ROADS to Metadata

by Debra Hiom[*]

Abstract
ROADS (Resource Organisation and Discovery in Subject-based Services) is a UK Higher Education funded project to design and implement a user oriented resource discovery system. The project is investigating the creation, collection and distribution of resource descriptions to provide a transparent means of searching for, and using resources on the Internet. The system is being piloted on a number of Internet subject gateways, namely ADAM (Art, Design, Architecture and Media), Biz/ed (Business Education on the Internet), IHR-Info (Institute of Historical Research), OMNI (Organising Medical Networked Information) and SOSIG (Social Science Information Gateway). The paper will discuss the background to the project, the type of metadata being collected by the subject-based gateways and the possibilities of cross searching distributed databases, with specific references to SOSIG.

The project uses a standard template for recording information about resources (this was originally based on the Internet Anonymous FTP Archive (IAFA) Template) which is a simple text based record using attribute-value pairs. The simplicity of the ROADS format also provides possibilities of mapping to and exchanging data with other metadata formats, for example the Dublin Core or other standards.

One of the aims of the subject based gateways is to encourage information providers to become involved in the creation of records about their own data in order to make their information as useful and accessible as possible; an approach to this will be discussed.

Background
ROADS[1] (Resource Organisation and Discovery in Subject-based Services) is a collaborative project funded by the Electronic Libraries Programme[2] (eLib) in the UK, to design and implement a user oriented resource discovery system. The ROADS partners are:

- ILRT (Institute for Learning and Research Technology) at the University of Bristol - responsible for user liaison and project management
- Loughborough University (Department of Computer Science) - responsible for the software development
- UKOLN (Office of Library and Information Networking) at the University of Bath - responsible for co-ordinating metadata requirements and issues

ROADS has been created to provide a set of software tools and standards for building and maintaining catalogues of Internet resources. The system allows resources to be catalogued and indexed and provides a searchable and browsable interface to the resource descriptions. The ROADS system is primarily being piloted on a number of eLib funded subject information gateways (under the Access to Network Resources (ANR) programme) who feed into the development of the software. The gateways using ROADS are:

- ADAM (Art, Design, Architecture and Media)[3]
- Biz/ed (Business Education on the Internet)[4]
- IHR-Info (Institute of Historical Research)[5]
- OMNI (Organising Medical Networked Information)[6]
- SOSIG (Social Science Information Gateway)[7]

The system is being used or evaluated by a number of other projects in the UK and interest in its use has also been expressed outside the UK. In addition, ROADS (along with SOSIG) is involved in the EU funded DESIRE[8] project (under the Fourth Framework Telematics Programme).

Each of the eLib ANR gateways is building a subject specific catalogue of Internet resource descriptions. The ROADS software attempts to be as modular as possible to allow the gateways to configure the 'look and feel' of the services; for example, in the way they present browsable listings of resources and search results. It also allows the gateways to pick and choose parts of the system appropriate to their service and 'plug in' their own applications; SOSIG plans to use this capability to add a thesaurus tool to the standard ROADS search facility.

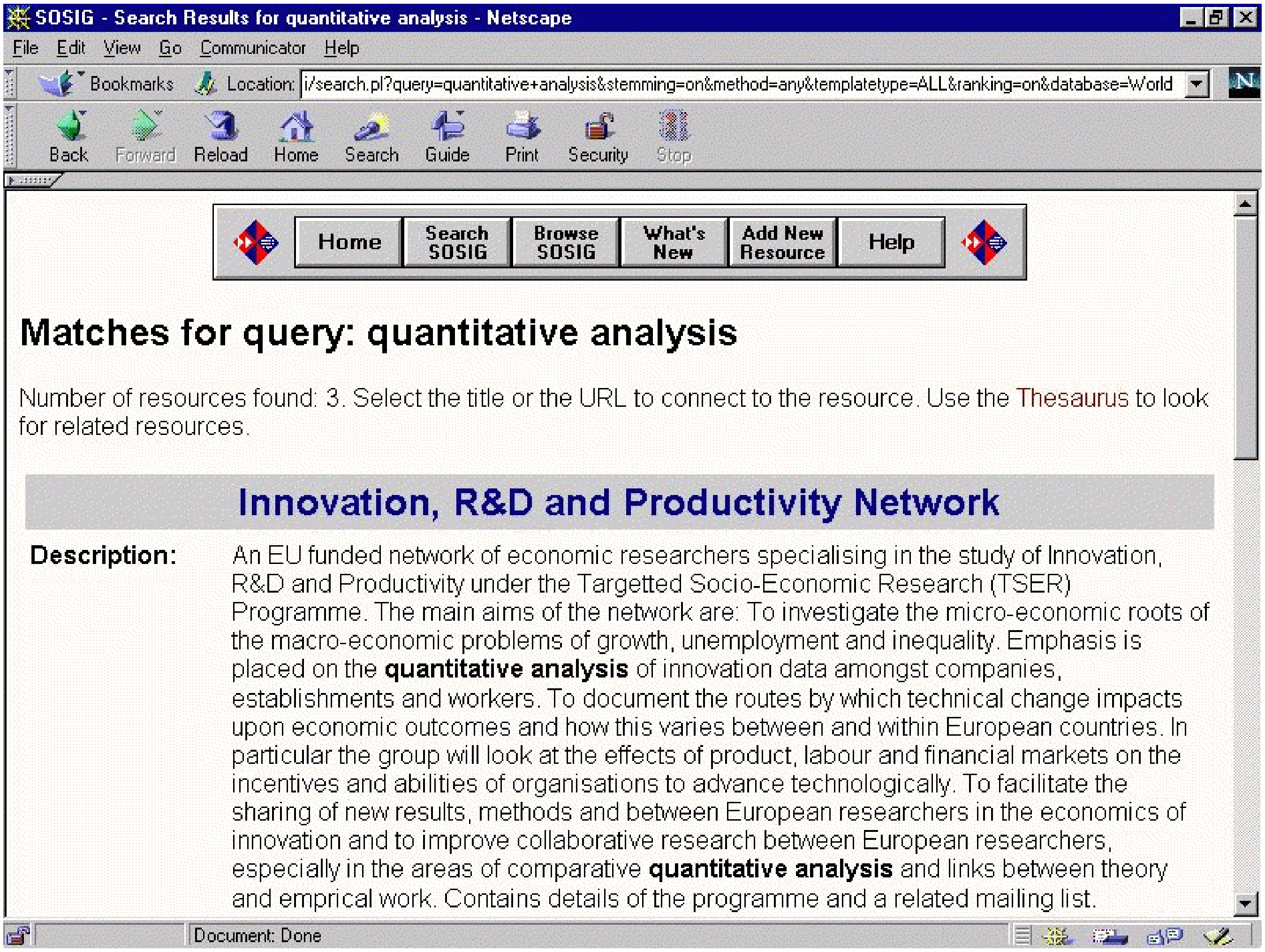
Each service may also differ on issues of selection policy, classification, etc. However, the gateways share a common metadata format for collecting information and end-users will ultimately be able to cross-search the gateways (using the WHOIS++ directory technology).

ROADS Templates

In addition to software development, the ROADS project is concerned with issues of metadata. The eLib ANR gateways are creating records for selected quality resources on the Internet using a standard template. The metadata format used by ROADS is based on the Internet Anonymous FTP Archive (IAFA) Template definitions[9]; which as the name suggests were originally designed to describe resources available through FTP sites. With the growth of the Web, these templates were extended to cover Internet resources in general. The templates have been extended further by ROADS based on the implementation experiences of the subject gateways (therefore the templates will be referred to throughout the paper as ROADS templates rather than IAFA). The template format was originally designed to be created by site administrators and therefore the emphasis is on simplicity and ease of creation. This simplicity also means that information skills are not essential and subject gateways can use the expertise of subject specialists as well library and information professionals to build their catalogues.

The template is a text-based record composed of a series of attribute-value pairs that describe a resource content, format and location. A number of different resource description template types exist:

- DOCUMENT
- IMAGE
- MAILARCHIVE
- ORGANIZATION
- PROJECT
- SERVICE
- SOFTWARE
- SOUND
- USENET
- USER


SOSIG - Search Results for quantitative analysis - Netscape
File Edit View Go Communicator Help
Bookmarks
Location: i/search.pl?query=quantitative+analysis&stemming=on&method=any&templatetype=ALL&ranking=on&database=World
Back
Forward
Reload
Home
Search
Guide
Print
Security
Stop
Home
Search SOSIG
Browse SOSIG
What's New
Add New Resource
Help
Document: Done

Matches for query: quantitative analysis

Number of resources found: 3. Select the title or the URL to connect to the resource. Use the Thesaurus to look for related resources.

Innovation, R&D and Productivity Network

Description: An EU funded network of economic researchers specialising in the study of Innovation, R&D and Productivity under the Targetted Socio-Economic Research (TSER) Programme. The main aims of the network are: To investigate the micro-economic roots of the macro-economic problems of growth, unemployment and inequality. Emphasis is placed on the **quantitative analysis** of innovation data amongst companies, establishments and workers. To document the routes by which technical change impacts upon economic outcomes and how this varies between and within European countries. In particular the group will look at the effects of product, labour and financial markets on the incentives and abilities of organisations to advance technologically. To facilitate the sharing of new results, methods and between European researchers in the economics of innovation and to improve collaborative research between European researchers, especially in the areas of comparative **quantitative analysis** and links between theory and emprical work. Contains details of the programme and a related mailing list.

Example of search results in SOSIG

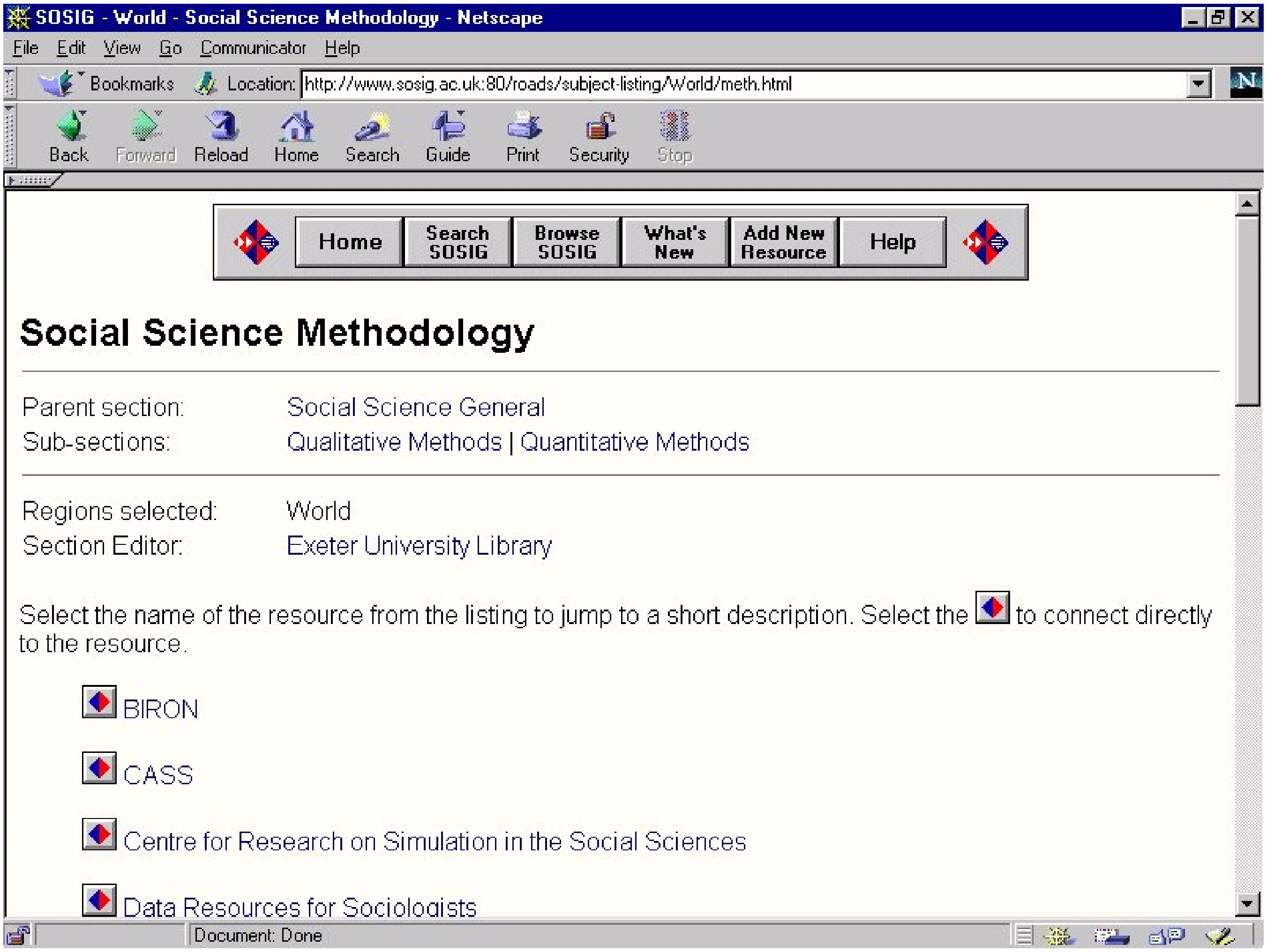
- VIDEO

Within a ROADS template, there are three kinds of attribute; plain, variant and cluster. Plain attributes describe the basic characteristics of a resource, such as Title, Description and Keywords. They contain information about a resource that is only required once. Variant attributes are repeated for multiple versions of a resource. Examples of variant attributes are Language and URI; if a document is available in English and French two sets of variant attributes would be used to record the language and URL of each version. Other examples of variant attributes include format and size of the resource. Cluster attributes record information that may be common to a number of resources, for example name, address and email details of individuals or organisations.

The original IAFA templates have been extended slightly based on requirements from the subject gateways. For example, the attributes Subject-descriptor and Subject-descriptor-scheme have been added. These attributes allow the resources to be classified using an appropriate classification scheme and this information is used to form the basis of browsable listings on the gateways. A range of administrative attributes was also added.

The subject gateways can choose which template types and attributes to use according to the requirements of their end-users. However, a minimum set of attributes exists to ensure a level of interoperability between the gateways. The core attributes are: Title, Description, Keywords, URI, Subject-descriptor and Subject-descriptor-scheme. This does not include any of the administrative attributes such as the record creation date as these are generated automatically. Individual gateways may choose to make other attributes mandatory according to the requirements of their user community. For example, SOSIG is currently extending its coverage of European resources and has made Language and Country mandatory attributes in order to support this.

A registry of the templates[10] is maintained at UKOLN.


SOSIG - World - Social Science Methodology - Netscape
File Edit View Go Communicator Help
Bookmarks
Location: http://www.sosig.ac.uk:80/roads/subject-listing/World/meth.html
Back
Forward
Reload
Home
Search
Guide
Print
Security
Stop
Home
Search SOSIG
Browse SOSIG
What's New
Add New Resource
Help
Document: Done

Social Science Methodology

Parent section: Social Science General
Sub-sections: Qualitative Methods | Quantitative Methods

Regions selected: World
Section Editor: Exeter University Library

Select the name of the resource from the listing to jump to a short description. Select the ◆ to connect directly to the resource.

◆ BIRON

◆ CASS

◆ Centre for Research on Simulation in the Social Sciences

◆ Data Resources for Sociologists

Example of browsable listing in SOSIG

This allows ROADS gateways to register the need for new template types or new attributes within existing templates if or when required. The registry will also contain some basic cataloguing rules to assist with interoperability.

Example of a ROADS Template

Template-Type: SERVICE
Handle: SOSIG472
Category: Database
Title: IBSS ONLINE
Alternative-Title: International Bibliography of the Social Sciences
URI-v1: telnet://bids.ac.uk
URI-v2: http://www.bids.ac.uk/ibss
Admin-Handle-v1:
Admin-Name-v1:
Admin-Work-Postal-v1:
Admin-Country-v1: uk
Admin-Work-Phone-v1: +44 (0)1225-826074
Admin-Work-Fax-v1:
Admin-Job-Title-v1:
Admin-Department-v1:
Admin-Email-v1: bidshelp@bids.ac.uk
Publisher-Handle-v1:
Publisher-Name-v1: Bath Information & Data Services
Publisher-Type-v1:
Publisher-Work-Postal-v1: University of Bath, Bath, BA2 7AY
Publisher-Country-v1: uk
Publisher-Work-Phone-v1:
Publisher-Work-Fax-v1:
Publisher-Email-v1:
Description: IBSS ONLINE provides electronic access to the database of the International Bibliography of the Social Sciences. It contains the bibliographic details of journal articles, book reviews books, and the chapters from selected multi-authored monographs. The database contains over 680,000 records covering publications appearing between 1981 and the present day, and is growing at the rate of approximately 100,000 items per annum. Subject coverage is based on the four principal disciplines of anthropology, economics, political science and sociology, but it also reflects the interdisciplinary nature of the social sciences. Material can be found which covers, for example, agriculture, archaeology, business studies, criminology, education, environmental issues, history, law, social policy, social work, and statistical methods. There is extensive coverage of international material. Records come from over 100 countries, and 95 different languages are represented in the database. The database is mounted at Bath Information & Data Services (BIDS). All members of UK higher education institutions (HEIs) funded by the HEFCs are eligible to use IBSS ONLINE free at the point of use. Users must register with their own institution's library.
Keywords: social science, sociology, politics, economics, anthropology
Authentication: Access to most databases is by username and password.
Registration: Users are required to register with a representative at their own HE institution.
Access-Policy: The IBSS ONLINE data may only be used by an employee, student or other person authorised by the institution or organisation which has taken out a licence to use the service.
Access-Times: All BIDS services are normally available 24 hours a day, 7 days a week.
Copyright:
Subject-Descriptor-v1: 3,301,32,33,572
Subject-Descriptor-Scheme-v1: UDC
Language-v1: en
Language-v2: en
ISSN:
Source:
To-Be-Reviewed-Date:
Record-Last-Verified-Email:
Record-Last-Verified-Date:
Destination: UK,WORLD
Record-Last-Modified-Date: Mon, 28 Apr 1997 17:21:46 +0000
Record-Last-Modified-Email: ecdh@aubergine.ilrt.bris.ac.uk
Record-Created-Date: Wed, 15 Jun 1995 13:22:00 +0000
Record-Created-Email: ecdh@ssa.bris.ac.uk

Mapping ROADS Templates

One of the main project objectives of ROADS is to 'implement and test emerging standards and to improve UK participation in international standards making activity'[11]. To this end the project closely monitors metadata developments and is very active in metadata standards initiatives, in particular the Dublin Core and Warwick Framework. Mapping ROADS templates to WHOIS++ templates has been done as part of the planned developments for distributed searching of ROADS databases (these are actually very similar in format to the ROADS templates). In addition UKOLN has produced several textual mappings from the ROADS templates to other metadata formats such as USMARC, Dublin Core, SOIF and the Z39.50 Bib-1 attribute set[12]. The templates map reasonably well on to these other formats although there may be some difficulties with syntax.

It would seem fair to assume that there will continue to be several metadata formats in use to describe Internet resources. However, ROADS has committed to providing conversion tools if they are required by the eLib gateways and some proof of concept work has already been carried out converting the ROADS templates into USMARC and other formats. As part of another project, an experimental Z39.50/WHOIS++ gateway has been built which allows users to search a ROADS database in parallel with a range of Z39.50 databases[13].

ROADS Developments

ROADS is presently in version 1 of its development cycle; this provides all the tools and software to build and maintain a gateway of Internet resources. ROADS version 2 (already in alpha development) will continue to enhance these tools in response to requirements from the subject gateways. In addition to this, the next version will include:

Cross Searching Distributed Databases

Currently each ROADS subject gateway is searchable in a standalone format although some experimental work on cross searching has already been carried out between some of the gateways. The project is using the WHOIS++ search and retrieval protocol developed by Bunyip Information Systems (who provided some industrial consultancy on the project). This allows distributed databases to be queried over the network and the next version of the software will fully support this cross searching mechanism. In addition to linking the databases together, the project will be investigating the use of a related technology - the Common Indexing Protocol (CIP) to provide a method of routing search queries to appropriate databases[14].

Cross searching will be particularly useful for the SOSIG and Biz/ed gateways whose subject areas (social sciences and business and economics) overlap; potentially causing confusion for end users trying to identify which gateway they should use. Once cross searching is implemented, SOSIG will no longer continue to catalogue business or economics resources but users of the SOSIG gateway will still be able to search for and find economics related resources. Because of the overlap, the two projects are also looking at ways of presenting browsable lists across the two projects.

As part of the DESIRE project SOSIG is also hoping to collaborate with social science institutions or libraries in Europe who want to set up national databases of networked resources. European institutions would be able to make use of the tools and documentation developed by ROADS and DESIRE to create national gateways. This model is currently being piloted by the Koninklijke Bibliotheek (National Library of the Netherlands) who are building a ROADS database of Dutch social science resources.

Harvesting Resources

The eLib ANR gateways concentrate on cataloguing high quality Internet resources and it is this human input that distinguishes them from other Web search tools such as AltaVista. Users of the gateways are not overwhelmed by thousands of matches to their queries but are presented with a small number of resources which have been through a careful process of selection and description. However, this process means that there is a high cost associated with the creation of the catalogue records. Consequently, the gateways tend to catalogue at a server level rather than at the level of individual documents or pages. ROADS is looking at incorporating a Harvest-type technology in order to try to bridge the gap between the 'hand picked' approach of the gateways and the so called 'vacuum cleaner' approach of the Web search engines. One approach is to use a harvested database to supplement the quality-catalogued records and ROADS is investigating ways to integrate and present the two.

One of the aims of ROADS and of the subject based gateways is to 'encourage information providers to become involved in the creation of records about their own data in order to make their information as useful and accessible as possible'[15]. Typically, information providers supply little or no metadata with their resources, due in part to a lack of standards or direction in this area. ROADS is promoting the idea of 'Trusted Information Providers' (TIPS) who would be identified by the individual subject gateways. The TIPS may be services or institutions whose information had been previously validated by the gateways that would provide metadata with their resources to be collected automatically.

This second level of approach to support the TIPS idea is to develop a tool that can be used to pre-populate ROADS databases by harvesting metadata from resources and inserting them into templates. The cataloguers can then 'add value' to the automatically generated template before finally submitting it to the database. The ROADS Harvester can be used to generate a single template based on one URL or it can be run recursively across a range of URLs as a 'Bulk Harvest'. The Harvester is still under development but in the longer term should help the gateways to redress the imbalance between quantity and quality.

Contact Details

Debra Hiom is a Research Officer on the SOSIG and DESIRE projects at the Institute for Learning and Research Technology, University of Bristol in the UK. She can be contacted at the following address:

Institute for Learning and Research Technology,
University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK.
Tel: +44 (0)117 928 8443
Fax: +44 (0)117 928 8478
Email: D.Hiom@bristol.ac.uk

Acknowledgements

This paper references the work of the ROADS project team; in particular Jon Knight and Martin Hamilton at Loughborough University, Rachel Heery, Michael Day and Andy Powell at UKOLN and Chris Osborne and Paul Hofman at the University of Bristol. Any inaccuracies are the author's own.

For More Information About ROADS
If you would like more information about the ROADS project and availability of the software, contact Paul Hofman at: <roads-liaison@bris.ac.uk>

References

1. Resource Organisation and Discovery in Subject-based services

<URL: http://www.ukoln.ac.uk/roads/>

2. Electronic Libraries Programme

<URL: http://www.ukoln.ac.uk/elib/>

3. ADAM (Art, Design, Architecture and Media Information Gateway)

<URL: http://www.adam.ac.uk/>

4. Biz/ed (Business Education on the Internet)

<URL: http://www.bized.ac.uk/>

5. IHR-Info (Institute of Historical Research)

<URL: http://ihr.sas.ac.uk/>

6. OMNI (Organising Medical Networked Information)

<URL:http://www.omni.ac.uk/>

7. SOSIG (Social Science Information Gateway)

<URL:http://www.sosig.ac.uk/>

8. DESIRE

<URL:http://www.nic.surfnet.nl/surfnet/projects/desire/>

9. Publishing Information on the Internet with Anonymous FTP

<URL:http://www.roads.lut.ac.uk/System-docs/Internet-drafts/draft-ietf-iiir-publishing-03.txt>

10. ROADS Template Registry

<URL:http://www.ukoln.ac.uk/roads/templates/>

11. ROADS Objectives

<URL:http://www.ukoln.ac.uk/roads/

12. Mapping between metadata formats

<URL: http://www.ukoln.ac.uk/metadata/interoperability/>

13. Zexi: Z39.50 Experimental Implementation

<URL: http://www.roads.lut.ac.uk/zexi/>

14. The Common Indexing Protocol

<URL:http://www.roads.lut.ac.uk/System-docs/Internet-drafts/draft-ietf-find-cip-new-00.txt>

15. ROADS Objectives

<URL:http://www.ukoln.ac.uk/roads/>

* Paper presented at IASSIST/IFDO '97, Odense, Denmark, May 6-9,1997.