



Implementation of OAI protocol by Spanish NFP

Ángel J. Martín Sesma
CEDEX

14 de noviembre de 2006



CONTENT

- Basic OAI concepts

- Open Archives Initiative (OAI)
- OAI Protocol for Metadata Harvesting (OAI-PMH)

- Cross search or harvest?

- Data and Service providers

- Roles
- Flexibility of deployment



CONTENT

- Technical Ideas of OAI-PMH
 - Basic functioning
 - Overview and structure model
 - Protocol details
- Implementing OAI-PMH on a Data Provider
 - Components and architecture
 - Flow chart
 - Resumption tokens
 - Testing



CONTENT

- HISPAGUA OAI-PMH implementation
 - Specifications
 - Access URL's
 - Testing demo

- References

Basic OAI concepts

■ Open Archives Initiative (OAI)

- The essence of the OAI is to enable access to web-accessible material through interoperable repositories for metadata sharing, publishing and archiving.
- Proposes a solution to access across heterogeneous data repositories.
- Establishes a standard in metadata exchange.

Basic OAI concepts

■ OAI Protocol for Metadata Harvesting (OAI-PMH)

- Defines a lightweight mechanism for harvesting records containing metadata from repositories.
- It is based on HTTP and XML standards.
- Harvested metadata may be in any format agreed by a community and the response is recommended to be compliant to unqualified **Dublin Core** format to provide a minimum level of interoperability.

Cross search or harvest?

■ Performance

- First attempts of collecting data from several sources showed that crossed searches were degraded due to slow responses from searched repositories.
- Differences between query language differences and search attribute variations introduced barriers to access all collectable data.
- Serviced based on harvested metadata reduces the number of nodes to search in to only one, giving significant performance benefits. Also, a value added service could be given using only one query language, set of attributes and ranking algorithm.

Data and Service providers

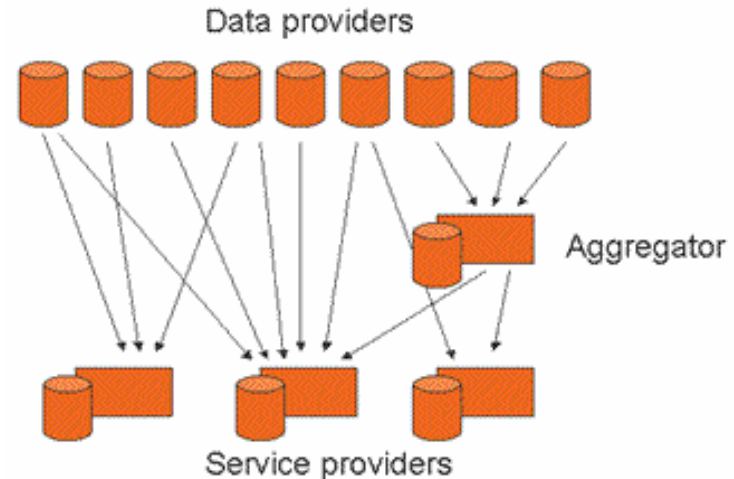
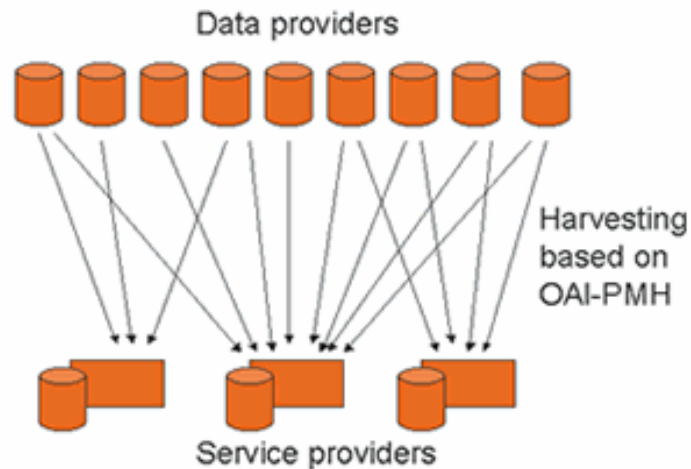
■ Two roles identified:

- **Data Providers (DP)** handle the deposit and publishing of resources in a repository and “expose” for harvesting the metadata about them. They are the creators and keepers of the metadata. Provides a machine-oriented interface to access data.
- **Service Providers (SP)** harvest metadata from DP's. They use the harvested metadata for providing services across all of them, as searching, sorting and review of records, etc. Provides human-oriented interfaces to access harvested data and value added services.

Data and Service providers

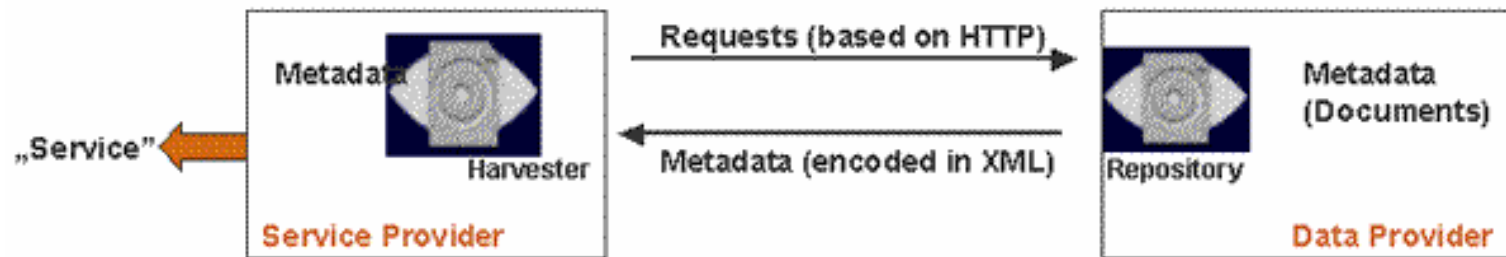
- Flexibility of deployment

- Systems can be deployed in a variety of configurations



Technical ideas of OAI-PMH

■ Basic functioning



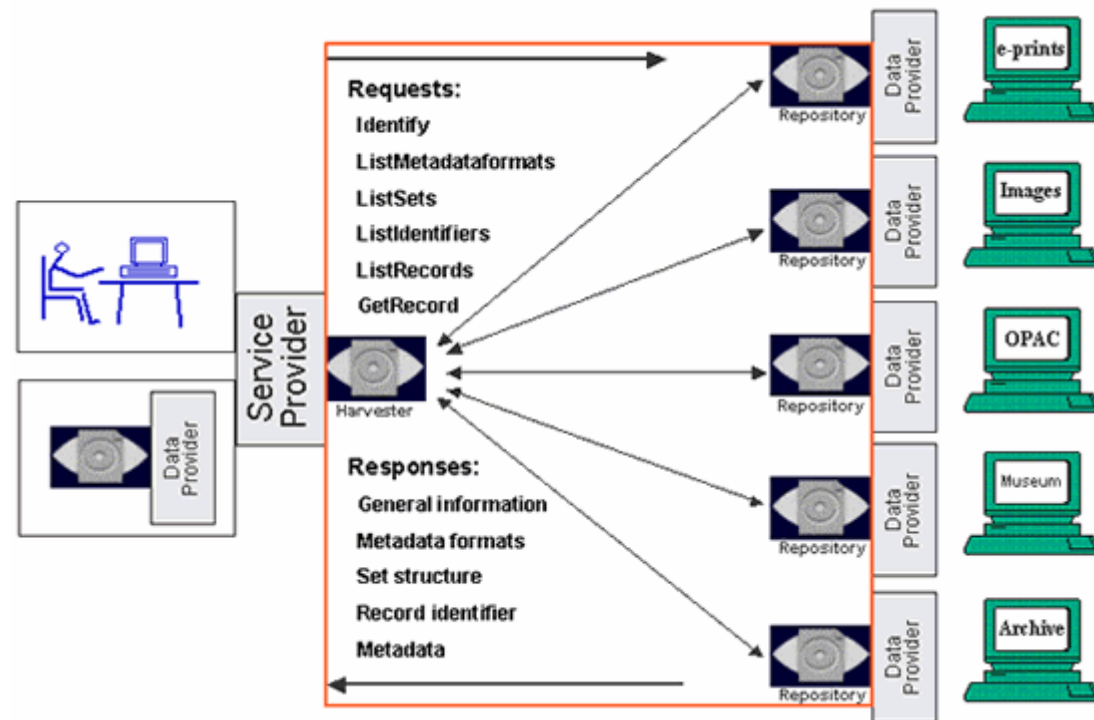
- **Data Providers** provide free access to metadata, and may, but do not necessarily, offer free access to full texts or other resources.
- **Service Providers** use the OAI interfaces of the Data Providers to harvest and store metadata. This means that there are no live search requests to the Data Providers. Offer value-added services on the basis of the metadata harvested, and they may enrich them in order to do so.

Technical ideas of OAI-PMH

■ Overview and structure model

The OAI-PMH protocol is based on HTTP. Request arguments are issued as **GET** or **POST** parameters. OAI-PMH supports six request types (known as "verbs"),

Responses are encoded in XML syntax. OAI-PMH supports any metadata format encoded in XML. Dublin Core is the minimal format specified for basic interoperability.



Technical ideas of OAI-PMH

■ Flow control

- Four of the verbs return a list of entries. Three of them may reply with 'large' lists.
- The response to a request includes:
Incomplete list, resumption token, expiration date, size of complete list
- Includes the next (which may be the last) section of the list and a resumption token. That resumption token is empty if the last section of the list is enclosed.

Technical ideas of OAI-PMH

■ Flow control



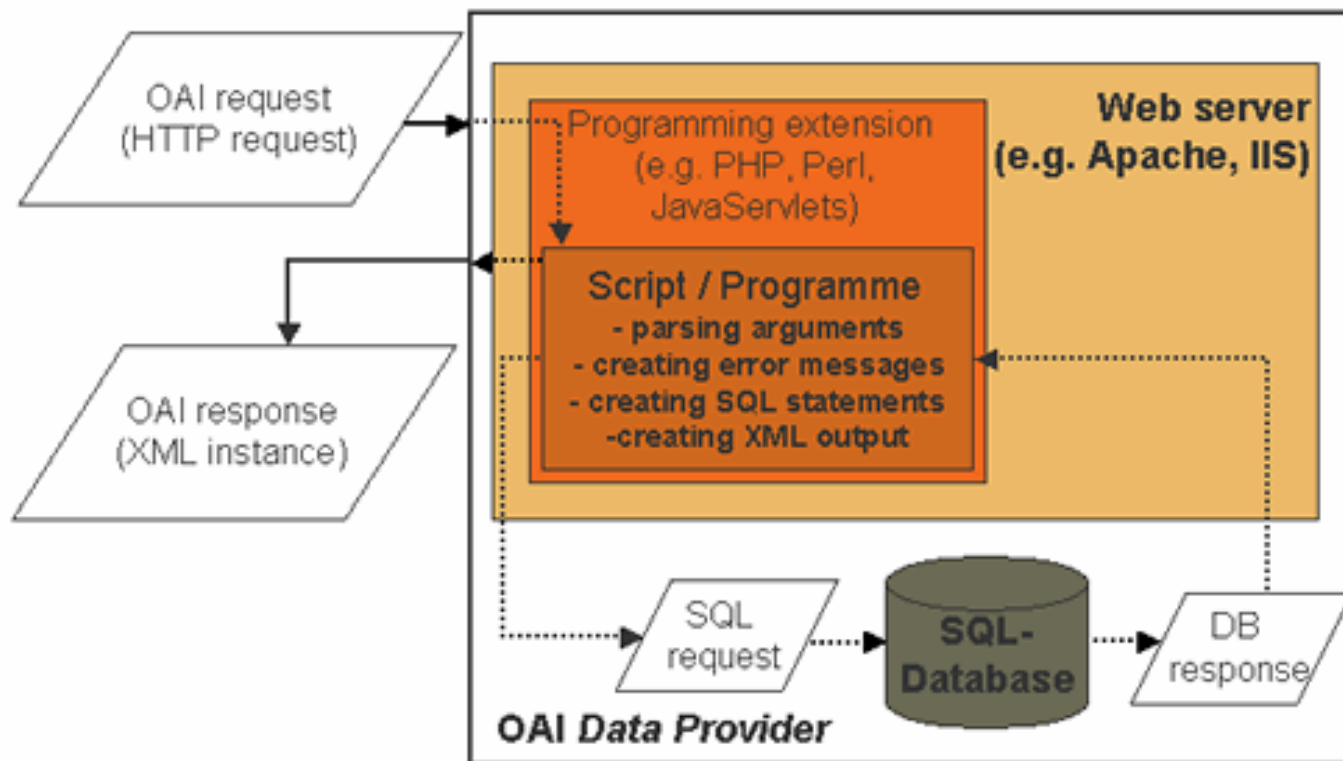
Implementing OAI-PHM on a Data Provider

■ Components and architecture

- **Argument Parser** validates OAI requests.
- **Error Generator** creates XML responses with encoded error messages.
- **Database Query / Local Metadata Extraction** retrieves metadata from the repository, according to the required metadata format.
- **XML Generator / Response Creation** creates XML responses with encoded metadata information.
- **Flow Control** realises incomplete list sequences for 'larger' repositories. It uses resumption token as the control mechanism.

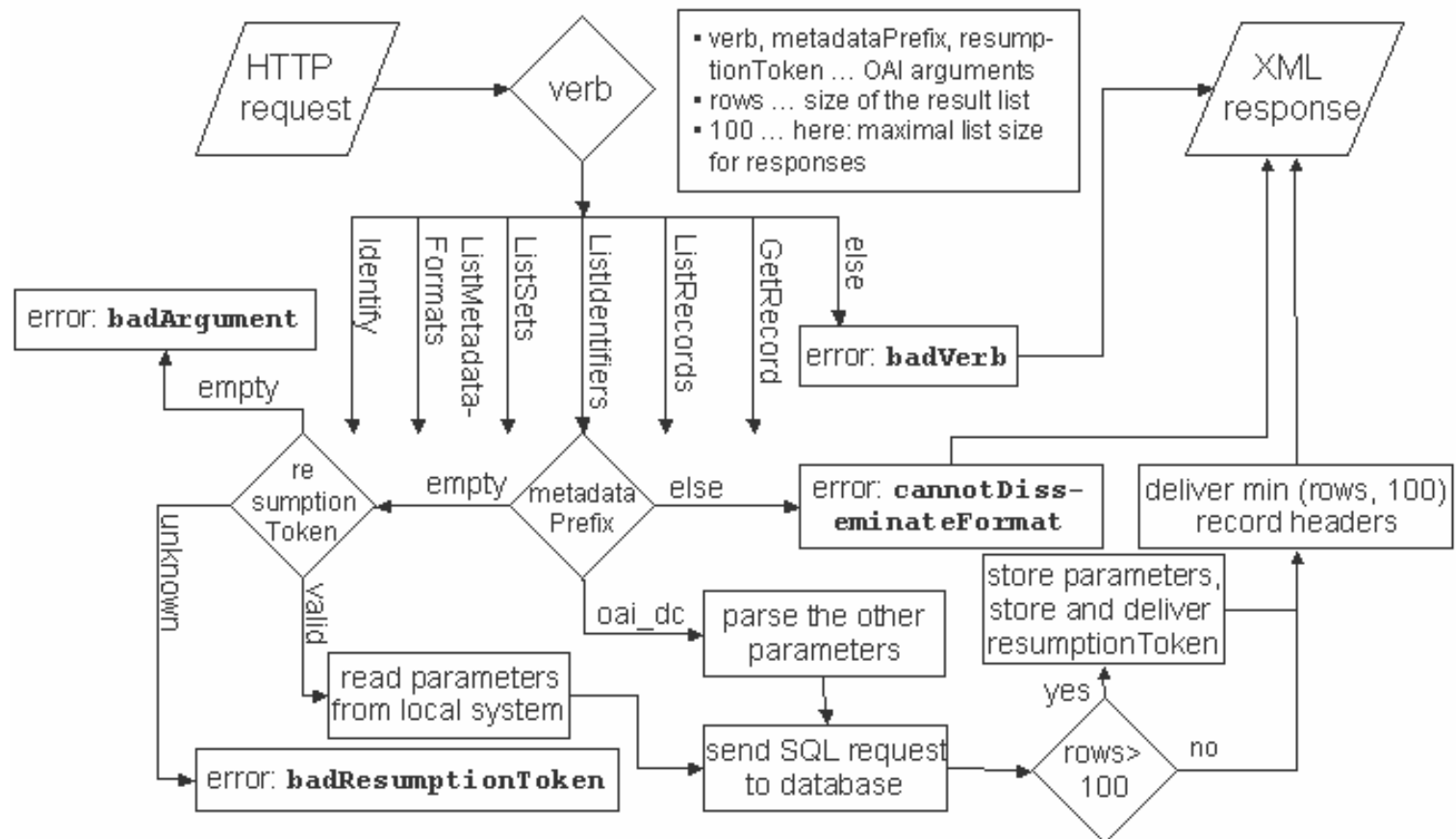
Implementing OAI-PHM on a Data Provider

■ Components and architecture



Implementing OAI-PHM on a Data Provider

■ Flow chart



Implementing OAI-PMH on a Data Provider

■ Resumption tokens (RT)

- RT should be implemented for handling "large" lists. It is initiated by Data Provider, and is used to store parameters (such as `set` or `from`) and the number of already delivered records.

DP needs to answer correctly if the most recent RT of a query is reissued. This feature allows SP to recover from network errors without having to reissue the complete list from the beginning.



Implementing OAI-PMH on a Data Provider

■ Testing

- It is possible to use the Repository Explorer to test the DP implementation (<http://re.cs.uct.ac.za/>) by browsing through the repository. Repository Explorer is an interactive, automatic compliance tester. It allows you to provide arguments via HTML forms. The responses are validated as conformant with OAI-PMH.
- It can check your repository against each of the OAI-PMH verbs in turn, setting parameters where required for date ranges, metadataPrefix, identifier, set, and resumption token. Thus, all aspects of the protocol can be tested, and the results of queries are checked for conformance with the expected syntax.

Implementing OAI-PHM on a Data Provider

■ Testing

The screenshot shows a web browser window titled "Open Archives Initiative - Repository Explorer". The address bar shows the URL: `http://oai.dlib.vt.edu/cgi-bin/Explorer/oa2.0/testoai`. The page has a green header with the title "Open Archives Initiative - Repository Explorer" and a logo on the left. Below the header, it states "explorer version - 1.43a ; protocol version - 1.0/1.1/2.0 ; March 2003".

The main content area is light green and contains the following text:

This site presents an interface to interactively test archives for compliance with the OAI Protocol for Metadata Harvesting [[Click here for details](#)]

JavaScript is required

Note: To avoid HTTP errors, please wait for each page to finish loading before clicking on any link.

Please enter the URL to the OAI interface (everything before the ?) or choose a predefined archive from the table
Then click on a verb from the list below to test that function (entering parameters as necessary)

URL :

A list of predefined archives is shown in a scrollable box:

- A Celebration of Women Writers
- Acervo de la Biblioteca -Dr. Jorge Villalobos Padilla, S.J.- (ITESO)
- AIM25 - Archives in London
- Alaska Native Language Center Archives

Below the list are two links: [[View Archive Website](#)] [[Test and Add an archive to this list](#)]

At the bottom, there is a table with two columns: "Verbs" and "Parameters".

Verbs	Parameters
Identify	from (eg., YYYY-MM-DD) : <input type="text"/> until (eg., YYYY-MM-DD) : <input type="text"/>

The browser's taskbar at the bottom shows the Start button, several icons, and the system tray with the date and time: 11:12.

HISPAGUA OAI-PHM implementation

■ Specifications

- Data Provider implemented to provide access to HISPAGUA databases metadata
- Written in PHP and supported by existing MySQL databases. Based on free PHP software “phpoi2”
<http://physnet.uni-oldenburg.de/oai/>
- Fulfils the requirements of OAI-PHM 2.0
- Results follow extended Dublin Core specifications
- 5 different DP: Congresses, Training Courses, Documents, Legislation and News

HISPAGUA OAI-PHM implementation

- Access URL's

News: http://hispagua.cedex.es/dp_noticias/oai2.php

Congresses: http://hispagua.cedex.es/dp_congresos/oai2.php

T. Courses: http://hispagua.cedex.es/dp_noticias/oai2.php

Documents: http://hispagua.cedex.es/dp_documentos/oai2.php

Legislation: http://hispagua.cedex.es/dp_legislacion/oai2.php

HISPAGUA OAI-PHM implementation

■ Testing demo

- It is possible to use the Repository Explorer to test the DP implementation (<http://re.cs.uct.ac.za/>) by browsing through the repository.

References

- Open Archives Initiative (OAI official Web site)
 - <http://www.openarchives.org/>
- Open Archives Forum (OA-Forum Web site)
 - <http://www.oaforum.org/>
- Dublin Core
 - <http://dublincore.org/>
- OAI-PMH implementation guidelines
 - <http://www.openarchives.org/OAI/2.0/guidelines.htm>
- PHP OAI Data Provider
 - <http://physnet.uni-oldenburg.de/oai/>