

ARTICLE
EXCERPTED
FROM:

ISQ INFORMATION STANDARDS QUARTERLY

WINTER 2012 | VOL 24 | ISSUE 1 | ISSN 1041-0031

SPECIAL EDITION: YEAR IN REVIEW AND STATE OF THE STANDARDS



Sloan Grant Supports Development of Resource Synchronization Standard

The Alfred P. Sloan Foundation has awarded a grant to NISO and the Open Archives Initiative to develop a new open standard on the real-time synchronization of Web resources.



PROBLEM

Increasingly, large-scale digital collections are available from multiple hosting locations and are cached at multiple servers. Examples of such collections include the Internet Archive's WayBack Machine, Twitter's collection of tweets, and Linked Data collections such as Freebase or DBpedia. In addition, high profile portals rely on resources originating in many distributed repositories. Examples include the Europeana portal, CiteSeer, HathiTrust, and OAIster. This proliferation of replicated copies of works or data on the Internet has created an increasingly challenging problem of keeping the repositories' holdings and the services that leverage them up-to-date and accurate. As we move from a Web of documents to a Web of data, synchronization becomes even more important: decisions made based on unsynchronized or incoherent scientific or economic data can have serious deleterious impact.

The OAI Protocol for Metadata Harvesting (PMH) 2.0 specification can be used to effectively synchronize the metadata about the resources, but synchronizing the resources themselves was never specified. Although some resource synchronization methods exist, they are generally ad hoc, arranged by the individuals involved, and cannot be universally deployed.

This proliferation of replicated copies of works or data on the Internet has created an increasingly challenging problem of keeping the repositories' holdings and the services that leverage them up-to-date and accurate.



PROPOSED SOLUTION

A standard will be developed for an interoperable, efficient, and lightweight mechanism to support synchronizing Web resources at scale. The standard will save time, effort, and resources by repository managers by automating the replication and updating process. It will increase the general availability of content from these repositories and will alleviate the variety of problems created by outdated, inaccurate, superseded content that exists on the Internet.

USE CASES

- Synchronization of Linked Data content
- Recurrently collecting Memento metadata from IIPC web archives to central aggregator
- arXiv mirroring

CORE TEAM

Cornell University & OAI:

Berhard Haslhofer, Carl Lagoze, Simeon Warner

Old Dominion University & OAI:

Michael L. Nelson

Los Alamos National Laboratory & OAI:

Martin Klein, Robert Sanderson, Herbert Van de Sompel

NISO:

Todd Carpenter, Nettie Lagace, Peter Murray

LINKS

Resource Sync Workroom

www.niso.org/workrooms/resourcesync/

I NR | doi: 10.3789/isqv24n1.2012.05