



Working Paper

Assessment of ROADS as HeadLine Resource Classification

Editor/Author:	Stephen Graham
Project Document Number:	HL-1998-12
Publication Date:	30 Oct 1998
Task:	2.1
Status:	draft for adoption
Version:	0.2
Keywords:	HeadLine Resource Classification
Abstract:	- This document presents an evaluation of the ROADS software, which is under consideration to be used as part of the Headline service/system. -
Confidentiality:	
Copyright:	LSE, LBS, UH
Edit source file:	C:\headline\ROADS.DOC
Reference URL/URI:	
Last saved:	6 May 1999

London School of Economics • London Business School • University of Hertfordshire

The HeadLine project is partly funded by the Joint Information Systems Committee (JISC) of the Higher Education Funding Councils, as part of its Electronic Libraries Programme (eLib)

Contents

Introduction	4
Related Documents.....	4
Scope and Aim of Report	4
What is ROADS?	5
Installation details of ROADS at the LBS and LSE	5
Access details for the LBS Installations:.....	5
Access details for the LSE installation:	5
Evaluation	6
Entering Records into the ROADS Database via the standard web form	6
Importing Records into ROADS from other Applications.....	6
Non Internet Resources.....	6
Defining Relationships Between Resources and Access Rights	6
Searching ROADS.....	8
Searching another ROADS database and Forward Knowledge.....	8
ROADS and Z39.50	8
Other Considerations.....	9
The Future of ROADS	10
Conclusion	11
Appendix 1 - An example of a "raw" service template.....	12
Appendix 2 - List of Default ROADS Template Types.....	15
Appendix 3 – A “suggested” ROADS template for Headline Resource Description.....	16
Bibliography	17
ROADS.....	17
Other Technologies.....	17

Introduction

This document presents an evaluation of the ROADS software, which is under consideration to be used as part of the Headline service/system. The paper will be continuously updated as long as Headline is using (or considering to use) ROADS as a software tool to help classify and organise resources at the three partner sites. Initially, the paper will be split into three parts, firstly an introduction to and an explanation of the ROADS software, an evaluation of the software based on testing at the partner sites, and finally, a look at ROADS features not yet fully explored, and a look at future developments.

Related Documents

Other relevant documents are the **Resource Description Model**, which outlines the schema for Headline resources, and the **Consolidated List of Resources**, which outlines the resources that may be included within Headline. A metadata glossary can be found at the following URL:

<http://ukoln.ac.uk/metadata/glossary/>

Scope and Aim of Report

This report seeks to give a comprehensive introduction and evaluation of the ROADS software, to determine if it would be a useful tool for Headline, and if so to recommend it for use within the project and the Hybrid Library. The primary aim of this evaluation is to try and put the recommended Headline Resource Description Model into practice using the ROADS templates. ROADS must be able to accommodate the Headline RDM in some shape. Other considerations are:

1. Field names can differ, but ROADS must be able to manage the relationships between different categories of resources.
2. Records must be able to be imported from other sources (e.g. existing records in other databases) into the ROADS format, and also exported, without too much trouble.
3. ROADS must be flexible, and be able to be manipulated and tailored to suit Headline needs. For example - the appearance, the type of metadata that needs to be entered and stored etc.
4. As well as using a standard metadata format and open protocols, ROADS must be "aware " of, and should be able to accommodate, other and new technologies. These are technologies such as Z39.50, RDF, Dublin Core etc.
5. ROADS should have a user-friendly front-end, for both end-users searching the database, and for administrators who enter the records.

Obviously, it is important to look at other relevant eLib projects, if only to avoid duplication of work already completed. It is hoped that by working with existing eLib products, that ideas and technologies will be further developed to advance the aim of the electronic library. The report will also examine other relevant technologies that may be used instead of/with ROADS.

What is ROADS?

ROADS (Resource Organisation and Discovery in Subject-based Services), is a software tool resulting from a previous JISC funded eLib project. The primary aim of ROADS is to act as a tool for users to create subject based information gateways on the web. Examples of services that have used ROADS to build such gateways are: SOSIG (<http://www.sosig.ac.uk>) - a social science information gateway, and BIZ-ED (<http://www.biz-ed.ac.uk>) - a business information gateway. End-users can search the gateways using a HTML form via the web, and results are also rendered into HTML, with hypertext links that lead to the chosen resource.

ROADS uses IAFA (Internet Anonymous FTP Archive) derived templates for the management of resource metadata. The templates are based on different types of resources e.g. services, software, FAQs etc. (See Appendix 2 for full list of template types). Within the templates the data is arranged into simple attribute-value pairs (e.g. TITLE (attribute) = Fame (value)). At the back-end of the ROADS software is a WHOIS++ server, which creates an inverted index using the Common Indexing Protocol (see bibliography for links to further info on these protocols). The database is searched and created by a series of web pages and CGI scripts that are written in Perl. The ROADS software is written and based around open standards.

Installation details of ROADS at the LBS and LSE

The ROADS software (version 2.1) has been set up at the London Business School on a Red Hat distribution of Linux (5.2, Kernel 2.0.36) for evaluation purposes. It has also been installed at the London School of Economics for trial purposes. It is useful to have two installations of ROADS at two different institutions to test out the cross searching/forward knowledge features of the ROADS software.

Access details for the LBS Installations:

URL for test search form: <http://gallifrey.lbs.ac.uk/ROADS/cgi-bin/seach.pl>

WHOIS++ server: Port Number: 8237
Server handle: gallifreylbsacuk01
Server name: gallifrey.lbs.ac.uk

URL for test search form: <http://gallifrey.lbs.ac.uk/TESTROADS/RoadsSearch.html>

WHOIS++ server: Port Number: 8999
Server handle: gallifreylbsacuk02
Server name: gallifrey.lbs.ac.uk

Z39.50 Access: Port Number: 2104
Database name: roadsrecords
Server name: gallifrey.lbs.ac.uk
Record syntax: GRS-1, US-MARC and SUTRS

Access details for the LSE installation:

URL for test search form: <http://decomat.lse.ac.uk/ROADS/cgi-bin/search.pl>

WHOIS++ server: Port Number: 8237
Server handle: decomatelseacuk01
Server name: decomat.lse.ac.uk

Evaluation

Entering Records into the ROADS Database via the standard web form

Records are entered into the database via existing pre-defined templates. These templates are essentially HTML forms with various fields organised into several sections, according to the type of information that you are recording about the resource. The different sections allow various types of information to be recorded about the resource, e.g. the vendor's/publisher's details, personal details concerning the administrator etc. These sections are known as clusters. Once a cluster has been created, it can be re-used in any other template. For example, if there are four products that have the same vendor, the template for the vendor only needs to be created once. The record for that vendor will be given it's own unique I.D. (handle). The "handle" can then be used as the "cluster handle", and entered instead of the entire vendor's details again. Once the record has been entered, it can be indexed straight away, or it stored to be indexed at a later time using a batch method, preferably overnight.

Templates can be tailored to include only the fields that are frequently used, and that are necessary for Headline. Primarily, the main template type that would be the most useful and relevant for Headline would be the Service template (see appendix 1). The main fields that are outlined in the Resource Description Model as being necessary metadata for Headline resources are available in the Service template type. These include fields such as the title, description, keywords etc. Visit the following URL to see the proposed ROADS template for the Headline Resource Description Model, or see Appendix 3:

<http://gallifrey.lbs.ac.uk/TESTROADS/admin-cgi/admincentre.pl?form=headlineadmin>

Importing Records into ROADS from other Applications

There are no standardised procedures for importing data into a ROADS database from other applications. It is usually left as an "exercise for the reader". However, because an individual ROADS record is just an ASCII file (see appendix 1 for an example), which is made up of attribute/value pairs, it is possible to develop scripts to convert existing system formats into a ROADS record. This could be achieved with Perl. (In fact, a Perl script has been developed at the LBS) Another alternative is to get the existing alternative system to spit out ROADS formatted records if it is possible. This has been achieved previously with MS Access databases. An important field to remember is the "Handle" field. This is the unique handle (this should be alpha-numeric) for an individual ROADS record. Every record should have it's own unique handle, usually this is automatically generated by the ROADS software (based on the time and date that the record is created). If records are created outside the ROADS system, it is important to note that there will have to be a method of generating a unique handle value for the record.

Non Internet Resources

Within the "Service" template there is a field entitled "Category", which is intended to store a value denoting what type of resource that particular item "belongs" to. An "Authority File" can be made for that field, and can be used, when searching, as a search value to only return resources that belong to a particular type. Thus, if desired the "category" field could be used to denote whether that resource is a commercial database, an electronic journal, or a free web site. This means that a ROADS databases doesn't just have to be a gateway to Internet/web based resources.

Defining Relationships Between Resources and Access Rights

ROADS uses the concept of "virtual databases" to show how one resource can belong to another resource. This is achieved by using the "Destination" field. For example, if a journal is available on BIDS, we would enter the value "BIDS" in the "Destination" field, or any other value which would connote "BIDS". However, what is more problematic is the representation of multiple instances of a "single" resource. An example would be if a particular journal is available in

several different places or formats. There would be two main ways in which the resource/s could be recorded; one record (with one title and multiple holdings data), or multiple records (each record having one title and one set of holdings data). The "one record model" is problematic in two ways. Firstly, there is a problem with the way that ROADS renders the HTML. This is because field values are presented in batches. Thus, if a resource has multiple URLs, the returned record from ROADS will look something like:

Title:

Description:

URL.1:

URL.2:

URL.3 :etc

Not like (which would be better):

Title:

Description:

URL1:

Description:

URL2:

Description:

URL3: etc

The second, and probably the more problematic issue, is the assigning of access rights to individual fields. Rights can be assigned in a very crude way at record level, by using the "Authentication" field in the "Service" template. A simple demonstration of this can be see at:

<http://gallifrey.lbs.ac.uk/TESTROADS/RoadsSearch.html>

However, it would appear that this basic approach would not work at field level. This would mean that it would not be possible (in the current ROADS set-up) to return a specific URL depending on what sort of user you are. Thus, if a user only has rights to use 2 out of three resources, it would not be possible to only return the two "allowed" URLs. With ROADS it would only be possible to return all 3 URLs.

Searching ROADS

ROADS is searchable from a HTML form on the web (e.g. <http://gallifrey.lbs.ac.uk/ROADS/cgi-bin/seach.pl>). Like the form for entering records into the database, the search form can be tailored as much as required. We can choose which field and templates the end-user can search by, we can decide which search options we want to allow, such as "case sensitive search", "follow referrals", and "stemming", and we can choose which fields are included in the returned search results. However, the returned search results have to have a URI. This is because the search result fields are controlled at template type level. Thus, if you create a search display view for (for example) the Service template, all records that are based upon the Service model will be displayed in that way. This means that, if we use the service template type for all none web-based electronic resources we need to decide what the URI will link to. Of course, one of the work packages for Headline is to create an application launcher, thus products will be launched from a web page. It must be noted that not all electronic products are networked, and are only available on stand alone machines (e.g. Bloomberg, Reuters 2000/3000), also some products are restricted to a specific physical location (e.g. Datastream, Reuters Business Briefing, Dow Jones Interactive to the LBS Library). In cases such as these the URI will have to link to a page of further information, or maybe a FAQ on the particular product.

Searching another ROADS database and Forward Knowledge

One particular feature of the ROADS database, is the option to search another ROADS database regardless of location. All that is needed is the WHOIS++ server details of the ROADS database that you want to search. Once these have been entered into the database configuration file locally, the database should appear in the search form. The user can search both databases at the same time, if they wish. This feature is being used on both the LSE's and the LBS's test databases. However, this is rather a crude method, as a search may take significantly longer whilst an "external" database is interrogated, only to find that the database has no relevant records. This is where the concept of "forward knowledge" querying is useful. A WHOIS++ server can be set up to visit a remote ROADS database, take a copy of the main inverted index (known as a centroid) and store it locally. Thus, when a user searches a remote database, the local copy of the index will be checked first to see if the remote database holds any relevant records. If it doesn't, the query will not be sent to the remote database. If it does hold relevant records the query will be sent and the relevant records will be retrieved. We have not yet explored this feature, but it may be useful if we are searching large "remote" WHOIS++ databases.

ROADS and Z39.50

As well as being searchable via the standard ROADS web form, the ROADS database can be made searchable from a Z39.50 client. This is achieved by using the add-on script, roads2gils.pl, and a Z-Server called Zebra. The add-on script converts native ROADS templates into GILS records. The Zebra System can make these converted ROADS records available in two structured formats (USMARC and GRS-1) and in unstructured format (SUTRS). The records are then indexed and can be searched from any Z-Client. The conversion is relatively straightforward process, and can be run automatically in batch mode over night.

ROADS also mentions use of a Z-product called Zexi. This product can be used to put a web front end on Z-targets. This uses a tailor made ROADS script (zoro.pl) . However, this is still experimental, and I have no success setting this up.

Other Considerations

The ROADS software is being developed all the time. The latest ROADS Annual Report outlines the majority of future developments and main issues for consideration. The following areas need to be explored further:

1. The "phasing out" of the WHOIS++ protocol. It seems that the WHOIS++ protocol is being superseded by new technologies, such as RDF (Resource Description Framework) (an application of XML) and Dublin-Core. However, these are relatively new and "un-complete" protocols. It is hard to compare ROADS with these, because there are no applications which exploit the new technologies. In fact RDF is still a draft proposal. In spite of the "newness" of these protocols, there is a ROADS script which will help convert ROADS records into Dublin Core format, and also the same for RDF. Because the ROADS template structure is so open (a ROADS record is just a plain ASCII file), it is easily manipulated into other formats. If ROADS is adopted by Headline, and then we decide to use another technology, it should be no problem to produce scripts (if there are none available), to transform the records into the desired format.
2. Local subject terms and thesaurus. There is the option of using the institutions' local classification scheme within the ROADS system to catalogue the resources. This involves using the "descriptor" template. Using a standard classification may prove a problem for cataloguing many of the electronic resources, as some resources/products cover several subject areas. For example, the Bloomberg system covers a wide range of areas such as news, macroeconomic data, stock market prices, and even the OED (Oxford English Dictionary). For the ROADS evaluation it has been sufficient to use a limited set of words and terms to describe the resources within the keywords field, and the idea of having a controlled vocabulary/thesauri has not been explored.
3. Z39.50. The bibliographic search and retrieval protocol could come in very useful for some Headline purposes. However, products such as ROADS, Unicorn, Decomate and Econlit are all Z39.50 compliant. The majority of Z-Clients that are available are not at all user-friendly, and are not suitable for the general end-user. What would be of more interest to Headline, and more suitable for the end-user would be a web based front-end that is capable of interrogating Z-Targets. It would be prudent to carry out more investigation on the Z capabilities of ROADS, and a Z39.50 capable web front-end in general (continuing looking at Zexi and products such as EuropaGate .
4. User data and ROADS. ROADS does have a template for end-user data. However, the user template is not very sophisticated, and only includes basic information concerning a user:

Name, work phone, work fax, work postal, country, job title, department, e-mail, home phone, home postal, home fax.

Obviously, this is not sufficient for Headline purposes, and a tailor made Headline user template would also miss the mark. ROADS does not have a built-in concept of authentication or authorisation , except for the generic BASIC authentication which is used by all httpd servers. There is some work being undertaken by the ROADS team, looking at LDAP (Lightweight Directory Access Protocol), and this may be worth keeping an eye on. However, the whole issue of interacting with existing MIS user data, is still an issue.

The Future of ROADS

The version of ROADS that is currently under evaluation is version 2. Version 3 is under development at the moment, and the alpha release is soon to be released. The following is a list of features that will be included in version 3. This list is developments outlined in the ROADS Annual Report 1998.

1. A Windows/X-Windows graphical user interface for ROADS administrators. This should work on Unix/Linux based systems and Win32 platforms (95/98/NT). This is to compliment the web based interface and is designed to make the adding, editing and removing records easier.
2. Interoperability with Windows NT Systems. This is still work in progress, and is rather vague. Interaction via ODBC and Microsoft Access databases, and Active Server Page technology (which is available with the MS IIS Web Server) are some of the area that the ROADS team are looking at.
3. RDF and XML. The ROADS team are looking at the latest developing technologies such as RDF, XML, COBRA and Dublin Core. Scripts have been written to convert native ROADS records into DC and RDF format - see the bibliography for more details, and also see the following page: <http://www.ukoln.ac.uk/metadata/> , where there are links to software tools for handling metadata. Also, there is a useful link to the "Mapping between metadata formats" page, which outlines the main metadata formats and how to convert them.
4. Z39.50. Work is continuing on using Z39.50 with ROADS.
5. Funding of ROADS and other ANR services. It is unclear that JISC will continue the long term funding of ANR (Access to Network Resources). JISC are currently waiting for various study groups to report back to them. This has a direct bearing on the future of the ROADS software. ROADS currently has funding for the next year.

Conclusion

The main obstacle, from the point-of-view of trying to implement the Headline Resource Data Model (RDM) using ROADS, is the "flat" structure of the IAFA/WHOIS++ data model. Even though ROADS uses unique handles to represent individual records, which then can be inserted into other records instead of re-entering the data, the Headline RDM requires a much more "relational" data structure. Another "weakness" of ROADS is the current lack of support for resources at collection level. (there is a collection template which has been recently released, but this is still "experimental" I.E. in draft format). The IAFA template that has the most relevance, for storing metadata for Headline resources is the Services template. However, whilst this does offer some degree of customisation, it is not sufficiently flexible to accommodate an adequate representation of the Headline RDM.

An example to illustrate the non-relational nature of ROADS is the inability to model the resource/resource instance relationship. A resource (such as a journal title) can belong to several collections (e.g. a database/product such as Proquest Direct, the hardcopy version in the Library collection etc), and the "holdings" information for each "instance" can (and almost always is) different. A journal may be available in full text from 1980 onwards via one collection, whilst on another it may be available in abstract only from 1993 onwards. Whilst the "holdings" information may be different, there are various other metadata about the resource which will be the same for each instance of that particular resource. (E.G. Title, Frequency of Publication, Publisher details etc.) Currently, within ROADS, to model this resource/resource instance relationship a separate record needs to be created for each resource instance. It is trying to model real-life relationships such as this (with one to many, as well as one to one relationships), that the unsuitability of a flat data structure for the Headline RDM is made clear.

It is recommended that the Headline project look at other "open source" database systems which are fully relational, such as PostgreSQL and MySQL to see if these systems are more adequate for modelling the recommended Headline RDM.

Appendix 1 - An example of a "raw" service template

Template-Type: SERVICE

Handle:907860726-1807

Template-Version:1

Title: ABI Inform

URI-v1:<http://gallifrey.lbs.ac.uk/abi.w3l>

Admin-Handle-v1:

Admin-Name-v1:

Admin-Admin-URI-v1:

Admin-Work-Phone-v1:

Admin-Work-Fax-v1:

Admin-Work-Postal-v1:

Admin-Country-v1:

Admin-Job-Title-v1:

Admin-Department-v1:

Admin-Email-v1:

Admin-Home-Phone-v1:

Admin-Home-Postal-v1:

Admin-Home-Fax-v1:

Admin-Destination-v1:

Owner-Handle-v1:

Owner-Name-v1:

Owner-Owner-URI-v1:

Owner-Type-v1:

Owner-Postal-v1:

Owner-City-v1:

Owner-State-v1:

Owner-Country-v1:

Owner-Email-v1:

Owner-Phone-v1:

Owner-Fax-v1:

Owner-Destination-v1:

Sponsoring-Handle-v1:

Sponsoring-Name-v1:

Sponsoring-Sponsoring-URI-v1:

Sponsoring-Type-v1:

Sponsoring-Postal-v1:

Sponsoring-City-v1:

Sponsoring-State-v1:

Sponsoring-Country-v1:

Sponsoring-Email-v1:

Sponsoring-Phone-v1:

Sponsoring-Fax-v1:

Sponsoring-Destination-v1:

Publisher-Handle-v1:

Publisher-Name-v1:

Publisher-Publisher-URI-v1:

Publisher-Type-v1:

Publisher-Postal-v1:

Publisher-City-v1:

Publisher-State-v1:

Publisher-Country-v1:

Publisher-Email-v1:

Publisher-Phone-v1:

Publisher-Fax-v1:

Publisher-Destination-v1:

Description: ABI Inform provides indexing and abstracting of over 1000 business and management periodicals. There are approximately 200,000 records; each record contains a full bibliographic citation and a 150 word abstract. It is available from 1971 to date. Updated monthly.

Authentication: Network Login

Registration:

Charging-Policy:

Access-Policy: All LBS Members

Access-Times: 24 Hours

Keywords: ABI Inform, journal abstracts, business, management, finance, articles

Subject-Descriptor-v1:

Subject-Descriptor-Scheme-v1:

Relation-Type-v1:

Relation-Target-v1:

Short-Title:

Alternative-Title:

Language-v1:

ISSN:

Discussion:

Source:

Category: Database

To-Be-Reviewed-Date: Thu, 31 Dec 1998 15:28:37 +0000

Record-Last-Verified-Email:

Record-Last-Verified-Date:

Comments: Networked CD-ROM

Destination:

Record-Last-Modified-Date: Thu, 08 Oct 1998 15:54:27 +0000

Record-Last-Modified-Email: unknown@163.119.251.238

Record-Created-Date: Thu, 08 Oct 1998 15:32:07 +0000

Record-Created-Email: unknown@163.119.251.238

Appendix 2 - List of Default ROADS Template Types

DATASET

DOCUMENT

DUBLINCORE - experimental

EVENT - experimental

FAQ - no longer in use, use DOCUMENT

IMAGE

MAILARCHIVE

PROJECT

SERVICE

SOFTWARE

SOUND

TRAINMAT

USENET

VIDEO

Clusters:

ORGANIZATION CLUSTER

USER CLUSTER

Other:

Codes for Subject-Description-Scheme element (DRAFT)

See the following URL for further information on all these templates:

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

Appendix 3 – A “suggested” ROADS template for Headline Resource Description

Title:**Description:** (Free Text)**Authentication:** (this field is being used with an “Authority File”, and lists the different types of users e.g. Students, staff, faculty etc)**Access-Policy:** (this field is being used with an “Authority File”, and lists the different locations that users can access services from e.g. Library, campus, anywhere etc)**Keywords:** (Free Text)**Alternative-Title:****Category:** (this field is being used with an “Authority File”, which list the different medium/types of resources e.g. Database, Journal, Free Web Sites etc)**Comments:** (General notes – free text)**Destination:** (this field is being used with an “Authority File”, and lists the various resources that other resources are available from to e.g. JSTOR, Super-Journal, Dow Jones Interactive etc)**URI:****Subject-Descriptor and Subject-Scheme:** (This allows an institution to use their local classification scheme to catalogue resources on ROADS – this is not currently be used in the LBS test database)**Publisher:** (name, address, e-mail etc – contact details for the different vendors)

* We may need to add a “holdings” field for journals. This field is not included in the stanard IAFA Service template. Thus, holdings data could be included in the description field.

Bibliography

Here is list of useful documents, that expand on the issue and technologies covered in this paper:

ROADS

ROADS Annual Report 1998, <http://www.ilrt.bris.ac.uk/roads/papers/annual98/roadsar98.html> (covers the future development of ROADS)

ROADS Cataloguing Guidelines, <http://www.ukoln.ac.uk/metadata/roads/cataloguing/cataloguing-rules.html> (covers existing cataloguing rules for Internet resources, and how ROADS fits in)

ROADS Template Registry, <http://www.ukoln.ac.uk/metadata/roads/templates/> (a complete list and description of all the standard ROADS templates)

Using ROADS for web-site metadata management, <http://www.ukoln.ac.uk/metadata/roads/metadata-mgmt/> (Looks at ROADS in relation to Dublin Core and RDF)

WHOIS++, <http://service.bunyip.com:8000/products/whois++/whois++.html> (An introduction to the WHOIS++ protocol)

The Common Indexing Protocol, <http://www.dstc.edu.au/AW3TC/papers/falstrom> (An introduction to the CIP which is used by WHOIS++)

The architecture of the Common Indexing Protocol, <http://www.ietf.org/internet-drafts/draft-ietf-find-cip-arch-01.txt>, (The Internet Engineering Task Force draft for the CIP)

Other Technologies

Dublin Core Metadata, http://purl.org/metadata/dublin_core (Introduction to DC metadata)

An Introduction to the Resources Description Framework, Eric Miller, D-LIB Magazine, May 1998, <http://www.dlib.org/dlib/may98/miller/05miller.html> (like the title says, an introduction to RDF)

What is.....XML?, Brian Kelly, Ariadne, Issue 15, <http://www.ariadne.ac.uk/issue15/what-is/> (simple introduction to XML, with good links to other relevant sites)