

Open Archives Initiative Protocol for Metadata Harvesting

From Wikipedia, the free encyclopedia
(Redirected from OAI-PMH)

OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) is a protocol developed by the Open Archives Initiative. It is used to harvest (or collect) the metadata descriptions of the records in an archive so that services can be built using metadata from many archives.

The protocol is usually just referred to as the OAI Protocol.

OAI-PMH uses XML over HTTP. The current version is 2.0, updated in 2002.

Contents

- 1 History
- 2 OAI registries
- 3 Uses
- 4 Software
- 5 Archives
- 6 Workshops
- 7 References
- 8 External links

History

This summary was largely drawn from (Lynch, 2001).

In the late 1990s, Herbert Van de Sompel (Ghent University) was working with researchers and librarians at Los Alamos National Laboratory (US) and called a meeting to address difficulties related to interoperability issues of e-print servers and digital repositories. The meeting was held in Santa Fe, New Mexico, in October 1999. A key development from the meeting was the definition of an interface that permitted e-print servers to expose metadata for the papers it held in a structured fashion so other repositories could identify and copy papers of interested from each other. This interface/protocol was named the "Santa Fe Convention".

Several workshops were held in 2000 at the ACM Digital Libraries conference (<http://www.jcdl.org/>) and elsewhere to share the ideas from the Santa Fe Convention. It was discovered at the workshops that the problems faced by the e-print community were also shared by libraries, museums, journal publishers, and others who needed to share distributed resources. To address these needs, the Coalition for Networked Information (<http://www.cni.org/>) and the Digital Library Federation (<http://www.diglib.org/>) provided funding to establish an Open Archives Initiative (OAI) secretariat managed by Herbert Van de Sompel and Carl Logoze. The OAI held a meeting at Cornell University (Ithaca, New York) in September 2000 to improve the interface developed at the Santa Fe Convention. The specifications were refined over e-mail.

OAI-PMH version 1.0 was introduced to the public in January 2001 at a workshop in Washington D.C., and another in February in Berlin, Germany. Subsequent modifications to the XML standard by the W3C required making minor modifications to OAI-PMH resulting in version 1.1. The current version, 2.0,

was released in June 2002. It contained several technical changes and enhancements and is not backward compatible.

OAI registries

The OAI Protocol has become widely adopted by many digital libraries, institutional repositories, and digital archives. Although registration is not mandatory, it is encouraged.

There are several large registries of OAI-compliant repositories:

1. The Open Archives list of registered OAI repositories (<http://www.openarchives.org/Register/BrowseSites>)
2. The OAI registry at University of Illinois at Urbana-Champaign (<http://gita.grainger.uiuc.edu/registry/Info.asp>)
3. The Celestial OAI registry (<http://celestial.eprints.org/>)
4. Eprint's Institutional Archives Registry (<http://archives.eprints.org/>)
5. Openarchives.eu The European Guide to OAI-PMH compliant repositories in the world (<http://www.openarchives.eu>)

Uses

Commercial search engines have started using OAI-PMH to acquire more resources. Google has started to accept OAI-PMH as part of their Sitemap Protocol, and they are using OAI-PMH to harvest information from the National Library of Australia Digital Object Repository (<http://www.nla.gov.au/digicoll/oai/>) . In 2004, Yahoo! acquired content from OAIster (University of Michigan) that was obtained through metadata harvesting with OAI-PMH.

The mod_oai project is using OAI-PMH to expose content to web crawlers that is accessible from Apache Web servers.

Software

OAI-PMH is based on a client-server architecture, in which "harvesters" request information on updated records from "repositories". Requests for data can be based on a datestamp range, and can be restricted to named sets defined by the provider. Data providers are required to provide XML metadata in Dublin Core format, and may also provide it in other XML formats.

A number of software systems support the OAI-PMH, including GNU EPrints from the University of Southampton and DSpace from MIT.

Archives

A number of large archives support the protocol including arXiv and the CERN Document Server.

Workshops

Since 2001 there has been a yearly workshop at CERN in Geneva.

References

- Carl Lagoze and Herbert Van de Sompel (2001). "The Open Archives Initiative: Building a Low-Barrier Interoperability Framework (<http://www.openarchives.org/documents/jcdl2001-oai.pdf>) ". *Proceedings of the first ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'01)*, 54-62.

- Lynch, Clifford A. (2001). "Metadata harvesting and the open archives initiative" (<http://www.arl.org/newsltr/217/mhp.html>) . ARL Bimonthly Report 217.
- Frank McCown, Xiaoming Liu, Michael L. Nelson and Mohammed Zubair (March/April 2006). "Search Engine Coverage of the OAI-PMH Corpus" (<http://library.lanl.gov/cgi-bin/getfile?LA-UR-05-9158.pdf>) ". *IEEE Internet Computing* **10** (2): 66-73.
- Herbert Van de Sompel and Carl Lagoze (2000). "The Santa Fe Convention of the Open Archives Initiative" (<http://dx.doi.org/10.1045/february2000-vandesompel-oai>) ". *D-Lib Magazine* **6** (2). DOI:10.1045/february2000-vandesompel-oai (<http://dx.doi.org/10.1045/february2000-vandesompel-oai>) .
- Herbert Van de Sompel, Jeffrey A. Young, and Thomas B. Hickey (2003). "Using the OAI-PMH ... Differently" (<http://dx.doi.org/10.1045/july2003-young>) ". *D-Lib Magazine* **9** (7/8). DOI:10.1045/10.1045/july2003-young (<http://dx.doi.org/10.1045/10.1045/july2003-young>) .

External links

- Protocol specification (<http://www.openarchives.org/OAI/openarchivesprotocol.html>)

Retrieved from

"http://en.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting"

Categories: Digital libraries | Internet protocols | Metadata | Open access

- This page was last modified 03:56, 5 January 2007.
- All text is available under the terms of the GNU Free Documentation License. (See **Copyrights** for details.)
Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a US-registered 501(c)(3) tax-deductible nonprofit charity.