

A Tour of Version 2.0 of the Open Archives Initiative Protocol for Metadata Harvesting

A proposal for a full day tutorial at the
2002 ACM/IEEE Joint Conference on Digital Libraries
<http://www.jcdl2002.org/>

Submitted by:
Michael L. Nelson
NASA Langley Research Center
Hampton VA 23681
+1 757 864 8511
m.l.nelson@larc.nasa.gov
<http://mln.larc.nasa.gov/~mln/>

Description

This tutorial will present the history, motivation, current status and future directions of the Open Archives Initiative (OAI) and its Protocol for Metadata Harvesting (OAI-PMH). Version 1.0 was released to the public in early 2001 in meetings in Washington DC and Berlin. Except for a slight change (version 1.1) to reflect the W3C's change in XML Schema, the OAI-PMH has been frozen from that time. Now, version 2.0 of the OAI-PMH is scheduled to be released to the public on May 1, 2002. While version 2.0 will reflect a gentle evolution of the protocol, there are many subtle yet important distinctions from version 1.1.

The tutorial will assume no prior knowledge of the OAI or metadata harvesting. It will assume familiarity with basic concepts and technologies of digital libraries (DLs): http, XML, and metadata.

Objectives

Tutorial attendees will meet the following learning objectives:

- a review of the "metadata harvesting" and "distributed searching" models;
- an explanation of the "data provider" and "service provider" concepts in OAI;
- an in-depth explanation and discussion of the OAI-PMH version 2.0;
- a migration strategy for OAI-PMH version 1.1 repositories to version 2.0;
- an outline of possible implementation strategies for host DLs;
- a discussion of advanced topics (e.g. flow control, sets, harvesting strategies).

Intended Audience

The intended audience for this tutorial includes technical staff and managers who:

- currently have an OAI repository and wish to upgrade to version 2.0;
- wish to add OAI capability to an existing or planned DL;
- offer services with metadata harvested from OAI repositories;

Outline

9am – 9:30am	Nelson	Introduction and overview of OAI <ul style="list-style-type: none">– history, motivation, scope and definitions– current state of OAI participation
9:30am – 11:00am	Van de Sompel	Detailed walk through of the OAI-PMH
11:00am – noon	Lagoze	Migrating from 1.1 to 2.0
1pm – 2:30pm	Warner	Advanced topics: <ul style="list-style-type: none">– sets– flow control– harvesting strategies
2:30pm – 3:30pm	Suleman	Review of existing software and tools: <ul style="list-style-type: none">– repository explorer– data provider software– service provider software
3:30pm – 5:00pm		Implementer's panel a chance for the audience to interact with a variety of people at different institutions who have implemented either service providers or data providers; including: <ul style="list-style-type: none">– a representative from Old Dominion University– a representative from NSDL– a representative from Southampton– a representative from the Library of Congress– a representative from OCLC– a representative from UIUC– others TBD.
5:00pm – 5:30pm	Lagoze	Future directions and closing

Materials

This tutorial will draw from the OAI-PMH document available from the OAI home page

(<http://www.openarchives.org/>) as well as slides prepared by the individual presenters. All materials will be available for download before the conference dates.

Presenters

Carl Lagoze <lagoze@cs.cornell.edu>

<http://www.cs.cornell.edu/lagoze/>

Carl Lagoze is Digital Library Scientist in Computing and Information Science at Cornell University. At Cornell he teaches and co-directs the Digital Library Research Group, which undertakes research in digital library, electronic publishing, and distributed information systems. His main research interests include metadata, digital library architecture, and policy expression and enforcement. Carl is a member of the OAI executive, steering and technical committees.

Michael L. Nelson <m.l.nelson@larc.nasa.gov>

<http://mln.larc.nasa.gov/~mln/>

Michael L. Nelson received his B.S. in computer science from Virginia Tech in 1991, and his M.S. and Ph.D. in computer science from Old Dominion University in 1997 and 2000. He has worked at NASA Langley Research Center since 1991, originally in distributed and parallel computing and then shifting to WWW and digital libraries in 1993. He is an adjunct assistant professor at Old Dominion University, and was a visiting assistant professor at the School of Information and Library Science at the University of North Carolina at Chapel Hill for the 2000–2001 academic year. Michael is a member of the OAI technical committee.

Hussein Suleman <hussein@vt.edu>

<http://purl.org/net/hussein/>

Hussein Suleman is a Ph.D. student working with Edward Fox at Virginia Tech. His research focus is on topics closely related to matters of interoperability. He is currently funded by NSF to work with the iLumina project (<http://www.ilumina-project.org>) which is building a federated digital library of resources for science and technology education, incorporating resources from the Computer Science Teaching Center (<http://www.cstc.org>). This and other projects utilize technology developed by the Open Archives Initiative. He has implemented the standards on various platforms and also developed and actively maintains the Repository Explorer software that is used by the Open Archives community to rigorously test archives for compliance with the standards. Hussein is a member of the OAI technical committee.

Simeon Warner <simeon@cs.cornell.edu>

Simeon Warner is one of the maintainers and developers of the arXiv e-print archive (<http://arXiv.org/>). He recently moved to the Computer Science Department at Cornell University. Before that he worked at Los Alamos National Laboratory and the Physics Department at Syracuse University. He has implemented and maintains an OAI interface for arXiv, along with test-suite of harvesting software. Simeon is a member of the OAI

technical committee.

Herbert Van de Sompel <hvdsomp@yahoo.com>

Herbert Van de Sompel graduated in mathematics and computer science at Ghent University, and in 2000, obtained a Ph.D. from Ghent University for his research on dynamic and context-sensitive reference linking, now commonly known as the OpenURL framework. From 1982 to 1998 he worked as Head of Library Automation at Ghent University. While at the Los Alamos National Laboratory in 1999, Herbert started the Open Archives Initiative with Paul Ginsparg and Rick Luce. With Carl Lagoze, Herbert forms the executive committee of the OAI; he also is on the technical and the steering committees. Herbert was a Visiting Professor in Computer Science at Cornell University in 2000 – 2001. Afterwards, he was the Director of e-Strategy and Programmes at the British Library. Currently, he is a digital library researcher at the Los Alamos National Laboratory.