Cover Sheet for Proposals	*JISC Capital Programme*
(All sections must be completed)	

Name of Capital Programme: Repositories and Preservation Programme

Bid for Call Area : (Please tick ONE BOX ONLY, as appropriate)

	Tools and Innovation (Strand B)	
	Call Area I – Tools and Innovation Projects	Please specify area of proposed project eg *'metadata generation and validation'*

	Discovery to Delivery (Strand C)	
	Call Area II – Discovery to Delivery Projects	**a)** Version identification framework
		b) Persistent identifier interoperability demonstrator
		c) Federated access management and repositories
		d) Semantic interoperability demonstrator

	Repository Start-Up and Enhancement (Strand D)	
	Call Area III – Repository Start-Up and Enhancement Projects	**a)** Repository start-up projects
		b) Repository enhancement projects

	Digital Preservation and Records Management (Strand H)	
	Call Area IV – Digital Preservation and Records Management Projects	**a)** Digital preservation across the lifecycle
		b) Models and implementation of preservation services
		c) Preservation tools development

	Shared Infrastructure Services (Strand I)	
	Call Area V – Shared Infrastructure Services Projects	**a)** Pilot implementation of licence registry
		b) Pilot national name and factual authority service
		c) Scoping an architecture to support digital policy management
		d) Scoping a terminology registry

Name of Lead Institution:

UKOLN, University of Bath

Name of Proposed Project:

Enhanced Tagging for Discovery

Name(s) of Project Partner(s):

UKOLN, University of Bath; University of Glamorgan; Intute (MIMAS The University of Manchester); CCLRC
Non-funded supporting partners: OCLC Office of Research, USA

Full Contact Details for Primary Contact:

Name: Rachel Heery
Position: Deputy Director UKOLN
Email: r.heery@ukoln.ac.uk
Address: Bath, BA2 7AY

Tel: 01225 386580
Fax: 01225 386838

Length of Project: **12 months**

Project Start Date:	1 March 2007	**Project End Date:**	29 February 2008

Total Funding Requested from JISC:	£110,000

Funding Broken Down over Financial Years (April – March):

Apr06 – Mar07	Apr07 – Mar08	Apr08 – Mar09
£9,167	£100,833	

Total Institutional Contributions:

Percentage Contributions over the Life of the Project:	**JISC 75%**	**PARTNERS 25%**

Outline Project Description

Enhanced Tagging for Discovery, a Semantic Interoperability Demonstrator, will investigate the combination and comparison of controlled and folksonomy approaches to semantic interoperability supporting resource discovery in repositories and digital collections. This project will demonstrate and evaluate the combination of both approaches in the context of repositories and digital collections, attempting to get the best of both worlds.

The project will investigate deploying demonstrators with two communities of use: at Intute, focusing mainly on tagging by readers (postgraduate users annotating resources with tags); and at CCLRC, focusing mainly on tagging by authors (when they deposit in the repository).

The project will evaluate the results by a range of quantitative and qualitative measures. It will evaluate scenarios for indexing and search. It will make recommendations with reference to resource discovery in digital repository networks

I have looked at the example FOI form at Appendix A and included an FOI form in the attached bid (Tick Box)	**YES**	

I have read the Circular and associated Terms and Conditions of Grant at Appendix B (Tick Box)	**YES**	

Enhanced Tagging for Discovery

1. Introduction

This is a proposal for **Enhanced Tagging for Discovery**, a Semantic Interoperability Demonstrator in response to the JISC Capital Programme call, Repositories and Preservation, Strand (C), Discovery to Delivery Projects. The project will be led by **UKOLN, University of Bath** partnered by **University of Glamorgan , CCLRC, Intute (MIMAS, The University of Manchester)**, with unfunded support from **OCLC Office of Research** and the **Royal School of Library and Information Science, Denmark** (expenses only). Funding is sought for **£110,000** with a planned **one year** duration from 1st March 2007 to 28 February 2008. The partners have internationally recognised expertise in terminology services technology and development, combined with academic and commercial experience applying Knowledge Organisation Systems (KOS) to indexing, search and discovery. Additionally, they bring communities of use to play an active role in creating tags and providing feedback on demonstrators.

Earlier this year, UKOLN carried out JISC commissioned reports relevant to this project: The Shared Infrastructure Synthesis Review and, in partnership with the University of Glamorgan, the Terminology Services and Technology Review. All partners add rich and long-term experiences with relevant national and international activities and projects. This project will be undertaken by staff who are **in post** and are experts in the field.

The demonstrator requested by this call will **investigate the combination and comparison of controlled and folksonomy approaches** to semantic interoperability in the context of repositories and digital collections.

Use of **controlled vocabularies** has been shown to provide benefits for search and discovery and to enable access via browsing and navigation. In the context of repositories, the addition of controlled vocabularies to repository content might be used as basis for effective layering of a subject view (or aggregation) over institutional repositories. Vocabulary control aims to reduce the ambiguity of natural language when describing and retrieving items. The semantic relationships in structured Knowledge Organisation Systems (KOS) provide pathways to connect a searcher with an indexer or author's choice of terminology and to facilitate mapping and semantic interoperability between different information systems. The semantic structure can also provide guidance to an indexer in deciding what aspects or facets to index. However, there are costs associated with use of controlled vocabularies – manual indexing is a significant resource, especially when performed by trained indexers.

On the other hand, **social tagging** is less concerned with consistency than with making it easier for end-users to describe information items and to have access to other users' descriptions. Social tagging applications, such as Flickr and Del.icio.us with their community-based user interfaces encouraging social tagging activity, currently attract much attention and are seen as key elements of new Web 2.0 services. They hold the promise of reducing indexing costs by drawing end-users into contributing this resource, adding value as part of their interaction with information services. However they have yet to be evaluated for JISC purposes and existing social tagging applications have not been designed with information discovery and retrieval in mind. The resulting folksonomies are completely uncontrolled, lacking even basic control of word forms, spelling, synonyms and disambiguation of homonyms.

This project will demonstrate and evaluate **the combination of both approaches in the context of repositories and digital collections**, attempting to get the best of both worlds. Can vocabulary control and the use of an established classification, such as the Dewey Decimal Classification (DDC), improve the usability and utility of free tagging for resource discovery and interoperability for general users of an information system? Can the folksonomic tag clouds be seeded with KOS elements, so as to produce a hybrid?

The combined approach could be used at various stages of the metadata creation workflow:
- by an author creating original metadata at time of deposit
- by a reader annotating (for colleagues/world or for recommendation for inclusion in a collection)
- by an intermediary
- by an automatic process at any stage

Due to constraints on time, this project will focus on the first two options for social tagging activity: **the user as author and the user as reader.**

The project will investigate deploying demonstrators with two communities of use: at Intute, focusing mainly on tagging by readers (postgraduate users annotating resources with tags)**;** and at CCLRC, focusing mainly on tagging by authors (when they deposit in the repository). **Two different KOS**, appropriate to the different use contexts, **will be combined with social tagging**. The DDC will be employed in context of Intute: Social Sciences. Value will be added from participation by OCLC Office of Research. OCLC Research will provide access to termninology resources including the electronic DDC 22 file with its deep vocabulary content and rich terminology

resources of synonyms and mappings. OCLC Research will provide access to Dewey Web Service APIs and will collaborate on their incorporation into the Demonstrator. In a complementary project at CCLRC, PACS (Physics and Astronomy Classification Scheme) will be employed in the context of the CCLRC ePublication Archive.

The project will evaluate the results by a range of quantitative and qualitative measures. It will evaluate scenarios for indexing and search. It will make recommendations with reference to resource discovery in digital repository networks. It is envisaged that the work will benefit development of the Information Environment and the Intute Repository Search Service. The project will input service descriptions to the e-Framework. Outcomes from the project will also inform advice given by the Repositories Support Project

2. Project Description

The aim of the project, as set out in the call is "to build a demonstrator to **test the combination and comparison of controlled and folksonomy approaches to semantic interoperability**." For the purpose of resource discovery, we wish to explore how to optimise social tagging systems for discovery and how to combine them in different ways with more controlled knowledge organisation and retrieval systems.

We aim to investigate whether vocabulary control and the use of an established KOS can assist in moving free social tagging beyond personal bookmarking to aid resource discovery in context of the JISC Information Environment and eFramework. The project will demonstrate use of tagging in different environments (within Intute and CCLRC) and will provide an interface that enables use of a traditional classification scheme to enhance free form tags. The project will consider issues, such as whether prompting with controlled terminology is beneficial. Is it possible or desirable to educate or guide the user to improve their 'tag literacy'? Or will this destroy the enthusiastic anarchy on which tagging occurs (Guy and Tonkin 2006)?

2.1 Initial review and requirements analysis

Although many sources are familiar from the Terminology Services Review, the project will briefly review existing social tagging literature and publically available software for tools and also methods of application. Leading applications, such as Flickr, Del.icio.us and Connotea (e.g. Hitchcock 2006; Lund et al. 2005), will be reviewed for project relevance. Particular attention will be paid to user interface visualization and prompting techniques with end-user supplied tags. For example, the (UKOLN) Guy and Tonkin D-Lib article reviews some common visualization methods, analyses problems with social tagging and proposes some methods for improvement. Trant (2006) discusses some of the first tests of tagging behaviour in the cultural heritage domain, as part of the steve.museum project.

The project will track and collaborate with, as appropriate, the two projects funded in the April round (Image Data Web and Semantic Tagging - we have good connections with these project teams). The project's findings on the combination of social tagging and structured Knowledge Organisation Systems, together with the terminology web service development, will be relevant to the e-Infrastructure programme (and September call) on knowledge organisation and semantic services and developments there will also be tracked.

Scenarios will be constructed in consultation with stakeholder groups to explore how combination of controlled vocabularies and social tagging might be introduced into the workflow to enhance resource discovery across repositories.

2.2 Demonstrator development

Informed by the requirements analysis, two experimental hybrid systems will be developed and evaluated. One investigation will involve an extract of Intute records and the development of an experimental hybrid controlled-folksonomy system that accesses DDC via calls to OCLC Research web services. Here, the user as reader will be studied. A complementary study at CCLRC will focus on the user as author.

Due to the limited resources available for the project, the larger development effort (at Glamorgan) will focus on developing a demonstrator that will be deployed in collaboration with Intute, on an Intute dataset with users from a postgraduate network applying tags. Glamorgan (with technical support from OCLC researchers) will implement a rich hybrid demonstrator, which will investigate different approaches to combining free end-user tags with a structured classification system. This will have access to the electronic DDC 22 file and mapped terminology, via OCLC Research terminology web services. A complementary study will be conducted at CCLRC, with a more limited development, extending the current author tagging system along social tagging lines. This will allow a general comparison with a repository context, a different controlled vocabulary and interface and a different user community. Conclusions, guidelines and implications for discovery will be drawn from both studies.

Since the first step must be to collect tagging data within the context of the JISC IE, the main experimental focus will be on gathering and analysing the different types of cataloguing and indexing data, along with an evaluation of usability issues of the hybrid tools. However, the aim of the analysis and conclusions will be towards the implications for search and discovery, taking account of cost benefit issues. These will be expressed via recommendations, options, guidelines, and search/discovery scenarios with analysis of retrieval implications of the different options considered. The existing search systems will be available in both studies.

2.3 Intute demonstrator

Project partner Intute will provide access to a selection of material from the government and politics sections of Intute: Social Sciences: http://www.intute.ac.uk/socialsciences/cgi-bin/browse.pl?id=120086 http://www.intute.ac.uk/socialsciences/cgi-bin/browse.pl?id=120104
OCLC Office of Research, as unpaid collaborators, will provide access to the electronic DDC 22 file, with its deep vocabulary content and rich terminology resources of synonyms and mappings. They will allow access to Dewey Web Service APIs and, via technical collaboration, facilitate their incorporation into the Demonstrator. This will allow very flexible use of DDC in the hybrid system and two way connection between the controlled vocabulary and project folksonomy. The potential of re-exporting the resulting indexing to Intute will be considered if appropriate.

The Intute dataset will be extracted to form a closed collection for purposes of the project. Existing indexing and DDC will be stripped (a copy with original indexing will be retained for evaluation purposes later), so end users can tag resources from an unindexed collection. Users will 'annotate' with tags – using their own free tags and DDC. This may enable a 'community' feel to develop as users annotate content in relation to their shared terminology. Intute will provide access to specialist subject groups, such as the Political Science Association's postgraduate network to trial the experimental hybrid system.

The experimental system will take as a starting point the existing operational myIntute tagging and simple search system http://www.intute.ac.uk/myintute/index.php and significantly extend it. This software connects to the Intute database and has a limited single user interface (with tag cloud). It is part of the broader myIntute personalization system. Intute will provide the open source code. Technical support will be provided by the myIntute developer, Chris Stephens. The current user interface will be completely reengineered for purposes of the project. Glamorgan will develop an experimental hybrid system operating on the extracted collection that allows both free social tagging and various combinations of social tagging with DDC. It will be applied to the extracted Intute dataset and a new community-based interface, with different visualizations and tagging options, will be developed. The system will connect to the Dewey classification system using OCLC Research Dewey Web service calls (supported by OCLC research staff members). We will experiment with different options, selecting from possibilities such as:-

Starting from DDC
- initially prompting with controlled terminology
- via OCLC Research web services, develop visualizations and navigation of DDC to show terms in their context
- seeding tag clouds from DDC
- investigate forms of structuring and ranking 'tag clouds' other than by popularity, for example using DDC as a structuring mechanism for tag clouds and other interface visualisations

Starting from user free tags
- improvement of user tags: spelling correction, synonym linking, singular/plural linking, treatment of compounds, etc.
- a 'suggest other terms' option drawing on DDC (as part of free tagging), for example mapping user tags to DDC entry vocabulary (and more loosely mapped terminology) at time of entry, thus allowing DDC suggestions (with different degrees of confidence)

post processing
- mapping of user tags to DDC, for purposes of expanding DDC entry vocabulary and also for further analysis

In short, we will develop a workbench of services that we can combine to investigate different aspects – comparing and combining KOS/Social tagging.

2.4 CCLRC demonstrator

The CCLRC institutional repository, ePubs http://epubs.cclrc.ac.uk/ covers the scientific output of CCLRC, and all the areas of science supported by the Lab. The main disciplines are in physics, materials science and chemistry, with some environmental science, computing, mathematics etc. One classification option offered is PACS (Physics and Astronomy Classification Scheme).

A case study with author end-users will be employed, with support from the CCLRC Library and Information Service. A limited extension to the current author indexing interface (currently allowing free keywords to be entered) will allow
 - a facility for allowing reader keywords.
 - a "tag cloud" based search interface (a corpus of suitable initial tags already exists).
 - an interface to a keyword prompting service (from PACS)
 - monitoring of use of the tagging/prompting system

The potential of SKOS as a common representation for vocabularies will be explored in the context of the project, along with consideration of the appropriate protocol to communicate with the vocabularies (see Terminology Services Review, section 6). The evaluation and analysis will follow the same general framework as Intute, as far as feasible. It will consist of a subset of the options, making allowances for the differences between the two studies

and in the integration of the demonstrator with the classification scheme, since potentially there may be a more limited range of connections.

2.5 Evaluation and analysis

The main focus of investigation in both studies will be the effect of the combined hybrid system: free tagging with no instructions versus tagging using a hybrid system and guidance for users.

Final details of the study methodology will be decided after examination of the available datasets at the start of the project and an initial pilot evaluation. Potential biases (such as learning effects) will be taken into account together with precise tasks, selection of any user subsets, the context (purpose) for tagging, degree of freedom/control in user tasks, etc. We will consider specification of the context of tagging: tagging for recommendation, tagging as part of book marking, tagging as part of information searching and relevance assessment, etc. There will also be an evaluation of the usability of the hybrid tool and its interface, navigation of structured KOS, perceived complexity of the KOS. A major focus of the analysis will be a quantitative comparison of the outcomes of the different tagging activities. This will be complemented by some qualitative evaluation, which will include follow-up interviews with taggers to get context and background information. Descriptive statistics will be analysed and account taken of cost benefit issues – eg, various factors are considered in the steve.museum social tagging study in the recent NRHM KOS issue (Trant 2006).

The particular research questions to investigate will be identified in the initial stages of the project. Possible factors to consider include differences between :-
- the effect of instructions to the taggers (and compliance with instructions),
  such as rules concerning exhaustivity, specificity, tagging for classification Vs indexing
- differences in results regarding exhaustivity and patterns in selection of facets
- differences in results regarding specificity,
- differences in results tagging for classification versus indexing
- number of errors in topic assignment from tagging
- nature of topics chosen (main, secondary; document, domain, user-oriented) - nature of terms employed (academic, layman, slang, abbreviations, special languages)

We will also have the original cataloguing data available for comparison in the analyses. Post processing of data and mapping to the classification scheme in question will be considered. For example, the Intute project potentially will allow some comparison of (all or some of)
- indexing of end-user free social tagging
- indexing of end-user controlled/improved tagging and DDC classification
- indexing of librarian expert cataloguers free keywords
- indexing of librarian expert cataloguers using DDC

2.6 Contextualising results in repositories context

Results of the evaluation will be interpreted to consider interoperability implications in the context of an inter-working network of repositories, with particular focus on the current UK deployed base of scholarly eprint repositories aggregated by the Intute Search Infrastructure project. The analysis will report on the potential for providing enhanced subject access to repository content. The implications for institutional repositories will consider feasibility of partitioning repositories using assigned controlled vocabulary terms. The implications for aggregators of repository content will also be considered, in terms of potential for layering subject access onto aggregated content. The costs and benefits of different approaches in the two demonstrators will be considered from a qualitative perspective. A briefing paper will be produced aimed at repository managers and developers.

3. Project Plan

Full account will be taken of issues relating to accessibility of Web-based systems and software and the outputs of this work will conform to published standards and guidelines.

3.1 Project Deliverables and Timetable

Major Deliverables	Month	Lead Effort
Project plan	1	**UKOLN** + partners
Demonstrators with Intute data	7	**GLAM** + partners
Demonstrator with CCLRC data	8	**CCLRC** + partners
Evaluation reports	11	**GLAM +** partners
Final report	12	**UKOLN** + partners
Dissemination	3-12	**UKOLN** + partners
Project management	1-12	**UKOLN** + partners

To carry out the work of the project, the following Work Packages are planned:

Work Package 1 - Project Management (Lead UKOLN)
Deliverables (1-12): Project Plan - Website - Progress Reports - Final Report
Project management and partner co-ordination will be provided by UKOLN and will be achieved by an initial project start-up meeting, a mid-term meeting and a closure meeting. Communication between partners will be supported by email, conference calls and informal methods.
Project staff will work in partnership with members of relevant JISC Development teams, provide progress reports as required and participate in programme evaluation activities.

Work Package 2 – Requirements analysis (Lead UKOLN)
Deliverables (1-4): Requirements summary report - Evaluation methodology summary
UKOLN and GLAM: review social tagging literature and software
Scenarios and use cases will be formulated and circulated to Project Research Team.
Initial requirements gathering from stakeholders:
- user groups at Intute and CCLRC
- service managers at Intute and CCLRC

UKOLN, GLAM and RSLIS: design evaluation methodology (and initial pilot)

Work Package 2 – Software development (Lead GLAM)
Deliverables: Intute and CCLRC demonstrators
Intute: select and extract datasets; select and set up access to user groups
CCLRC: select and extract datasets; select and set up access to user groups
OCLC: provide documentation and access to OCLC terminology web services

GLAM (1-12): Design and implement hybrid demonstrator and user interface, accessing DDC via OCLC's terminology web services and investigating different options for linking DDC to social tagging. Subsequent basic search system for purposes of illustrating search scenarios. Software development necessary for post-processing social tagging mapped to DDC.
Intute and OCLC: Provide technical assistance.

CCLRC (6-8): Extend current CCLRC author tagging system along social tagging lines, applying the Glamorgan social tagging user interface to the PACS system. CCLRC will also explore the development of a new search interface based on the use of "tag clouds".

Work package 3 – Evaluation and analysis (Lead GLAM)
Deliverables (7-11): Evaluation analysis report
Conduct initial pilot to evaluate methodology. Evaluate demonstrators at Intute and CCLRC. Gather quantitative and qualitative data and analyse. Further details given in the project description.
SLIS: Provide assistance in methodology and analysis.

Work package 4 - Write up and analyse results (Lead UKOLN)
Deliverables (11-12): Briefing paper on implications for repositories – Recommendations to JISC - search/discovery scenarios

Work package 5 – Dissemination (Lead UKOLN)
Deliverables (3-12): web site – presentation at conference – participation in JISC programme events –journal article
The project will work with the JISC programme to disseminate results in a timely fashion. The project web site will be kept up to date with progress on deliverables. Project partners will participate in JISC programme events, and will disseminate through conferences.

3.2 Risks

Risk	Level	Likelihood	Contingency
Recruitment difficulties	Medium	Low	Existing staff will work on study
Loss of a team member	High	Low	Multiple staff at each site have the expertise and skills required.
Project is over-ambitious in scope and/or over-runs	Low	MediumLow	Agree scope with JISC by means of project plan

4. Impact

Outcomes of the project will inform future enhancement of subject access to repositories and digital collections. There is a challenge facing institutional repositories, and aggregators of institutional repository content, as to how to layer 'subject views' onto the aggregated content of repositories. The project outcomes will inform the community of potential usefulness of the combined social tagging and controlled vocabulary approach.

The collaborative proposal addresses Welsh priorities concerning promotion of research capability and collaboration. Its outcomes will support the development of e and distance learning/research through the enhancements to current Repository and Intute capabilities concerning the combination of social tagging and controlled vocabularies, such as DDC.

CCLRC serves the wider research community as well as its own scientists, so improvements to its services will benefit the wider community. The project will enhance the usability of the CCLRC ePublications archive with a novel and usable facility which will make the discovery of relevant new science simpler for the scientists using the facilities of CCLRC and its wider research community, thus encouraging the development of new science. As appropriate, this work will be transferred to the other data archives and catalogues provided by CCLRC, notably the data catalogues provided to support the large-scale scientific facilities of the CCLRC (e.g ISIS Neutron source, Diamond Light Source, Central Laser Facility), again potentially adding to the usability of these facilities and encouraging new science.   The work of the project will be used to inform the ongoing standardisation effort of the W3C to produce the SKOS recommendation for Knowledge Organisation Systems exchange and interoperability.

The resulting experimental system will be available for Intute to apply/adapt/reengineer, subject to continued availability of the Dewey web services and Dewey licence. Similarly CCLRC may decide to further develop their demonstrator system to embed it into their production services.

5. Intellectual Property

The project will comply with the terms of the JISC Funding Agreement. The IPR of material generated as part of the project will remain with the respective creators. All outputs, including documentation and code, created during the fulfilment of this project will be disseminated to the wider HE community with the expectation that it will be made freely available under an appropriate open source or creative commons licence as appropriate. Outputs will be made publicly available in a timely manner to ensure current information about the project is available throughout its life.
Budget

The proposed project budget has been constructed on a full economic cost (fEC) basis using the Transparent Approach to Costing (TRAC).

As a demonstrator project, the quantitative benefit for the main partners (UKOLN and Glamorgan) is of less importance than the qualitative benefits. These partners will benefit from the complementary fit of this project with other work within their departments related to terminology services, registries and repositories. Intute and CCLRC will benefit from evaluation of the demonstrator informing future development of their services. On these grounds the proposal requests a total budget which represents just 75% of fEC.

6. Partnership and expertise

This bid is led by UKOLN at the University of Bath. Project management will come from UKOLN and be carried out by Traugott Koch, with leadership and direction from Rachel Heery as UKOLN Research team leader. Traugott will also carry out research focusing on requirements, design of methodology, and analysis. Traugott has internationally recognised expertise in terminologies and subject access, as outlined below. Work on this project will be complemented and informed by UKOLN's involvement in other repositories projects, including the Intute Search Infrastructure Project, the Repository Support Project, and the Repositories Research Team. In addition,

UKOLN have proven expertise in management of successful projects.

Development work will be led by Glamorgan, with Douglas Tudhope acting as technical lead. Douglas has experience of leading successful development projects, including the EPSRC funded FACET project, which investigated thesaurus-based query expansion, in collaboration with the Science Museum and Museum Documentation Association. His area of research is Knowledge Organisation Systems and Services and he was an author of the recent JISC State of the art review on Terminology Services and Technology. He supervised Jim Moon's PhD at Glamorgan, which saw major object-oriented software development. Jim will act as the main software developer. He researches and teaches software engineering at Glamorgan and has previous commercial development experience.

Brian Matthews leads the work at CCLRC and has twenty years experience of Computer Science projects at UK and European level with some 70 publications. He has been involved in EC projects on digital repositories and the Semantic Web, including the Limber project on Social Science, and the SWAD-Europe (Semantic Web Advanced Development in Europe) where he lead the effort on KOS. Alistair Miles founded the SKOS (Simple Knowledge Organisation Systems) community, an open collaboration of semantic web and knowledge organisation experts, and has led the further development of standardisation initiatives within the domain of knowledge organisation and the semantic web, extending the work done within the remit of SWAD-E.

Further support will be provided by Chris Stephens (the developer of myIntute) and Marianne Lykke Nielsen, who has significant commercial and research experience of KOS and vocabulary focused user studies. OCLC researchers are engaged in several projects involving the DDC and other terminology resources, including
• DeweyBrowser – Supports searching and browsing of collections of library resources organized by the DDC
• Terminology Services project -- provides access to multiple terminologies as Web services. The services are available as production services and as research prototypes.
• FAST – Faceted Application of Subject Terminology – A rich controlled vocabulary based on the terminology of Library of Congress Subject Headings (LCSH) that employs a simplified application syntax.

This partnership provides a strong basis for successful outcomes, combining teams with proven expertise in relevant areas..

UKOLN, University of Bath
Rachel Heery is Deputy Director, she leads the research team and works on the Repository Research team and the repository support Project. She supervises staff on various JISC projects including shared infrastructure services, IEMSR and IESR. In the past she has worked on repository projects including the Open Archives Forum and ePrints UK, and has co-authored a repositories review and roadmap, and a terminologies services review for JISC. She is UKOLN's representative on the EC DRIVER project and has worked with the JISC eFramework initiative.

Traugott Koch works as a Research Officer. He has carried out many digital library R&D projects during the last fifteen years at European, Nordic and national levels. His areas of special expertise are knowledge organization, terminologies and subject access, semantic interoperability, metadata and resource discovery and retrieval. He is involved in the development and maintenance of Dublin Core Metadata Initiative standards, has published more than 60 scholarly papers, edited one book and five journal special issues. He is a co-author of the DELOS Semantic Interoperability report and the JISC Terminology Services and Technology review. Since their beginning, he has been active in both the US and European NKOS network and has acted as organiser and committee member for many NKOS workshops and activities. At UKOLN Traugott has been working in the IESR and eBank UK projects, and has co-authored the JISC Terminologies Services Review.

University of Glamorgan
Douglas Tudhope is Reader in the Faculty of Advanced Technology, University of Glamorgan and leads the Hypermedia Research Unit. He was PI on the EPSRC FACET project and is PI on a forthcoming AHRC project (STAR) in collaboration with English Heritage. He is Editor of the journal, New Review of Hypermedia and Multimedia and acting Theme Editor, Information Discovery, Journal of Digital Information (JoDI). He has co-organised 7 workshops on Networked Knowledge Organisation Systems/Services at ECDL, JCDL, DCMI conferences. He has over 50 refereed publications.

Jim Moon is a Senior Lecturer in Software Engineering at the University of Glamorgan. He has a PhD from Glamorgan. His research interests include the Unified Modelling Language (UML), software engineering, Multi-Agent Systems and simulation. Jim has developed software for commercial full mission ship simulators, UML tutoring systems and RFA software that has obtained a leading position in lower bounding work using dynamic tabu search (world class results for Cost259 at "FAP web").

Intute
Chris Stephens is a software developer for Intute and works for the Humanities Computing Unit at Oxford University. He has developed various web based projects, such as the Humbul system including the development

of a personalization service. He has developed a personalization system for the new cross-discipline Intute site (myIntute) and continues to support and develop this. He is interested in ways of giving users a stake in the site and personal reuse of Intute resources.

CCLRC
Brian Matthews is the Group Leader of the Information Management Group within the eScience Centre, leading a research and development team with projects in Digital Libraries, Semantic Web, Digital Curation and Grid Middleware. He is a co-investigator on the JISC project Claddier, and is part of the team involved in the European Project CASPAR on Digital Curation. He is deputy manager of the W3C Office for the UK and Ireland, and a part-time lecturer at Oxford Brookes University.

Alistair Miles has an M.A. in Natural Sciences from the University of Cambridge. He is a research associate in the Information Management group at CCLRC. Since joining in 2002 he has worked on the Semantic Web Advanced Development for Europe (SWAD-E) project as co-ordinator and lead developer for the Thesaurus Activity work package. He was co-ordinator of the W3C Semantic Web Best Practices and Deployment Working Group's Thesaurus Task Force, and the principal author of both the SKOS Core Vocabulary Specification and Guide, published as W3C Working Group Notes,and is now editor of the forthcoming W3C Recommendation on SKOS.

Royal Library School, Denmark
Marianne Lykke Nielsen is associate professor at the Royal School of Library and Information Science. Marianne has agreed to collaborate on an expenses only basis, with particular regard to indexing and evaluation issues. She lectures and researches on the design and evaluation of systems for knowledge organisation. She is co PI in the US NSF Pathway Project that investigates user-centred indexing methods based on semantic components of documents. She recently finished a comparison study of controlled and free indexing in an enterprise setting, in collaboration with and financed by H. Lundbeck (pharmaceutical industry). She has acted as consultant for many corporate and government organizations, designing or evaluating metadata systems, taxonomies and thesauri. She is collaborating on evaluation and user aspects of the Glamorgan STAR project, co-edited the 2006 NRHM special issue on KOS and co-organised NKOS workshops in 2004, 2005 and 2006.

Key Contact
Rachel Heery   Tel: 01225 386580
Deputy Director   Fax: 01225 386838
UKOLN   Email: r.heery@ukoln.ac.uk
University of Bath
Bath, BA2 7AY   http://www.ukoln.ac.uk

7. References

CCLRC institutional repository, ePubs http://epubs.cclrc.ac.uk/
Connotea http://www.connotea.org/
Delicious http://del.icio.us/
FACET Project. http://www.comp.glam.ac.uk/pages/research/hypermedia/Facet/facetproject.html
Flickr http://www.flickr.com/
Guy M., Tonkin E. 2006. Folksonomies: Tidying up Tags? D-Lib Magazine, 12 (1).
    http://www.dlib.org/dlib/january06/guy/01guy.html
Hammond T., Hannay T., Lund B., Scott J. Social Bookmarking Tools (I): A General Review. D-Lib Magazine, 11(4), 2005. http://www.dlib.org/dlib/april05/hammond/04hammond.html
Hitchcock H. 2006. Nature expands visibility of EPrints. EPrints news article.
    http://www.eprints.org/news/features/connotea.php
Intute www.intute.ac.uk
Lund B., Hammond T., Flack M., Hannay T. 2005. Social Bookmarking Tools (II): A Case Study - Connotea. D-Lib Magazine, 11(4), 2005. http://www.dlib.org/dlib/april05/lund/04lund.html
myIntute system http://www.intute.ac.uk/myintute/index.php
OCLC Terminology Services. http://www.oclc.org/research/projects/termservices/
Trant J. 2006. Exploring the potential for social tagging and folksonomy in art museums: proof of concept. New Review of Hypermedia and Multimedia, 12(1), 83-105.
SKOS Core Vocabulary Specification, W3C Working Draft , Editors:Alistair Miles , and Dan Brickley ,
    http://www.w3.org/TR/swbp-skos-core-spec

Appendix A: Project partners

Appendix B: Letters of Support

Letters of support from all partners.

Appendix A: Project Partners

UKOLN, University of Bath
UKOLN is a national centre of expertise in digital information management. It provides services to the library, information and cultural heritage communities. UKOLN currently hosts the Repositories Research Team, who support the JISC Digital Repositories Programme and the Repositories and Preservation strand of the Capital Programme. UKOLN is also a partner in the Digital Curation Centre and has either led or is a partner in the following projects: ePrints UK (JISC), Intute Repository Search (JISC), Repository Support Project (JISC), e-Bank (JISC), Grand Challenge in Engineering (EPSRC), Driver (EU) and Delos Digital Library Network of Excellence (EU).

Hypermedia Research Unit, University of Glamorgan
The Hypermedia Research Unit (HRU) has been working in the area of Knowledge Organisation Systems from 1991 and is part of the Faculty of Advanced Technology, at the University of Glamorgan. The HRU is a member of the EC FP6 DELOS Network of Excellence on Digital Libraries and participates in activities within Cluster 5 on Knowledge Extraction and Semantic Interoperability. There are currently three full-time members of staff, two research fellows and several PhD students in the unit.

Royal School of Library and Information Science is the Danish higher education institution, responsible to the Danish Ministry of Cultural affairs for education, research and development to the highest level in the field of Library and Information Science (LIS). The school is the only institution in Denmark that offers courses leading to the qualifications of Bachelor and Master of Library and Information Science, Librarian, and PhD. The total number of students enrolled is about 1,000, and the school has some 70 full-time academic staff members.

CCLRC
The Council for the Central Laboratory of the Research Councils is one of the 8 UK research councils with responsibility for the provision of facilities and services to the UK academic community. The eScience Centre within CCLRC provides high-performance computing services, library and information services, advanced software development and research into IT to support the science programme of the CCLRC. The research and development programme includes research into Digital Libraries (inciuding support and development of the CCLRC Institutional Repository, ePubs), Digital Curation and the Semantic Web. The eScience Centre hosts the W3C Office for the UK and Ireland and has been leading the development of the W3C Recommendation for the SKOS format for Knowledge Organisation Systems.

OCLC Office of Research
OCLC is a supporting, non-funded partner. OCLC Research is one of the world's leading centers devoted exclusively to the challenges facing libraries in a rapidly changing information technology environment. Since its beginning in 1978, the Office has investigated trends in technology and library practice to identify technological advances that will enhance the value of library services and improve the productivity of librarians and library users.