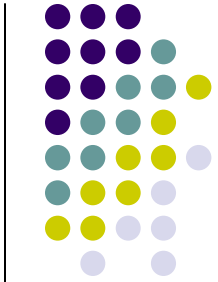
Centro de Información y Documentación Científica

CSIC

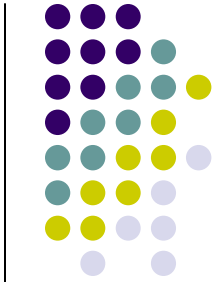


Séminaire d'information et de formation pour les responsables du contenu des PFN. CEDEX. Madrid, 22-25,-11-2004

Proposition d' architecture pour le partage de contenu

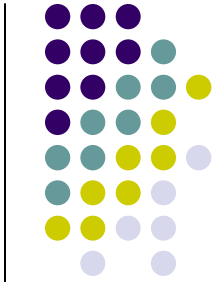
Elena Fernández
CINDOC-CSIC

Objetives



- Interoperability to facilitate the Internet disemination of contents.
- Automation in sharing data with minimum changes in the existing PFN.The process will be based on technologies transparent to the system.

Open Archives



- System that allows the storage of items (articles, preprints, reports..)
- The system collects metadata on these items and makes them avaliable through an interface.
- A possible interface is OAI-PMH

Some History: e-prints.

- **arXiv.org** (1991). System disigned by Paul Ginsparg to store and distribute pre-publications and reports in Physics, Mathematics. It was first at Los Álamos and after at Cornell University. Centralized architecture with worldwide mirrors

- **Networked Computer Science Technical Reference Library**. (1997). Repository containing technical reports on computer science. Distributed architecture, the documents are stored by the partnes institutions

- **RePEC: Research Papers in Economics** (1997). 100.000 documents in full text from 400 Economics Departaments. The architecture is for the first time implemented by service providers versus data providers

Some History: e-prints.

- CogPrints: A Southampton University Project. Similar to arXiv
- NDLTD: Networked Digital Library of Theses and Dissertations. Its main purpose is to build a PhD digital library

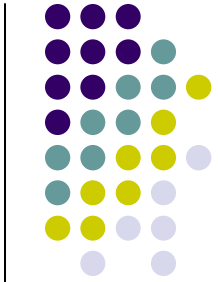
Santa Fe Convention (1999)

- Simple Protocol (HTTP+XML)
- Architecture: data provider- service provider
- Model based on the metadata harvesting
- Use of a common Metadata format (Dublin Core)
- Posibility to use other formats depending on the subject area

OAI – PMH

Open Archives Initiative- Protocol for Metadata Harvesting

Results

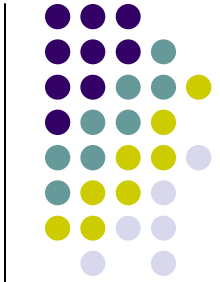


Open Archives Initiative- Protocol for Metadata Harvesting (OAI-PMH)

Currently:

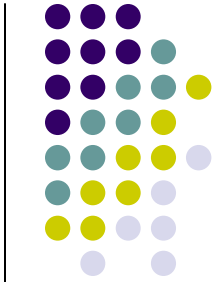
- Adopted by over 100 institutions
- Software development
- Several research projects funded by public bodies (USA y EU)

Two roles

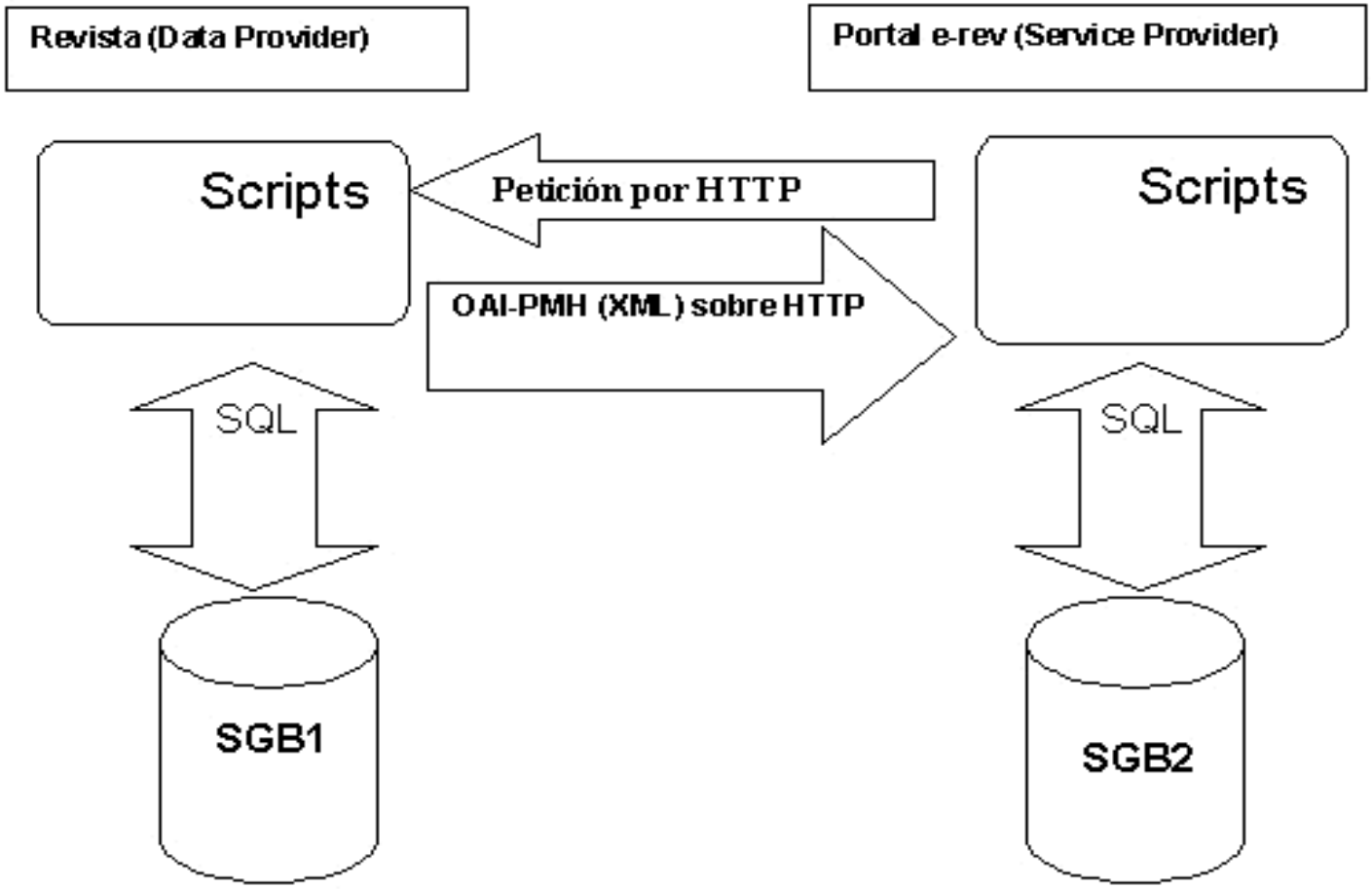


- Data Providers: They host a repository containing resources to be published and they include the resources metadata so that the service provider may recover them
- Service Providers : They recover the metadata produced by the data providers and use them to facilitate the use of the data

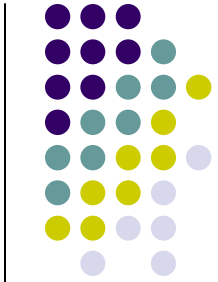
OAI- PMH Architecture



- “Harvesting” approach at metadata level
- Global division in data provider and service provider
- OAI-PMH uses the HTTP protocol
- The requests are on HTTP, GET o POST proccesses
- The repplies are validated XML documents

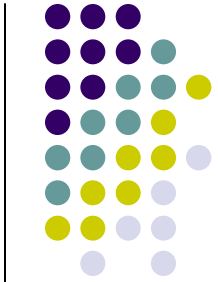

Revista (Data Provider)
Portal e-rev (Service Provider)
Scripts
Petición por HTTP
Scripts
OAI-PMH (XML) sobre HTTP
SQL
SQL
SGB1
SGB2

DATA provider requirements



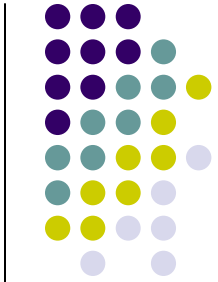
- Metadata storage support
- web support: web server
- API - Application Programming Interface – manteinance
- URL
- Metadata Format . Dublin Core
- Metadata initial date/modification date
- Logic Support for hierarchical sets
- Flow control: allows data provider and service provider dialog

SERVICE Provider requirements



- Metadata storage support : Access to Database through SQL requests
- Web server: Apache, IIS, …
- API - Application Programming Interface – manteinance

Software del SERVICE provider



Software ARC harvester and search engine: ARC

url: qrc.cs.odu.edu

Developed by: Digital Library Research Group, Old Dominium University

Development in Java

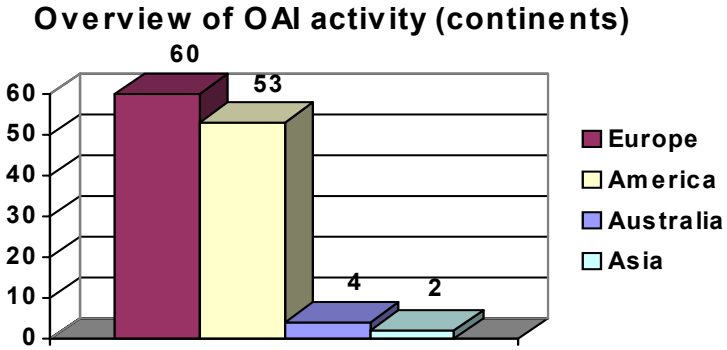
*Applications Server:*Tomcat

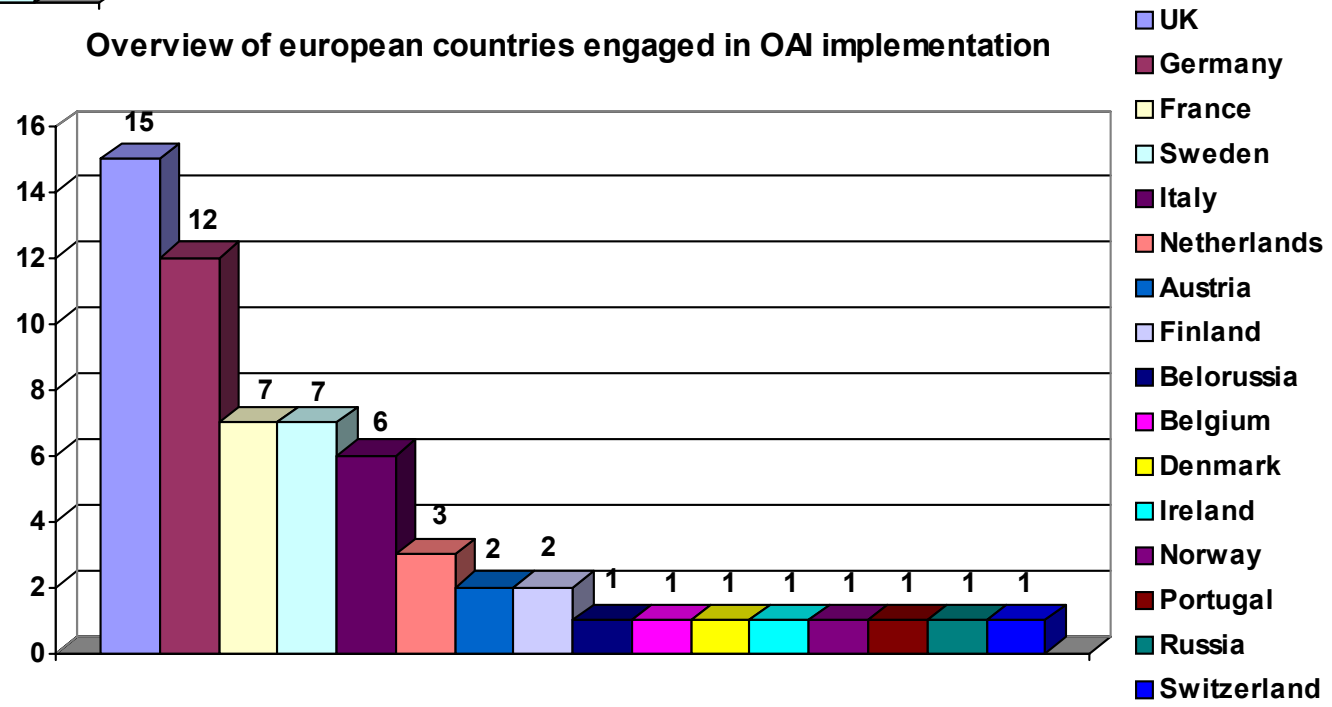
Database managing system: Oracle and MySqlhave been tested

Operative system : Windows, Linux, Solaris, etc…

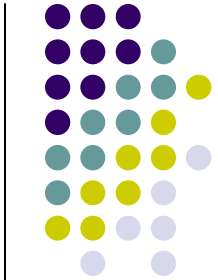
OAI Activities in Europa

(January, 2003)





References



- http://www.oaforum.org
- Community of interests in OAI
- Workshops
- Tutorial (very good)
- Database:
- Data Providers
  - Projects
  - Software