



Project Plan Cover Sheet

Project Information			
Project Acronym	EnTag		
Project Title	Enhanced Tagging for Discovery		
Start Date	1 September 2007	End Date	31 August 2008
Lead Institution	UKOLN - The University of Bath		
Project Director	Michael Day		
Project Manager & contact details	Koraljka Golub UKOLN 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; Intute (MIMAS The University of Manchester); STFC Non-funded supporting partners: OCLC Office of Research, USA; Danish Royal School of Library and Information Science		
Project Web URL	http://www.ukoln.ac.uk/projects/enhanced-tagging/		
Programme Name (and number)	Repositories and preservation programme		
Programme Manager	Balviar Notay		

Document Name			
Document Title	Project Plan		
Author(s) & project role	Project Team		
Date	29 January 2008	Filename	EnTag- PPlan-v2.doc
URL	http://www.ukoln.ac.uk/projects/enhanced-tagging/documentation/		
Access	<input type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
Draft 1	21 January 2008	Submitted to Programme Manager
Draft 2	29 January 2008	Submitted to Programme Manager



JISC Project Plan for EnTag

Overview of Project

1. Background

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.

Social tagging applications, such as Flickr (<http://www.flickr.com/>) and Del.icio.us (<http://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, 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. Existing social tagging applications have not been designed with information discovery and retrieval in mind. The resulting folksonomies (tags collections) are completely uncontrolled, lacking even basic control of word forms such as spelling variants, synonyms and disambiguation of homonyms. Many users use tags only to organise own documents, and not to help the community. On the other hand, natural language tags could cover aspects that are not available in a controlled vocabulary, especially when it comes to new concepts; as such, they could help update the controlled vocabulary. Taking all these aspects into account, potential benefits of social tagging for JISC purposes need to be evaluated.

2. Aims and Objectives

2.1. Aims

The project will investigate the combination and comparison of controlled and folksonomy approaches to semantic interoperability in the context of repositories and digital collections.

2.2. Objectives

The project will:

- Investigate indexing aspects when using only social tagging versus when using social tagging in combination with a controlled vocabulary;
- Investigate above in two different contexts: tagging by readers and tagging by authors; and,
- Investigate influence of only social tagging versus social tagging with a controlled vocabulary on retrieval.

3. Overall Approach

The main focus of investigation will be the effect of the combined hybrid system: free tagging with no instructions versus tagging using a hybrid system and guidance for users. The larger development effort will be at Glamorgan in collaboration with Intute and OCLC. Intute will provide a dataset and users from a postgraduate network, and OCLC access to electronic DDC 22 file with technical support. Glamorgan will implement a rich hybrid demonstrator, which will investigate different approaches to combining free end-user tags with the structured classification system.

A complementary study will be conducted at STFC, with a more limited development, extending the current author tagging system along social tagging lines. Here the focus will be on the user as author.

The two studies will allow a general comparison of a repository versus digital collection context, a different controlled vocabulary and an interface and a different user community. Two major methods predicted to be used to collect user data are a questionnaire and log analysis. The results will be evaluated by quantitative and qualitative approaches, both in the context of indexing and retrieval. Evaluation will include the usability of the hybrid tool and its interface and navigation of structured KOS.

Taking account of cost benefit issues, conclusions, implications and recommendations for both indexing and searching will be drawn from the studies. Potential biases (e.g., learning effects, interface design) will be taken into account together with selection of users and data, the context (purpose and motivation) for tagging, degree of freedom/control in user tasks, scenarios and tasks definition, etc. Use scenarios will be collected from the existing users of Intute: Social Sciences. 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.

Critical success factors include the following: reaching conclusions as to whether including a controlled vocabulary in social tagging activities improves indexing, implications as to whether it enhances retrieval, and gaining implications of the interface design for enhanced tagging both in terms of indexing and retrieval.

4. Project Outputs

WP	Tangible Deliverables	Knowledge and Experience to Build and Share
WP2	Review of social tagging literature and software.	Knowledge gained from literature on social tagging especially related to enhancing tagging with controlled vocabularies.
	Scenarios and use cases.	Examples of use scenarios in Intute and STFC contexts.
	Designed user study and evaluation methodology.	
WP3	Intute demonstrator.	
	STFC demonstrator.	
WP4	Finalized methodology based on pilot testing.	Suggested methodology for similar studies.
	Quantitative and qualitative data from user studies.	
WP5	Paper reporting analysis and results.	Knowledge of KOS enhanced tagging behaviour in the two contexts, both in terms of indexing and searching.

5. Project Outcomes

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. Outcomes of the project will inform future enhancement of subject access to repositories and digital collections, especially as to potential usefulness of the combined social tagging and controlled vocabulary approach. These enhancements will in turn better support the development of e- and distance learning and research.

STFC 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 STFC e-Publications archive with a novel and usable facility which will make both the indexing and the discovery of relevant new science simpler for the scientists using the facilities of STFC and its wider research community, thus encouraging the development of new science.

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

Social tagging is very popular today and users seem to enjoy it; if it at the same times proves to be possible to enhance it for improved retrieval, indexing costs would be lowered and information discovery would be enhanced. This could lead to larger acceptance of controlled vocabularies among computer science researchers developing search engines, which could then want to utilize them more and improve information discovery even further.

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
Digital librarians and other information professionals	Ability to reduce indexing costs and improve retrieval.	high
End users	Improved subject access to information.	medium
Discovery and delivery service providers	Improved ways to reduce indexing costs and improve retrieval. Guidance on how to build systems to support the above.	high
Information scientists	New findings related to subject access to information.	medium
Standards organizations and bodies in charge of controlled vocabularies	New ways for updating and maintaining controlled vocabularies.	medium

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing - Recruitment difficulties	2	3	6	Existing staff will work on study
Staffing - Recruitment difficulties for users	2	4	8	Involve Intute and STFC repository staff to facilitate access
Organisational – loss of team member	1	2	2	Multiple staff at each site have skills and expertise required
Technical - Project over-ambitious and/or over-runs	2	2	4	Agree scope with JISC by means of project plan
Technical – Short timescale for development project	3	3	9	Prioritise development work on key goals. Monitor progress.
External suppliers	N/A	N/A	N/A	N/A
Legal	N/A	N/A	N/A	N/A

8. Standards

Name of standard or specification	Version	Notes
DDC (Dewey Decimal Classification)	22	The most widespread classification scheme, freely available for research
UML	1.4	Most texts currently support version 1.4. Version 2.0 has recently been released and some literature is now available. It is under

		investigation.
SQL	MySQL 5.0	This is a widely used open source SQL server.

9. Technical Development

The short duration of the project means that maximising productivity is essential, while providing software products that are of a high quality and may be used as the basis for further research.

As far as possible, the software is designed using the Unified Modelling Language (UML):

- UML is the standard design language for software systems;

- UML supports Use Case specification, architectural design and behavioural design.

- A wide range of support resources are available including definitive OMG specifications, web resources, texts, developer articles and journals.

Note: UML is primarily a design language for Object Oriented Systems, some extensions have been applied to permit the design of web based applications.

The web applications are being developed using Active Server Pages (ASP) .NET

- ASP .NET provides rapid application development;

- ASP .NET provides a rich class framework;

- ASP .NET permits easy integration of Web Services;

- ASP.NET has a wide tool base supporting technologies such as AJAX and MySQL.

- A wide range of support resources are available for ASP .NET, including web resources, up to date texts, developer articles and journals.

The MySQL database system is being employed:

- MySQL is freely available as an open source database product under the GNU General Public License;

- MySQL is a multi-user, cross-platform product;

- MySQL is a fast database server accessed via robust, proven, Structured Query Language (SQL). SQL is the de facto standard for database access;

- MySQL provides a simple upgrade path to Enterprise;

- MySQL is supported by a very large user group;

- MySQL is a well proven product that meets all our expected database requirements.

10. Intellectual Property Rights

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. OCLC will make available Dewey terminology resources and web services to Glamorgan for the research purposes of the project, in particular as input to inform the demonstrator. IPR of Dewey itself remains with OCLC.

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 license as appropriate.

Project Resources

11. Project Partners

UKOLN, University of Bath

Koraljka Golub leads the UKOLN effort and is responsible for EnTag project management and the project website. She will carry out research focusing on design of evaluation methodology, and analysis.

University of Glamorgan

Douglas Tudhope is Professor in the Faculty of Advanced Technology, University of Glamorgan and directs the main development work, which Glamorgan will lead on.

STFC

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. STFC are sub-contracted to develop and evaluate the STFC demonstrator.

Intute

Debra Hiom, responsible for Intute Social Sciences will coordinate access to Intute data for the project and facilitate access to users for the Intute demonstrate.

OCLC

OCLC will provide access to the Dewey (DDC) and also assist with web services as appropriate, in collaboration with Glamorgan. Diane Vizine-Goetz is main contact.

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, design of the evaluation study and evaluation issues.

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 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-based discussions and further telephone 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 demonstrator leader	dstudhope@glam.ac.uk School of Computing, University of Glamorgan, Pontypridd, CF37 1DL tel: 01443 482271 fax: 01443 482715

Jim Moon	Software developer	jinmoon@glam.ac.uk School of Computing, University of Glamorgan, Pontypridd, CF37 1DL tel: 01443 483603 fax: 01443 482715
Intute		
Debra Hiom	Intute data and users coordinator	d.hiom@bristol.ac.uk Institute for Learning and Research Technology, University of Bristol 8-10 Berkeley Square Bristol, BS8 1HH tel: 0117 928 7117 fax: 0117 928 7112
Chris Stephens	Intute software developer	christopher.stephens@oucs.ox.ac.uk Intute: Arts and Humanities Research Technologies Service Oxford University Computing Services 13 Banbury Road Oxford OX2 6NN United Kingdom tel: 01865 283 347 fax: 01865 273 275
STFC		
Brian Matthews	STFC demonstrator leader	B.M.Matthews@rl.ac.uk STFC Rutherford Appleton Laboratory Chilton, Didcot, Oxfordshire OX11 0QX tel: 01235 446648 fax: 01235 445831
OCLC		
Diane Vizine-Goetz	Supporting officer for DDC	vizine@oclc.org OCLC 6565 Kilgour Place Dublin, Ohio 43017-3395, USA tel: +1 614 764 6084 fax: +1 614 764 6096
Andrew Houghton	Software supporter for DDC	houghton@oclc.org OCLC 6565 Kilgour Place Dublin, Ohio 43017-3395, USA tel: +1 614 764 6003 fax: +1 614 718 7075
Danish Royal School of Library and Information Science		
Marianne Lykke Nielsen	Research officer	mln@db.dk RSLIS Fredrik Bajers Vej 7K 9220 Aalborg Øst, Denmark tel.: +45 98 15 79 22 fax: +45 32 84 02 01

13. Programme Support

Invitations to events on subject access to information.

Invitations to events on social tagging and Web 2.0.

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

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
Month 6	Development of the Intute demonstrator	Is the demonstrator ready to be used in the user study?	Pilot testing	Pilot testing shows that the demonstrator is ready
Month 6	Designing user study for Intute	Is the user study well designed?	Pilot testing	Pilot testing shows that there the study is appropriate and well designed
Month 7	Development of the STFC demonstrator	Is the demonstrator ready to be used in the user study?	Pilot testing	Pilot testing shows that the demonstrator is ready
Month 7	Designing user study for CCLRS	Is the user study well designed?	Pilot testing	Pilot testing shows that there the study is appropriate and well designed
Month 7	User study on Intute	As included in the study	Questionnaires and data logging	All data are collected and properly stored
Month 8	User study on STFC	As included in the study	Questionnaires and data logging	All data are collected and properly stored
Month 11 and 12	Reports	Are issues important to stakeholders addressed?	Check with stakeholders through personal contact	Production of report that represents interests of stakeholders

17. Quality Plan

Output	WP2				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Months 1 to 3	Appropriate sampling of the literature and use cases.	Feedback from other partners and colleagues.	Positive feedback.	Koraljka Golub, Doug Tudhope	
Output	WP3				

Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Months 6 to 7	Demonstrators ready to be used in the user study	Testing	Pilot testing	Doug Tudhope, Brian Matthews, Marianne Lykke Nielsen, Koraljka Golub	
Output	WP4				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Months 7 to 10	Scientific Appropriateness	Using established study methods and sampling	Pilot testing	Doug Tudhope, Brian Matthews, Marianne Lykke Nielsen, Koraljka Golub	
Output	WP5				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Months 10 to 12	Scientific Appropriateness and Stakeholders Interests Covered	Frequent discussions among partners, JISC, and other colleagues	Successful completion of external peer review	Doug Tudhope, Brian Matthews, Marianne Lykke Nielsen, Koraljka Golub	

Feedback and peer review from project partners, people at events and JISC throughout the project.

18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
End of project	Report on the advantages of enhancing and using social tagging, if any	Repositories, digital collections	To inform beneficial developments and motivate buy-in	Whether (enhanced) social tagging can be useful
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 access to demonstrators	

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Knowledge on enhanced social tagging	Further dissemination	Further research in other contexts

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
Demonstrators		Investigate stakeholder's interest	Seek further funding

20. References

Why ASP.NET? <http://www.planetmagpie.com/w3services/asp-net.aspx>

OMG UML resources <http://www.uml.org/>

IBM/Rational UML resources <http://www-306.ibm.com/software/rational/uml/>

MySQL reference manual <http://dev.mysql.com/doc/refman/5.0/en/>

Appendixes

Appendix A. Project Budget

Appendix B. Workpackages

See separate file.