Proposal for a tutorial at Open Repositories 2013

The ResourceSync Framework for Resource Synchronization

Simeon Warner (Cornell University, simeon.warner@cornell.edu)

Abstract

This tutorial will give an overview and a practical introduction to ResourceSync [1], a synchronization framework consisting of modular capabilities that a server can selectively implement to enable third party systems to remain synchronized with its evolving resources. Motivations for the ResourceSync approach will be explained by outlining several synchronization use cases including re-use of information in scholarly repositories, and sharing of linked data knowledge bases. The tutorial will detail the underlying concepts and implementation of ResourceSync capabilities, discovery mechanisms, and serialization based on the widely adopted Sitemap protocol. It will describe extensibility of the synchronization framework beyond the core use cases, for example, to include communication of references to mirror locations of synchronization resources, transferring only patches for changed resources, and offering historical data.

The ResourceSync standards effort is a project of NISO [2] and the Open Archives Initiative [3]. ResourceSync is funded by the Alfred P. Sloan Foundation. UK participation is supported by Jisc.

[1] http://www.openarchives.org/rs
[2] http://www.niso.org/workrooms/resourcesync/
[3] http://www.openarchives.org/

Learning Outcomes

Tutorial attendees will realize the following learning outcomes:
- Understanding of the motivations for creating ResourceSync as a synchronization framework and the roles of Sources and Destinations in the synchronization process
- Familiarity with the underlying concepts of the ResourceSync capabilities
- Understanding of the Sitemap-based XML documents in the framework
- Understanding of integration strategies with existing systems
- Understanding of discovery mechanisms and issues

In addition, the hands-on session will give:
- Familiarity with the ResourceSync python library and client
- Experience using ResourceSync to synchronize, update and check a copy of an example set of web resources

Topical Outline

- Motivating use cases involving resource synchronization
- Concepts behind ResourceSync capabilities
- Recap of the Sitemap protocol as specified and used by major search engines
- Implementing ResourceSync capabilities
- Discovering ResourceSync capabilities
- Advanced examples within ResourceSync

Length of session

I propose a 2h tutorial with an optional 1h hands-on session using the ResourceSync python library and command-line client for attendees that have their own laptop (with python 2.6 or 2.7) and internet access.

Attendees

The intended audience includes technologists, programmers, and managers with some technical knowledge. The tutorial will assume a basic level of familiarity with notions of digital data repositories, Web resources, and XML. No prior experience with synchronization protocols or the Sitemap document format will be assumed.

I suggest a minimum of 10 attendees to run the tutorial. The maximum would depend mainly on room size.

Technology and facility requirements

- Computer projector
- Internet access would be useful for the main tutorial
- Internet access for the presenter and attendees would be essential for the hands-on session

Presenter

Simeon Warner is Director of the Repositories Group at Cornell University Library. Current projects include development of an archival repository, the arXiv e-print archive (http://arxiv.org/), and Project Euclid (http://projecteuclid.org/). He was one of the developers of arXiv and his research interests include web information systems, interoperability, plagiarism detection, and open-access scholarly publishing. He has been actively involved with the Open Archives Initiative (OAI) since its inception and was one of the authors of the OAI-PMH, OAI-ORE and ResourceSync specifications.