Proposal to Microsoft Corporation for the OAI-ORE Acceleration Project

Summary and Purpose of Proposed Project

The Open Archives Initiative Object Reuse and Exchange Project (OAI-ORE) is developing standardized, interoperable, and machine-readable mechanisms to express information about compound information objects on the web. The OAI-ORE standards will make it possible for web clients, agents, and applications to reconstruct the boundaries of compound objects, the relationships among their internal components, and their relationships to other resources in the web information space. This will provide the foundation for the development of value-added services for analysis, reuse, and re-composition of compound objects, especially in the areas of e-Science, e-Scholarship, and scholarly communication, which are the target areas of OAI-ORE.

OAI-ORE is currently funded for two years (September 2007 – August 2009) by a grant to Cornell University from the Andrew W. Mellon Foundation. The co-directors of OAI-ORE are Carl Lagoze of Cornell University, the principal investigator of this proposal, and Herbert Van de Sompel of the Los Alamos Research Library. Lagoze and Van de Sompel both co-led the highly successful OAI Protocol for Metadata Harvesting (OAI-PMH) project – OAI-PMH is now a de-facto international standard for exchange of structured metadata.

The goal of the Mellon funding to OAI-ORE is the production of a tested and stable protocol specification that has been tested and deployed in target communities. This work is being undertaken by the OAI-ORE technical committee (ORE-TC), an international group of experts in web technologies, library and archive systems, and e-Science. This group is led by Van de Sompel and Lagoze and conducts business via face-to-face meetings, conference calls, and e-mail exchange. The work of this group is advised and reviewed by the ORE Liaison Group, which includes representatives of key international projects (including Savas Parastatidis from Microsoft), and the ORE Advisory Committee.

Experimental deployment of the developing OAI-ORE standards is a vital component of the success of the work. In order to facilitate such experimentation, Microsoft's Technical Computing Initiative under Tony Hey and staff members Lee Dirks and Savas Parastatidis organized an eChemistry Project Planning meeting in April 2007. This meeting brought together Lagoze and Van de Sompel and key members of the chemistry community to explore how OAI-ORE might be deployed to that community and catalyze advances in its scholarly communication and research process. The two-day meeting successfully completed with an agreement on ideas for initial use of OAI-ORE models to represent molecular information and an outline of a proposal for a two-year OAI-ORE eChemistry experimental project starting in 3rd quarter 2007. (A proposal for this project will be submitted to Microsoft in July 2007).

Following the eChemistry Project Planning meeting, contacts were established between Herbert Van de Sompel and Adam Farquhar of the British Library. Adam Farquhar leads the international Planets digital preservation project, and expressed a strong interest in the exploration of the OAI-ORE interoperability layer within the Planets context. Initial conversations with Lee Dirks and Savas Parastatidis about the possibility of such an exploration project revealed enthusiasm on behalf of the Microsoft, especially

since Microsoft is already involved in Planets. Follow-up conversations will soon be initiated aimed at determining the possibility for a Microsoft-funded project regarding Planets and OAI-ORE synergy.

A key catalyst for both the two-year eChemistry experiment and the possible Planets/OAI-ORE exploration is the availability of draft standards form OAI-ORE at the time of project initiation. This proposal for an *OAI-ORE Acceleration Project* addresses that issue. The goal of this project is to produce by September 30, 2007 a draft specification of a model for the representation of compound information objects and mechanisms for serialization of this model for exchange among cooperating systems. This model will be based on the concept of Named Graphs[1], an idea developed in the W3C semantic web community. This model is the result of current work of the ORE-TC.

In order to address this goal, we request funding from Microsoft for a face-to-face meeting in early August 2007 and subsequent follow-up work through the remainder of August and September 2007 of a specially-selected group of ORE-TC and LG members with additional representation from the semantic web community. This group, heretofore called the OAI-ORE Acceleration Group or ORE-AG, will work intensively during this August/September project period to produce an alpha specification. This specification will be vetted by the ORE-TC to ensure that the normal process of the OAI-ORE will not be disrupted by this acceleration project. We expect, therefore, that the results of this acceleration project will transfer seamlessly to the main OAI-ORE process. The alpha specification will provide initial input to the proposed eChemistry experiment, and the possible Planets/OAI-ORE exploration. These projects ccaan subsequently leverage any refinements of the OAI-ORE standards developed in later OAI-ORE work.

Membership of the OAI-ORE Acceleration Group (ORE-LG)

- Julie Allinson (http://www.ukoln.ac.uk/ukoln/staff/j.allinson/) – Repositories Research Officer, UKOLN. Allinson leads JISC repositories programs and is a member of the OAI-ORE LG.
- Dr. Chris Bizer (http://sites.wiwiss.fu-berlin.de/suhl/ueber_uns/team/chris_bizer.htm) – Graduate Research Associate, Web-Based Systems Group, Free University of Berlin. Bizer co-developed the Named Graph concept is closely tied with W3C semantic web activities
- Carl Lagoze (http://www.cs.cornell.edu/lagoze) – Senior Research Associate, Computing and Information Science Cornell University. Lagoze co-directs OAI-ORE and is Principal Investigator of this proposed project.
- Dr. Michael Nelson (http://www.cs.odu.edu/~mln/) – Assistant Professor, Computer Science, Old Dominion University. Nelson co-authored the OAI-PMH specifications and is an active participant in OAI activities including the ORE-TC.
- Dr. Robert Sanderson (http://www.csc.liv.ac.uk/~azaroth/) – Professor, Computer Science, University of Liverpool. Sanderson is active in the XML and Data Mining Community and is a member of the ORE TC.

---

[1] Jeremy Carroll, Christian Bizer, Pat Hayes, Patrick Stickler. Named Graphs. 2005. (http://www.websemanticsjournal.org/ps/pub/2005-23)

- Dr. Herbert Van de Sompel (http://public.lanl.gov/herbertv/) – Digital Library Researcher, Research Library, Los Alamos National Laboratory. Van de Sompel co-directs OAI-ORE and will co-direct this acceleration project.
- Dr. Simeon Warner (http://www.cs.cornell.edu/people/simeon/) - Research Associate, Computing and Information Science Cornell University. Warner co-authored the OAI-PMH specifications and is an active participant in OAI activities including the ORE-TC. Warner also co-directs the arXiv ePrint archive with Paul Ginsparg.

We also expect that Savas Parastatidis from the Microsoft Technical Computing Group will join us depending on his availability.

Plan of Work

- August 6,7,8 – Face-to-face meeting, Cornell University, Ithaca, NY
- August 9 – September 7 – Specification writing and refinement (team internal)
- September 8 – September 24 – Vetting and refinement by ORE-TC
- September 30 – Alpha specification finalization and delivery