Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting

Protocol Version 2.0 of 2002-06-14
Document Version 2005/05/03T22:51:00Z
http://www.openarchives.org/OAI/2.0/guidelines.htm

Editors

Carl Lagoze (OAI Executive; Cornell University - Computer Science)
Herbert Van de Sompel (OAI Executive; Los Alamos National Laboratory - Research Library)
Michael Nelson (Old Dominion University - Computer Science)
Simeon Warner (Cornell University - Computer Science)

1. Introduction

These guidelines are a supplement to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) Version 2.0 and are intended to be read in conjunction with that specification.

It is anticipated that the guidelines will be added to or updated while the protocol specification remains unchanged. To facilitate this, the guidelines are separated out into a set of documents for which this document is a table of contents.

2. General Guidelines

2.1 Guidelines for Repository Implementers

These guidelines describe best practices and discuss implementation issues for repository implementers. Topics include: minimal repository implementation; datestamps and granularity; best practices for the use of `resumptionTokens`; collection and set descriptions; flow control and load balancing; response compression; and error handling.

2.2 Guidelines for Harvester Implementers

These guidelines discuss issues relating to implementing and operating harvesting software. Topics include: running harvesting software; datestamps and granularity; sets; flow control, load balancing and redirection; incomplete lists and `resumptionToken`; and response compression.

2.3 Guidelines for Aggregators, Caches and Proxies

These guidelines discuss issues specific to systems that act as both harvesters and repositories. Topics include: identifiers; provenance; and datestamps.

3. Guidelines for Optional Containers

There are a number of places in OAI-PMH responses where XML complying with any external schema may be supplied. These containers are provided for extensibility and for community specific enhancements. The

following sections list the optional containers and link to existing schemas.

3.1 Repository-level `<description>` container

The response to an `Identify` request may contain description containers that can be used to express properties of the repository that are not covered by the standard response to the `Identify` verb. The following guidelines are provided:

`oai-identifier`: a specification describing a specific, recommended implementation of unique identifiers which repositories may adhere to;
`rightsManifest`: part of the guidelines for conveying rights expressions than can be used to include a manifest of rights statements that apply to metadata from items in the repository;
`eprints`: a schema that can be used to provide collection-level metadata for eprint repositories;
`friends`: a recommended schema allowing a repository to list confederate repositories as a means to support automatic discovery of repositories by harvesters;
`branding`: a schema for repositories to provide branding information;
`gateway`: a schema than can be used to describe a gateway which is acting as an OAI-PMH repository.

3.2 Set-level `<setDescription>` container

The response to a `ListSets` request may contain `setDescription` containers, which provide an extensible mechanism for communities to describe their sets. Dublin Core metadata may be used for this purpose. A schema for unqualified Dublin Core is provided in the protocol document. The Guidelines for Repository Implementers makes further recommendations about schema that can be used for this purpose. Other schema that may be used at the set level include:

`rightsManifest`: part of the guidelines for conveying rights expressions than can be used to include a manifest of rights statements that apply to metadata from items in the set;
`branding`: a schema for repositories to provide branding information.

3.3 Record-level `<metadata>` container

Dublin Core metadata with `metadataPrefix oai_dc` is mandatory and is described in the protocol document. In addition, schema descriptions for the following metadata formats are provided:

`rfc1807`: a schema for rfc1807 format metadata;
`marc21`: a recommended schema for MARC21 metadata, provided by the Library of Congress;
`oai_marc`: a schema for MARC format metadata.

3.4 Record-level `<about>` container

A record may contain `<about>` containers that provide information about the `<metadata>` part of the record. It is expected that community specific standards may arise in addition to the following specifications:

`provenance`: a schema that is recommended for the description of the provenance of metadata that is re-exposed by a repository, i.e. metadata that has previously been harvested before being exposed by the repository;
`rights`: a specification for conveying rights expressions that pertain to the `<metadata>` part of the record.

4. Static Repositories and Static Repository Gateways

The [Specification for an OAI Static Repository and an OAI Static Repository Gateway](#) provides a simple approach for exposing relatively static and small collections of metadata records through the OAI-PMH. It is recommended that Static Repository Gateway implementers use the `gateway` container to include descriptions of the gateway.

5. Community Guidelines

Individual OAI user communities are encouraged to endorse particular formats and schemas, perhaps including some of those listed above, and to define additional standards that meet their own needs. The following community specific guidelines are available:

- **[Open Language Archives Community](#)** - a partnership of institutions and individuals who are creating a worldwide virtual library of language resources. Includes a specific metadata format and instructions for data providers.
- **[Repository description for the `eprints` community](#)** - a schema released with OAI-PMH v2.0 to describe e-print repositories.

Acknowledgements

Support for the development of the OAI-PMH and for other Open Archives Initiative activities comes from the [Digital Library Federation](#), the [Coalition for Networked Information](#), and from the National Science Foundation through Grant No. IIS-9817416. Individuals who have played a significant role in the development of OAI-PMH version 2.0 are [acknowledged in the protocol document](#).

Document History

2005-05-03: Added guidelines for conveying rights expressions about metadata.
2003-10-10: Added Static Repository Gateway and Gateway Description guidelines.
2002-05-08: Added `branding`, wording changes.
2002-03-31: Release of initial version of OAI-PMH v2.0 guidelines documents.