

# Scottish Distributed Information Infrastructure Research: Scottish Cultural Portal Pilot and Public Libraries Integration Initiatives

## 1. Background and Overview

This proposal, agreed in outline at an informal meeting with SLIC, is a request for funding to conduct basic research into distributed information infrastructure requirements for the two year Scottish Cultural Portal Pilot, and the one year Public Library CAIRNS Integration proposal, and, where appropriate, to develop associated pilot facilities. The aim would be to ensure that the infrastructure proposals would be sufficiently generic to be applicable to other Scottish information portals that might be considered in future, whether by SLIC, the CDLR itself, or others in the Scottish cultural information community. Some pilot facilities are also envisaged. The proposal builds on SLIC's own proposals for seamless access to learning, research, and cultural resources in Scotland for all citizens, and on an earlier survey conducted by the CDLR on behalf of SLIC (see Appendix A below). A key element of the work will be compliance with accepted and developing standards in the area in order to ensure Scotland-wide, UK-wide, and World-wide inter-operability in respect of distributed infrastructure elements (see section 13 below). The bid also encompasses a charge for two days a week technical advice and training covering the set up, development and ongoing maintenance of the Cultural Portal interface and database.

## 2. Specific Research and Development Threads and Deliverables

Specific project objectives are as follows:

1. Determination of requirements and working mechanisms for the seamless integration of the Cultural Portal Pilot into CAIRNS and related initiatives such as the SCONE collections facility, the collection strengths landscaper, and the SCAMP updates facility in such a way as to ensure that the Cultural Portal Pilot can act as a front end to these facilities and that they can also act as a front end to it, thereby offering two access routes into the Cultural Portal Pilot.
2. Research leading to a specification for the development of a pilot cultural facets (from heritage to ballet to football to street theatre) landscaping facility, and implementation of a working pilot as part of the pilot portal. Associated guidelines on resource creation and metadata.
3. Preparation of a specification for the development of a pilot user-type landscaping facility, and implementation of a working pilot as part of the portal. Associated guidelines on resource creation and metadata to allow the presentation of different versions of documents for different user levels (e.g. primary school children, life long learners, researchers, tourists)
4. A pilot Cultural Portal Pilot 'interoperability focus' that aims to ensure interoperability within the portal environment at all levels, including, in particular technical standards such as Z39.50, metadata standards and terminologies
5. Research into, and preparation of, a specification for a pilot terminologies server and implementation of illustrative facilities. This work would draw on the work of the HILT Project's proposed Terminologies Route Map or TeRM server (see appendix B for outline functional sketch).
6. Determination of guidelines to ensure optimal interoperability between the Cultural Portal Pilot and other relevant facilities such as SCRAP
7. Specification for, and development of, extension of SCAMP pilot facilities to facilitate a multi-domain mechanism for professionals involved in building and sustaining cultural 'collections' to collaborate online
8. Research into, and design of, alternative 'look and feel' themes that might be used in the portal for consideration by the Cultural Portal Advisory Group
9. Research into, and design of, alternative 'look and feel' themes for associated infrastructure elements in the CAIRNS, SCONE and SCAMP pilots
10. Research and development work to permit facilities in the co-operative infrastructure like CAIRNS, SCONE and SCAMP to serve as central facilities and data sources, not only for the Cultural Portal Pilot but for other similar portals in future
11. Formation and co-ordination of a COSMIC-based cross-sectoral and cross-domain cultural professionals group to guide the pilot portal; associated 'action research' activities to help elicit data on the best organisational support structures

12. Establish and report on the best mechanisms for exploiting the CORC shared cataloguing facility for collaborative working and standards assurance in the Cultural Portal Pilot.
13. Research leading to guidelines to ensure international interoperability, for example with the CDLR's US Portal partners, 'LOOK'
14. Identification of digital and non-digital collections relevant to the Cultural Portal for SCONE
15. Technical advice and training covering the set up, development and ongoing maintenance of the Cultural Portal interface and database.
16. Research project to extend the CAIRNS pilot to FE and Public Libraries
17. Determination of requirements and working mechanisms for the seamless integration of a significant number of Public Library catalogues into CAIRNS and the web enabling of the catalogues in question (likely to entail a single process)
18. Similar integration of Public Libraries into related initiatives such as the SCONE collections facility, the collection strengths landscaper, and the SCAMP updates facility.
19. Determination of requirements and of working mechanisms to allow any proposed Public Library portals to utilise data held centrally in CAIRNS and SCONE and to encompass, where possible, Cultural Portal elements in a similar way
20. Extension of the CAIRNS cataloguing standards for use in Public Libraries.
21. Specification and pilot facilities to enable CAIRNS to offer dynamic landscaping facilities suited to Public Library requirements
22. A pilot Public Libraries 'interoperability focus' that aims to ensure interoperability at all levels, including, in particular technical standards such as Z39.50, metadata standards and terminologies
23. Determination of Public Library terminologies server requirements
24. Specification for, and development of, extension of SCAMP pilot facilities for Public Libraries
25. Experimental formation and co-ordination of a COSMIC-based cross-sectoral and cross-domain public libraries professionals group to guide development within the distributed infrastructure
26. Establish and report on the best mechanisms for exploiting the CORC shared cataloguing facility for collaborative working and standards assurance in Public Libraries, with particular reference to the cataloguing of internet resources

### **3. Project Plan and Schedule**

A full and detailed Project Plan will be prepared in the first three months of the project and will show both the Cultural Portal Pilot and Public Libraries Integration elements as separate but inter-related and sometimes overlapping project threads. SLIC will be invited to evaluate and agree this detailed plan before it is finalised. Once agreed, the plan will guide project activity and scheduling.

### **4. Project length and proposed start date**

The project duration will be 18 months, commencing January 2003 if funding details are completed in time, or as soon after that as possible if not. The Cultural Portal Pilot infrastructure research will run from January 2003 to June 2004. The Public Libraries infrastructure research will overlap with this, running from January 2003 to December 2003.

### **5. Total Costs**

The total cost of the project over the two years will be £xxxxx, this sum primarily entailing staff costs. The breakdown is shown at 11 below.

### **6. The project team**

The project team will draw on time and expertise from current staff experienced in these areas, including Dennis Nicholson, Gordon Dunsire, and Emma McCulloch. However, additional staff will be taken on to supplement this expertise.

## 7. Relationship to other initiatives

The programme will be co-ordinated with other relevant initiatives such as Cultural Portal Project itself, the HILT High Level Thesaurus Project, the CAIRNS distributed catalogue, the SCONE Scottish Collections Database, HaIRST and CC-Interop.

## 8. Methodology

In line with the 'whole environment' approach taken to the issues entailed in distributed environments by the CDLR [1], the aim will be to stagger the start dates of the various key processes (see under 2 above) but to then allow them to proceed in parallel over most of the timescale of the project so that interim results coming out of any one process can inform developments in the others. Specific details on methodology will be included in the Project Plan (see 3 above).

## 9. The Centre for Digital Library Research

The Centre for Digital Library Research (CDLR) was formed in August 1999, with the aim of bringing together long-standing University research interests in the digital information area previously spread across two University departments: the University's Directorate of Information Strategy, which includes the Library and the Centre for Educational Systems, and the Department of Information Science. Its primary aim is to conduct quality academic research but with a practical focus. The CDLR is involved with a number of projects in the area of distributed information system infrastructure and has accumulated expertise in the areas of interoperability, distributed systems, Z39.50, collection description and management and standards via the CAIRNS, SCONE, SEED, CC-Interop, HaIRST and HILT<sup>1</sup> projects.

## 10. Project management

Everyday operations will be the responsibility of the project staff and the Project Director. This Project Team will report to a Project Management Group as described at 1.5 above. In addition, there will be a representative Steering Group encompassing senior staff across Scottish cultural information scene. At an early stage, the Project team will consult with SLIC on the membership of the group.

## 11. Costs

### Total Outline Budget: Staffing, Travel and Sundries, Software and Hardware<sup>2</sup>

	Costs
Researcher: Cultural Portal Aspect (18 months @ xxxxx p.a. mid range 2A)	£xxxxx
Researcher: Public Libraries Aspect (12 months @ xxxxx p.a. top end 1B)	£xxxxx
CP Technical advisor and developer (18 months @ 2 days a week)	£xxxxx
Hardware and software upgrade costs	£xxxxx
Travel and sundries (including dissemination, telephone etc)	£xxxxx
<b>Total</b>	<b>£xxxxx</b>

## 12. Accessibility

The project will ensure that deliverables will conform with the Disability Discrimination Act and Human Rights legislation regarding accessibility and the wider access policies of the Funding Councils. Web interfaces will conform to W3C current guidelines (<http://www.w3.org/WAI/Resources/#gl>).

## 13. Standards

Throughout the proposed project, CDLR will take due note of the various actual and emerging standards that are appropriate to the work, such as Z39.50, MARC21, the Bath Profile, the e-gif interoperability framework,

<sup>1</sup> See further information under specific projects at <http://cdlr.strath.ac.uk/projects/projects.html>

<sup>2</sup> Note: These are approximate figures. Small amounts may be transferred between funds should this be required

various Collection Description Schemas such as the UKOLN CLD Scheme, the CAIRNS cataloguing and indexing standards, NOF Technical Standards, and the developing JISC IE standards. Where doubt arises in any given case, the project team will consult SLIC on the preferred approach.

**14. Dissemination**

Dissemination of information will be via a project web-site and through the web-sites of related projects such as are detailed in section 7 above. In addition, postings will be made to appropriate e-mail lists, including the cross-domain scotslink list. Deliverables such as guidelines, reports and project outcomes will be available via the website, and the site will be linked to the service interface. Papers and news items will be submitted to professional publications and presentations at seminars and conferences will be made. Key progress reports would be sent to all relevant organisations and institutions in Scotland and elsewhere. Special regional workshops will be held to help involve local staff in decision making and other project processes. Leaflets will be produced under the CoSMIC banner on key issues. An attempt will be made to identify contacts in every relevant Scottish library or similar institution and close contact with these staff should ensure ongoing dissemination to libraries and similar organisations across Scotland

**15. Exit Strategy**

The final report and recommendations will indicate requirements for the support of a future robust distributed infrastructure entailing all necessary centralised elements but focussing as much as is practical on a decentralised approach based on local support where this is feasible and the most sustainable solution.

**16. Key contact**

Dennis Nicholson (Project Director)  
Director,  
Centre for Digital Library Research, University of Strathclyde, 101 St James Road, Glasgow G4 0NS.  
Phone: 0141 548 2102 FAX: 0141 548 2102 [d.m.nicholson@strath.ac.uk](mailto:d.m.nicholson@strath.ac.uk)

**References**

Law, D., Nicholson, D., Dunsire, G. Whole Environment Research on Distributed and Collaborative Digital and Non-digital Networked Libraries in Scotland. in *Bibliothek.: Forschung und Praxis*, V.26, 2002, Nr.2

## Appendix A: Results of a Survey of FE and Public Libraries Conducted in 2001

	FE College Libraries	Public Libraries
Q1: Can your staff access your library catalogue electronically?	85.71% - yes 14.29% - no	95.83% - yes 4.17% - no
Q2: Can your users access your library catalogue electronically?	85.71% - yes 14.29% - no	95.83% - yes 4.17% - no
Q3: Does your library catalogue have a URL?	42.86% - yes 57.14% - no	41.67% - yes 58.33% - no
Q4: If you have answered 'no' so far, do you have any future plans for an automated/online catalogue?	93.33% - yes	100%
Q5: Is your library catalogue Z39.50 compliant?	60.71% - yes 39.29% - no	70.83% - yes 29.17% - no
Q6: Which company supplies your library system?	See Appendix 4 for details	See Appendix 4 for details
Q7: Do you know what your supplier would charge to implement Z39.50 access to your catalogue?	See Appendix 4 for details	See Appendix 4 for details
Q8: Are any named special collections held by you or any branch libraries?	28.57% - yes 71.43% - no	75% - yes 25% - no
Q9: Does your library list or catalogue Internet resources for the benefit of your users?	82.14% - yes 17.86% - no	41.67% - yes 58.33% - no
Q10: Would you be interested in becoming involved in a cooperative approach to cataloguing free quality Internet resources?	85.71% - yes 14.29% - no	83.33% - yes 16.67% - no
Q11: Would your library find it useful to be able to conduct online cross searches of other library catalogues in Scotland and elsewhere?	100% - yes	100% - yes
Q12: Would your library find it useful to be able to conduct an online search of a database of Scottish special or local collections?	82.14% - yes 10.71% - possibly 7.14% - no	100% - yes

## Appendix B: HILT Phase II Discussion Document: Initial Service Specification Draft

	Process	Notes
	User enters system at TeRM server or at some other site that uses TeRM data	TeRM is built around an SQL compliant database using the ORACLE RDBMS
	▼	
Task not subject related, so go elsewhere	◀ User specifies task from drop down list	The Copac/clumps project, cc-interop will begin to identify tasks
	▼	
	Task is subject related	Obvious example is user says she wants to do a search by subject, but there may be other tasks that imply a subject search
	▼	
	User is prompted for subject terms	
	▼	
	TeRM interaction disambiguates subject and maps to DDC	In Wordmap user is given optional meanings of terms and asked to pick which she means (e.g. lotus, the flower, the software, the car, or...)
	▼	
	Task and DDC used to search CLD and CS collections database (a clone of SCONE for the pilot) that also contains service details	Implies a SCONE-like database, but with (1) DDC based collection strengths added, where they exist, for both special collections and general collections, with the latter mapped for strength below the general level where data exists, and (2) ▼
	▼	
	Collections service identifies sub-group of collections relevant to task and to subject using DDC number, truncating it to find collections	(2) Task relevance specified. As in SCONE, the top level of any collection hierarchy will, if it is an online service, have connection details attached. Services returned will be ranked according to the granularity level of ▼ their subject strength coding, with those with lower granularity levels ranked highest

	described at higher level of granularity	
	▼	
	User 'trims' list of collections to suit her purpose, deselecting those she decides not to use initially	User may spot services not relevant to her or simply wish to search less than the full list. Option to test results from highest ranked service may be available
	▼	
	Collections service details include subject scheme(s) used in each collection, and connection details	Specifying subject schemes necessary so that only appropriate terms are sent to any specific service, thereby avoiding the false drops that will occur if all terms from all schemes are sent to all services
	▼	
	Further interaction with TeRM based on 'trimmed' list of services provides terminologies set needed to search for the user's subject in each chosen service, including standard terms from scheme(s) and (possibly) common UK alternative terms or service specific alternative terms	User clicks button which says something like 'get terms', TeRM sends back appropriate terms for each service. Different levels of service are possible, beginning with terms from standard unchanged scheme used by service, to this plus common UK alternatives, with or without standard terms, to service specific alternatives. The addition and maintenance of UK alternatives will raise costs, and the addition of service specific terms will raise costs even further. Costs against benefits of each approach will be measured by HILT and compared against alternative approaches such as clustering
	▼	
	User gets option of switching off all but one of the various schemes used by a given service and the user's own terms before searching	User will have the option to use her own terms, plus the standard term, plus UK or service alternatives for each service, or to switch any of these alternatives off.
	▼	
	User searches services, either singly or in groups, using preferred term sets	This is assumed to be a broadcast search using Z39.50 plus, where facilities exist, other alternative protocols.