, should it be the case, might result in conflict in the harvesting of IRIS\_DC metadata by DRIVER.

### DC Element Variable: Format

Driver	Iris	Both
Specifies scheme "IANA registered list of Internet Media Types"	"Any guidance offered by the eprints application profile should be adopted"	
Specifies to format by type/subtype		

### Analysis

In specifying the IANA registered list of Internet Media Types to be used, DRIVER\_DC requires the use of this scheme; although it is wholly possible DRIVER might be able to map non-standard descriptors into its own schema, it might be recommended that IRIS\_DC should adopt the DRIVER vocabulary to avoid such potential conflict.

### DC Element Variable: Identifier

Driver	Iris	Both
No significant issues, both schemes identify the use of this field to provide a formal means of identifying the item described. The URI of the resource is given as the recommended use of this field.		

### Analysis

No differences or issues have been identified.

### DC Element Variable: Source

Driver	Iris	Both
Optional use	"Should not be used" –	

Reference source through string or number corresponding to formal identification system (e.g., URI)	instead use dc.description	
---	----------------------------	--

### Analysis

While IRIS\_DC advises this field should not be used, that DRIVER makes its use optional would not cause conflict when harvesting from IRIS to DRIVER. No other issues are identified.

### DC Element Variable: Language

Driver	Iris	Both
Use recommended Repetition allowed for documents in more than one language Scheme: ISO 639-3	"may be used ... but may be omitted for English language resources"  "dc.language is not repeatable"  Scheme: ISO 639-2	

### Analysis

While in DRIVER, use is "recommended", in IRIS, it "may be omitted for English language resources". Given that DRIVER is a proposed Europe-wide infrastructure, it might be recommended that IRIS\_DC records should include the identification of English language resources. Also while DRIVER allows repetition, IRIS advises "dc.language is not repeatable"; however, this would not cause conflict should DRIVER harvest IRIS\_DC records. Finally, a difference exists in the International Standard each identifies for use, DRIVER prescribing ISO 639-3 and IRIS ISO 639-2.

### DC Element Variable: Relation

Driver	Iris	Both
"can be used to indicate different kinds of relations between metadata records"	"primary use of dc.relation should be for recording the publication title, volume and issue number for journal articles and conference papers"	

### Analysis

There may exist a difference in interpretation of the use of this element; however this should not cause conflict in the harvesting of IRIS\_DC by DRIVER.

### DC Element Variable: Coverage

Driver	Iris	Both
Can also use for jurisdiction (e.g. Named administrative entity)	"dc.coverage should always be qualified with either .spatial or .temporal"	Used for temporal/spatial information

No instructions for qualification		
-----------------------------------	--	--

### Analysis

While IRIS advises the use of this element, qualified, for spatial/temporal information, DRIVER also advises it can be used for information regarding jurisdiction, and gives no instructions for qualification. Again this should not cause conflict in the harvesting of IRIS\_DC by DRIVER.

### DC Element Variable: Rights

Driver	Iris	Both
No differences		

### Analysis

No differences or issues have been identified.

### DC Element Variable: Audience

Driver	Iris	Both
"A class of entity for whom the resource is intended or useful"	Not used	

### Analysis

DRIVER also identifies the use of this sixteenth (i.e. non-simple core) DC element and gives directions for use while IRIS does not identify this element. As in previous cases, however, this should not cause conflict in the harvesting of IRIS\_DC by DRIVER.

## Summative Analysis

In general, IRISScotland is more specific in its recommendations/instructions for use; as such, were IRIS\_DC metadata records to be harvested by DRIVER this would result in the loss of some of this specificity (such as the dropping of some qualifiers not recognised, etc); however, while this would not be desirable, neither should it adversely affect the harvested metadata which will, in most instances, remain as complete as other records created specifically according to the DRIVER\_DC recommendations/instructions.

However, in the two instances of the recommendation/instruction for use of differing controlled vocabularies (i.e. occurring in **dc.type** and **dc.format**) which have been identified, there is a potential for conflict. It might be recommended, therefore, that IRIS\_DC prefer to recommend the use of the controlled vocabularies identified for use by DRIVER.

In the identification of the possible difference over the repeatability of **dc.type** a conflict may exist.

The differing instructions as to the necessity of the **dc.language** element should be noted. In IRIS\_DC the recommendation is that English language resources need not be identified as such, potentially conflicting with DRIVER's instructions. Obviously, in Scottish institutional repositories the vast majority of items are likely to be in English, this will not, however, be the case in a Europe-wide metadata aggregation such as DRIVER intends. Therefore, it may be identified that this recommendation for the exclusion of such language information be removed.

## Appendix H: Technical configuration

The following hardware and software is currently used for the pilot service.

### Hardware

All software is currently running on PCs running Windows 2000.

- The ARC package and MySQL database run on a development server.
- ColdFusion and MS Internet Information Services run on a live web server.
- MS Access database runs on a desktop PC.

### Software

#### ARC harvester

This is a component of the ARC harvester and search engine<sup>29</sup>. It carries out the actual harvesting of metadata records from repositories, using OAI-PMH, and stores the harvested records in a MySQL database that is supplied along with the ARC software. Version 0.95 of ARC is currently being used, although version 1.0 is now available.

#### ARC parser

This component of the ARC package processes the metadata held in XML format as a single field in the harvested records into a separate table holding 15 fields, corresponding to the 15 metadata elements specified in the simple Dublin Core format which is mandatory for compliance with OAI-PMH. The table also includes a few additional fields to record the item identifier, repository name, status, processing date etc.

It is possible to use an alternative parser alongside the ARC harvester, but at present the ARC parser is still in use.

#### MySQL

A MySQL database is used to store the harvested and parsed metadata. The database implementation is supplied as part of the ARC package, although MySQL itself must be installed separately.

#### MySQL Control Center

This is a graphical user interface to MySQL that is useful (though not essential) for configuring the ARC harvester, for monitoring the success of harvesting, and for interrogating and manipulating the database of harvested records.

#### ColdFusion

Adobe ColdFusion MX Standard version 6.1 is currently used to retrieve information from the MySQL database and to provide the user interface for the pilot service.

ColdFusion is more powerful and flexible than the website interface provided with the ARC package that was used for the HaIRST Experimental discovery service.

---

<sup>29</sup> ARC harvester and search engine. Available at: <http://sourceforge.net/projects/oaiarc/>

## **Microsoft Access**

MS Access is used as a convenient means of carrying out some cleaning and manipulation on the metadata held in the MySQL database.

An Access module (written by CDLR) can extract metadata from the MySQL database, manipulate it and then update the database with the modified metadata. This is particularly useful for ensuring that the Identifier field contains a valid URL, and can also be used for the substitution of certain non-standard characters, such as smart quotes and long dashes, with standard equivalents.