



Project Plan Cover Sheet

Project Information			
Project Acronym	TRSS		
Project Title	Terminology Registry Scoping Study		
Start Date	7 Feb 2007	End Date	7 Aug 2008
Lead Institution	UKOLN - The University of Bath		
Project Director	Michael Day		
Project Manager & contact details	Koraljka Golub UKOLN The University of Bath Bath, BA2 7AY Tel: +44 (0) 1225 383619 Email: k.golub@ukoln.ac.uk		
Partner Institutions	UKOLN - The University of Bath; University of Glamorgan Non-funded supporting partner: OCLC Office of Research, USA		
Project Web URL	http://www.ukoln.ac.uk/projects/trss/		
Programme Name (and number)	Repositories and preservation programme		
Programme Manager	Tom Franklin		

Document Name			
Document Title	Project Plan		
Author(s) & project role	Project Team		
Date	31 Jan 2008	Filename	PPlan-TRSS-v3.doc
URL	http://www.ukoln.ac.uk/projects/trss/documentation/		
Access	<input type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
Draft 1	21 January	Submitted to Programme Manager
Draft 2	29 January	Paper version given to Programme Manager
Draft 3	31 January	Submitted to Programme Manager



JISC Project Plan

Overview of Project

1. Background

A terminology registry lists, describes, identifies and points to sets of vocabularies available for use in information systems and services. It can cover free and publicly available, fee-based and restricted, or organisation-internal vocabularies. The registry allows discovery of suitable schemes for information or, potentially, use, by exposing rich metadata about them for navigation and retrieval. The metadata can hold information allowing the selection of schemes suitable for different purposes, address information for contacting owners and maintainers, hypertext-links to connect to the vocabularies or maintainer sites, information to differentiate between versions and identifiers, names and labels to unambiguously refer to a given scheme. Terminology registries can hold scheme level information only, or comprise the member terms, concepts and relationships as well, or even list services based on terminology (such as automatic classification, term expansion, disambiguation, translation, semantic reasoning). Registries should, if used as a digital infrastructure service, make their content available for both comfortable human inspection and for machine-to-machine access.

Shared infrastructure services are essential building blocks for an efficient and effective information and communications environment. These are services that operate as underlying machine-to-machine (M2M) services, although they may well in addition provide access for human users (preferably using the same underlying M2M services). The recent JISC Shared Infrastructure Services Synthesis Study makes several recommendations regarding development of shared services and these will be used to inform the Terminology Registry scoping study.

The project will build on the rich experiences of the partners in the area of terminologies and Networked Knowledge Organization Systems and Services (NKOS) and on significant work regarding registries. The lead partner UKOLN is active in the two JISC Shared Infrastructure Services projects most relevant to the Terminology registry: IEMSR and IESR which both aim to inform humans and to support m2m services. HILT is another relevant service, aimed at terminology mapping, with which collaboration would be established. In 2007, UKOLN carried out JISC commissioned reports relevant to this study: The Shared Infrastructure Services Review and, in partnership with the University of Glamorgan, the Terminology Services and Technology Review.

The study will be informed by lessons learned from previous registry efforts. Before Web usage became popular in the early nineties, comprehensive lists of KOS were only available in special printed volumes gathered by publishers or large organisations. The description/metadata provided was usually rather poor and did not very well support decisions about which KOS to use. National, regional, local and domain organisations often created and maintained lists of KOS in use by their own organisation.

There have been a number of efforts to maintain such lists. One of the larger and more recent such lists, the Thesaurus Guide [Thesaurus guide], published by the EU Commission, and containing about 700 vocabularies available in at least one of the EU languages, was also available as a database between 1993 and 1998. More than 2000 classification schemes, subject heading lists, and thesauri in the English language are physically collected at the University of Toronto and catalogued as publications in the Subject Analysis Systems (SAS) Collection [Subject analysis] in the University of Toronto Library online catalogue, and can be retrieved from there. WorldCat (OCLC) [WorldCat] contains many publication records cataloguing terminologies. Since 1996 several lists of online available KOS in digital formats have been created, however none is consistently enlarged or

maintained [e.g. Koch; HILT]

The commercial company Synapse (now: Factiva) started "Taxonomy Warehouse" [Taxonomy Warehouse] in 2003, a directory of taxonomies, thesauri, classification and categorization schemes from around the world, initially with about 200 records. It has just been relaunched in October 2006. Here, a simple metadata schema is used to describe vocabularies. Taxonomy Warehouse focuses on taxonomies for corporations and offers more than 550 taxonomies, arranged in 73 subject domains, produced by 260 publishers in 39 languages. More than 100 of these taxonomies can be licensed directly through Taxonomy Warehouse. It would be useful for the study to examine how well the scope and services offered fit with requirements of the JISC IE.

For some time the Dublin Core Metadata Initiative (DCMI) developed and tested a registry of "vocabulary encoding schemes", alongside its metadata registry, featuring a simple metadata schema to describe and label/name available vocabularies to be used in metadata records. For different reasons, mostly related to governance issues, this effort was cancelled in 2004. However, in the USA the NSDL Registry is now being developed using a highly similar approach.

The NKOS network [NKOS] started an effort to design a terminology registry in 1998, emanating from discussions at the second NKOS workshop at the ACM Digital Library Conference. A small task force led by Linda Hill subsequently developed a very detailed metadata schema for the purpose, containing most of the information one would need to make an informed decision about the selection of an appropriate vocabulary. Version 2 was published on the NKOS website in Nov. 1998 [NKOS Registry Version 2]. Prior to the NKOS workshop 2001, Diane Vizine-Goetz from OCLC Research developed a more formal document as a draft, converting most of the descriptive data selected in the prior versions into a Reference document for data elements, based on Dublin Core elements described according to the ISO 11179 standard [NKOS Registry Version 3]. As yet no suitable host has been identified to fund and maintain the development of such a terminology registry. Terminology registries were one of the main topics at the NKOS Special Session at DC 2005, bringing together the (DC) metadata and NKOS communities, featuring a main presentation by Rachel Heery [Heery]. The 2006 European NKOS workshop, again discussed the need for a registry.

Government is another application sector which has shown a lot of interest in the terminology registration issue. The Canadian Government [Libraries] runs an internal registry of vocabularies in use. US Government agencies (DoD, EPA, USGS, National Cancer Institute, Lawrence Berkeley National Lab. etc. including some European partners such as EEA) have engaged in a large five year project started 2004/5 called XMDR, eXtended Metadata Registries [XMDR]. It builds upon and contributes to the further development of the ISO 11179 Metadata Registries family of standards [ISO 11179]. This effort has close links to the Language Engineering community and most related ISO subcommittees (SC 32, TC 37/SC 4). Compared with the other efforts mentioned here, the focus seems to be more on a registry of individual terms than on vocabulary schemes and collections [Bargmeyer].

Apart from the motivations behind the initiatives described above, the need for terminology registries has been underlined by a number of current initiatives. The study will consider how current initiatives might contribute to meeting JISC requirements.

The UK museum and heritage sector has begun to take steps (following the early NKOS registry approach), to progress from initial unsystematic terminology descriptions on the "WordHoard" web pages to a more easily managed online reference source containing systematic, consistent and complete descriptions of relevant terminologies and subsequently moving towards a formal registry of some kind [Lee]. Such a registry is needed to assist in creating resource metadata according to the UK museums standard SPECTRUM and the historic environment standard MIDAS, as developed by the MDA and the Forum Information Standards in Heritage (FISH). Today, the SPECTRUM Terminology Bank offers limited metadata about relevant terminology resources [SPECTRUM Terminology].

The US NSF funded National Science Digital Library (NSDL) project has started to develop both a

metadata and a vocabulary registry [NSDL Registry], in one common registry. At this time, primarily a few education vocabularies are registered. The metadata about each scheme is very limited. The registry project aims, however, to address term history, vocabulary versions and SKOS encoding and provides some use cases.

During a renewed effort to discuss and further develop a typology of KOS at the NKOS workshop at ECDL 2006 [Tudhope], the extensive discussion constantly went back to underline the need for a terminology registry of individual instances of KOS systems, rather than, or in parallel to, a general typology. It is likely, however, that a typology would be required for an effective large scale registry of vocabularies. A task force of NKOS is expected to take this work forward again.

2. Aims and Objectives

2.1. Aims

The study's overall aims are:

- To inform the development of shared infrastructure for resource discovery;
- To describe the scope and potential use of a terminology registry;
- To analyse requirements for services based on a terminology registry; and,
- To help stakeholders understand the need for this component of a shared infrastructure.

2.2. Objectives

In order to meet these aims the study will have the following objectives:

- To develop a set of usage scenarios and use cases that demonstrate how and why a terminology registry as a shared infrastructure service is required;
- To gather requirements from various sources, such as documentation from JISC projects and IE architecture papers, prior work elsewhere, contact with key stakeholders;
- To synthesise the outcomes of efforts to date, from JISC activities and the wider context, including their current status, and, if ended, reasons that led to their termination;
- To include the international and commercial context; and,
- To analyse the potential costs, benefits and risks of terminology registries as shared infrastructure service.

3. Overall Approach

3.1 Scope and boundaries of the work

In order to achieve the aims and objectives, the study will focus on identifying relevant information available from prior efforts (see Background section) and project documentation, supplemented by information obtained through email, telephone calls and a small number of face-to-face meetings. The study will also use expertise available at UKOLN and the partners to inform the study in appropriate areas:

- Consultation will take place with key JISC services, projects and executive across digital library, research and learning domains (to include representation of repositories and digital preservation in particular).
- Due to the short timescale, the study will concentrate on gathering and prioritising requirements from the most relevant stakeholders in the wider community (to include key vocabulary owners relevant to JISC).
- An international perspective will be sought via email and possible telephone interviews using UKOLN and the partners' contacts.

3.2 Strands included

The study will comprise the following strands:

Requirements: These will be expressed through scenarios and use cases, set in context of current and anticipated work practice and prioritised. This activity will draw on interviews with key stakeholders and existing collections of scenarios and use cases within the UK and beyond. The

study may also construct new scenarios and use cases as appropriate. The study will review the use context and lessons learned from current and previous efforts at terminology registries.

Role of registry: The requirements will be analysed and potential functions of TRs within the JISC Information Environment and e-Framework will be outlined. This will include consideration of different TR configurations and the relationship of potential TRs to relevant JISC projects (including IEMSR, IESR, HILT) and potential stakeholders. Relationships with existing metadata and service/collection registries will be explored. The potential for co-ordination, integration and reuse will be considered. Previous and existing terminology registries will be analysed, both UK and international. Potential registry functionality will be prioritised. Potential registry content will be prioritised, including vocabulary metadata, individual vocabulary elements and any possible typology of vocabularies. Any special relevance of TRs to particular subject domains or application areas will be considered. The methods for entering content will be considered.

Architecture and Technologies: Potential underlying technologies, representation and access protocols and their associated standards will be examined. Different general options for realising TRs will be outlined, their integration into the JISC IE compared and their cost/benefits analysed. These options will include simple versus advanced solutions, centralised vs. distributed solutions; together with relevant standards. The study will consider representation and formats of vocabularies, terms and concepts and access protocols, in case the individual elements of a vocabulary are addressed by or accessed from the registry.

Governance and Organisational issues: The study will include consideration of policies regarding the use of a registry and its terminologies, including IPR, cost, sustainability and quality control issues. Different possibilities for creation, hosting and maintenance will be considered. Business models for the provision of infrastructure is a complex issue. The study will look at existing comparable examples and discuss different options in principle, including potential costs and benefits. While quantitative calculations may not be possible to provide, qualitative reasoning about these important considerations is certainly necessary.

4. Project Outputs

Project plan
Intermediate report (draft)
Final report

5. Project Outcomes

In common with other registry services, once a Terminology Registry is populated with useful information a number of services might be built upon it exploiting the content. These services might have relevance across domains. The registry might support a range of tasks such as metadata creation, information management, knowledge organisation, discovery and retrieval, and it might well be relevant beyond the area of repositories, for example in digital curation and preservation, in e-learning, in e-infrastructure, in digital library, semantic web, museum and archive communities and in commercial and open e-publishing. The scoping study will discuss common interest and overlap of effort with these related communities and propose options for cooperation. Whilst taking the wider perspective into account, the scoping study's main focus will be on the role of such a registry in delivering enhanced resource discovery within the JISC Information Environment.

6. Stakeholder Analysis

An indicative stakeholder analysis is included here to illustrate potential range of stakeholders.

Stakeholder	Interest / stake	Importance
JISC projects (e.g. IESR, IEMSR, HILT, GeoXwalk)	Potential modification of future direction. Cooperation.	High
JISC services (e.g., MIMAS, EDINA)	Potential contributor and user of registry services	High
Digital Curation Centre	Cooperate with representation information registry	Medium
e-framework	Definition of registry services	High
e-science/infrastructure	Collaborator, data exchange, re-use of registry software	High
Other domains: Museums Libraries and Archives Council (MLA), NHS, e-government	Collaborator, data exchange, re-use of registry software	medium
Vocabulary owners	Cooperation	High
International registry initiatives	Cooperation	Medium

In addition, the collaborative proposal addresses Welsh priorities concerning promotion of research capability and collaboration. Progress towards a terminology registry will support the development of e- and distance learning/research through enhancements to ongoing Repository, JISC Information Environment and eFramework initiatives.

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Project is over-ambitious in scope and/or over-runs	2	2	4	Agree scope with JISC by means of project plan
Difficulties in getting feedback from wider community	4	3	12	Prioritise drawing up structured interviews. Arrange interviews early in project

8. Standards

N/A

9. Technical Development

N/A

10. Intellectual Property Rights

The project will comply with the terms of the JISC Funding Agreement. It is expected that the final report will be made openly accessible with Creative Commons license as appropriate.

Project Resources

11. Project Partners

UKOLN, University of Bath

Koraljka Golub leads the UKOLN effort and is responsible for TRSS project management and will carry out research for the project.

University of Glamorgan

Douglas Tudhope is Professor in the Faculty of Advanced Technology, University of Glamorgan and directs and carries out the research work.

OCLC

OCLC will provide assist with advice as appropriate. Diane Vizine-Goetz is main contact.

A consortium agreement is being prepared and will be signed and sent to Programme Manager shortly.

12. Project Management

Project management and partner co-ordination will be provided by UKOLN. Communication between partners will be supported by email-based discussions, telephone meetings and a few face-to-face meetings. Project reports will be supplied and co-ordinated by the UKOLN. The project manager will spend 10% on the management.

Project team

UKOLN		
Michael Day	Project Director	m.day@ukoln.ac.uk UKOLN University of Bath, Bath, BA2 7AY tel: +44 (0) 1225 383923 fax: +44 (0) 1225 386838
Koraljka Golub	Project Manager/Research Officer	k.golub@ukoln.ac.uk UKOLN University of Bath, Bath, BA2 7AY tel: 01225 383 619 fax: 01225 386 838
Sally Lewis	Project Administrator	s.lewis@ukoln.ac.uk UKOLN University of Bath, Bath, BA2 7AY tel: 01225 386 250 fax: 01225 386 838
University of Glamorgan		
Douglas Tudhope	Glamorgan Research Officer	dstudhope@glam.ac.uk Faculty of Advanced Technology, University of Glamorgan, Pontypridd, CF37 1DL tel: 01443 483609 fax: 01443 482715
OCLC		
Diane Vizine-Goetz	Contact for OCLC research	vizine@oclc.org OCLC 6565 Kilgour Place Dublin, Ohio 43017-3395, USA tel: +1 614 764 6084 fax: +1 614 764 6096
Andrew Houghton	Contact for OCLC research	houghton@oclc.org OCLC 6565 Kilgour Place Dublin, Ohio 43017-3395, USA tel: +1 614 764 6003 fax: +1 614 718 7075

13. Programme Support

General alerts on other JISC projects and reports which are particularly relevant to TRSS.

14. Budget

See Appendix A.

Detailed Project Planning

15. Workpackages

See Appendix B.

16. Evaluation Plan

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Months 2-4	Interviews	Are we getting data wanted	Structure interview questions; Pilot test interview questions	
Months 3-6	Reports	Are important issues for stakeholders addressed?	Takes into account input from stakeholders	Production of report that represents interests of stakeholders

17. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
End of project	Final report	JISC	To inform beneficial developments and motivate buy-in	Whether and how TR(s) useful and feasible
Throughout project and afterwards	Presentations at conferences and other events	Information services providers, researchers	To foster further collaborations and ensure buy-in	
Throughout project	Web site	All of above	All of above, enable wider access findings	

19. References

- Bargmeyer**, Bruce (2005). eXtended Metadata Registries (XMDR). Presentation at the 7th NKOS Workshop at JCDL 2005. <http://nkos.slis.kent.edu/2005workshop/Bargmeyer.ppt>
- Heery**, Rachel (2005). (Metadata and) Vocabulary Registries. Presentation at the NKOS Special Session at DC-2005. <http://www.ukoln.ac.uk/terminology/events/NKOSatDC2005/NKOSatDC2005/heery-nkos-v2.ppt>
- HILT**: Vocabulary resources. <http://hilt.cdlr.strath.ac.uk/Sources/vocabulary.html>
- ISO 11179**. <http://metadata-standards.org/11179/>
- Koch**, Traugott (1996-2005). Controlled vocabularies, thesauri and classification systems available in the WWW. <http://www.lub.lu.se/metadata/subject-help.html>
- Lee**, Ed (2004). Terminology resource discovery for the UK heritage sector – a case study. Presentation at the 3rd European NKOS Workshop. <http://www2.db.dk/nkos-workshop/pp%20presentationer/NKOS-Ed.pdf>
- Library and Archives Canada**: Classifications, Thesauri and Controlled Vocabularies. 1996- <http://www.collectionscanada.ca/8/4/r4-280-e.html>
- MARC Code List**: PART III: Classification Sources. <http://www.loc.gov/marc/relators/relaclas.html>
- MARC Code List: PART IV: Term, Name, Title Sources. <http://www.loc.gov/marc/relators/relasour.html>
- NKOS** network. <http://nkos.slis.kent.edu/>
- NKOS Registry. Reference document for data elements**. Version 3, Draft, 14 June 2001 http://staff.oclc.org/~vizine/NKOS/Thesaurus_Registry_version3_rev.htm
- NKOS Registry, Version 2** with Documentation for Data Elements - Draft, 11/24/1998 <http://nkos.slis.kent.edu/registry2.htm>
- NSDL Registry**. <http://metadataregistry.org/>
- SPECTRUM Terminology Bank**. <http://www.mda.org.uk/spectrum-terminology/termbank.htm>
- Subject Analysis Systems (SAS)** Collection. Faculty of Information Studies Inforum Library, University of Toronto. http://www.fis.utoronto.ca/index.php?option=com_content&task=view&id=386&Itemid=134
- The physical collection is catalogued and searchable in the University of Toronto Library Online Catalogue at: <http://webcat.library.utoronto.ca/>
- Taxonomy Warehouse** (2003-). Factiva. <http://www.taxonomywarehouse.com/>
- Thesaurus guide**: analytical directory of selected vocabularies for information retrieval, 1992. 2nd ed. / prepared by EUROBrokerS for the Commission of the European Communities. Luxembourg : European Communities, 1993. 1033pp. (EUR/92/14006) ; (Rapports EUR 14006).
- Tudhope, Douglas** (2006). A tentative typology of KOS: towards a KOS of KOS? Presentation at the 5th European NKOS Workshop. <http://www.ukoln.ac.uk/nkos/nkos2006/presentations/tudhope.ppt>
- WorldCat** (Open Web Access). <http://www.worldcat.org/>
- XMDR**. <http://www.xmdr.org/>

Appendixes

Appendix A. Project Budget

Appendix B. Workpackages

See separate file.