



## Bibliothèque numérique de l'enssib

Extending the network: libraries and their partners, 17 au 20 juin 2003  
32<sup>e</sup> congrès LIBER

---

### **Open archive solutions to traditional archive/library cooperation**

CASTELLI, Donatella  
ISTI-CNR

CASTELLI, Donatella. Open archive solutions to traditional archive/library cooperation. In *32nd LIBER Annual General Conference, Extending the network: libraries and their partners, Rome, du 17 au 20 juin 2003* [en ligne]. Format PDF.

Disponible sur : <<http://www.enssib.fr/bibliotheque-numerique/notice-1191>>

Ce document est « **tous droits réservés** ». Il est protégé par le droit d'auteur et le code de la propriété intellectuelle. Il est strictement interdit de le reproduire, dans sa forme ou son contenu, totalement ou partiellement, sans un accord écrit de son auteur.

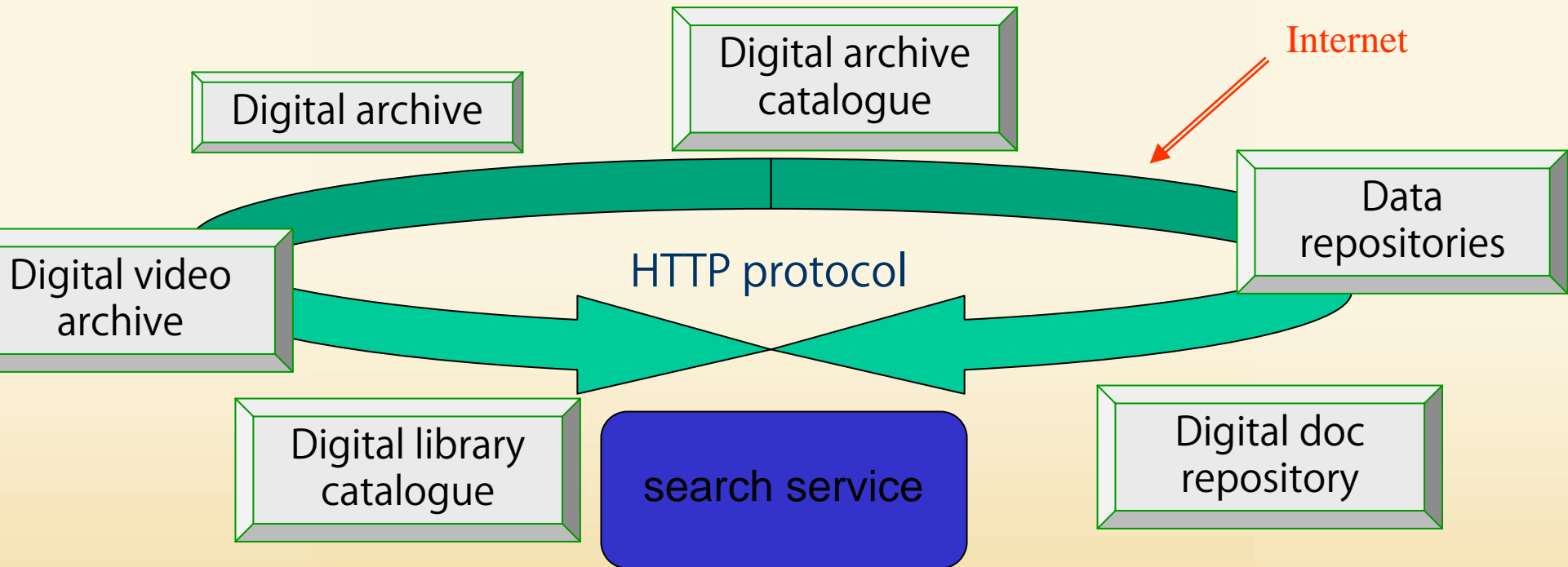
L'ensemble des documents mis en ligne par l'enssib sont accessibles à partir du site : <http://www.enssib.fr/bibliotheque-numerique/>

# Open archive solutions to traditional archive/library cooperation

Donatella Castelli  
ISTI-CNR  
Pisa (Italy)  
castelli@isti.cnr.it



# Internet – digital objects



# Cross-search and access services

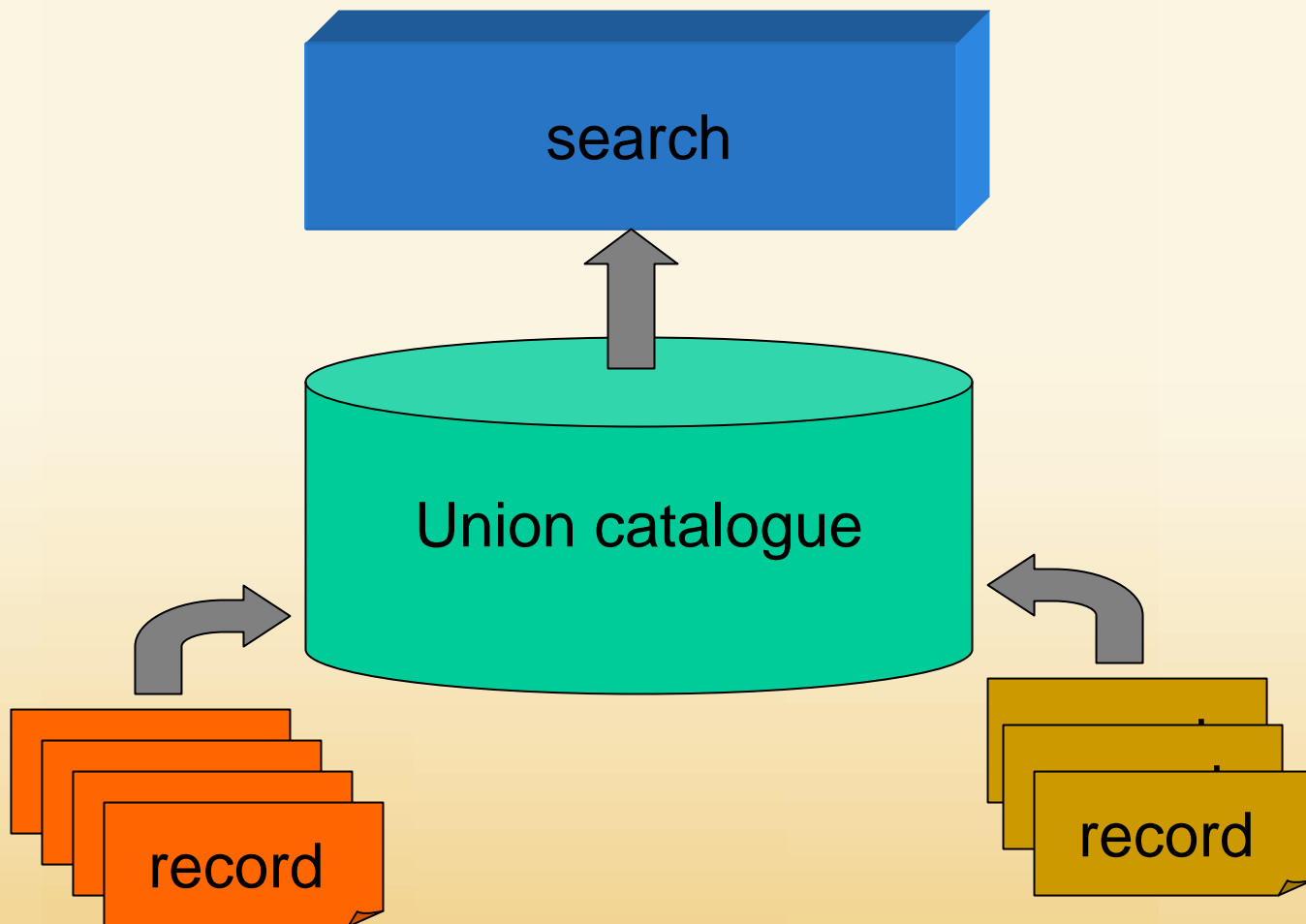
Require the solution of different issues

- technological
- organizational
- economical

# Focus of this talk

Technological issues

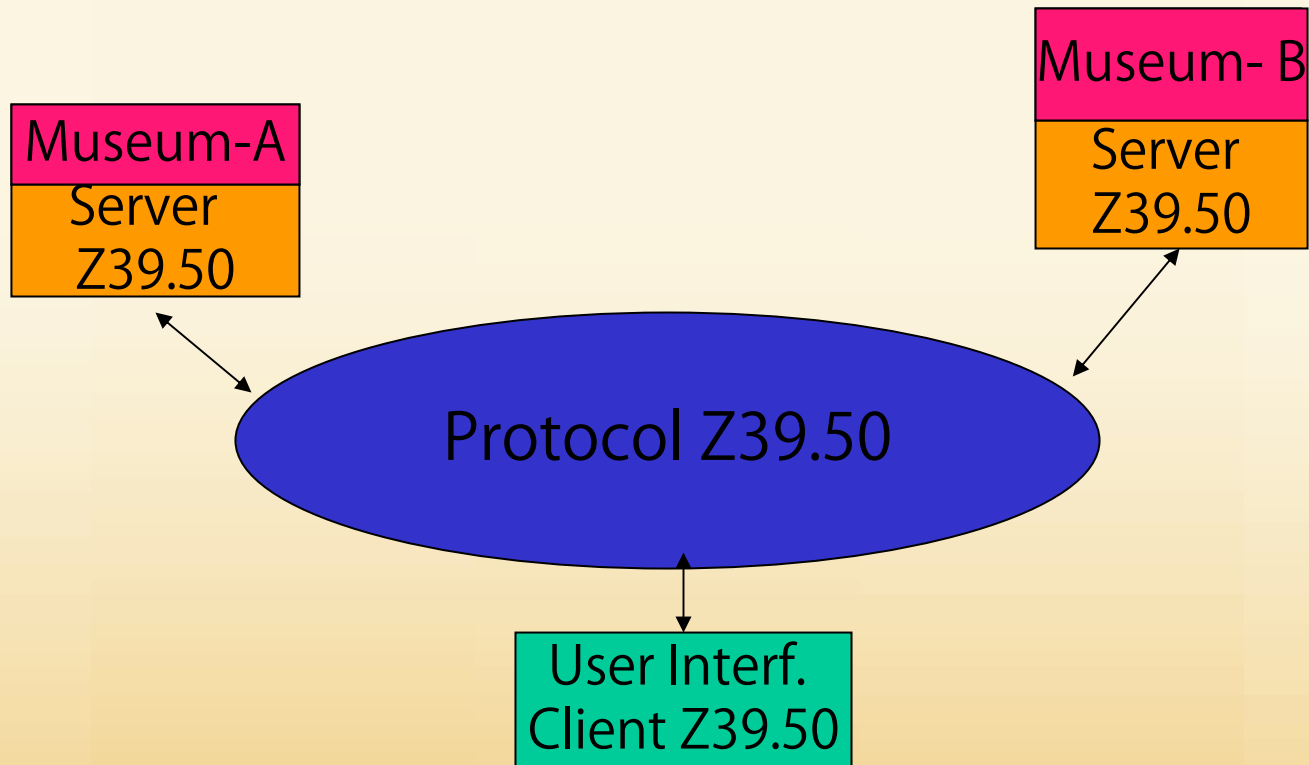
# Most used technical solutions...



# Main problems

- Scalability
- Maintainability
- One single metadata format

# Most used technical solutions...

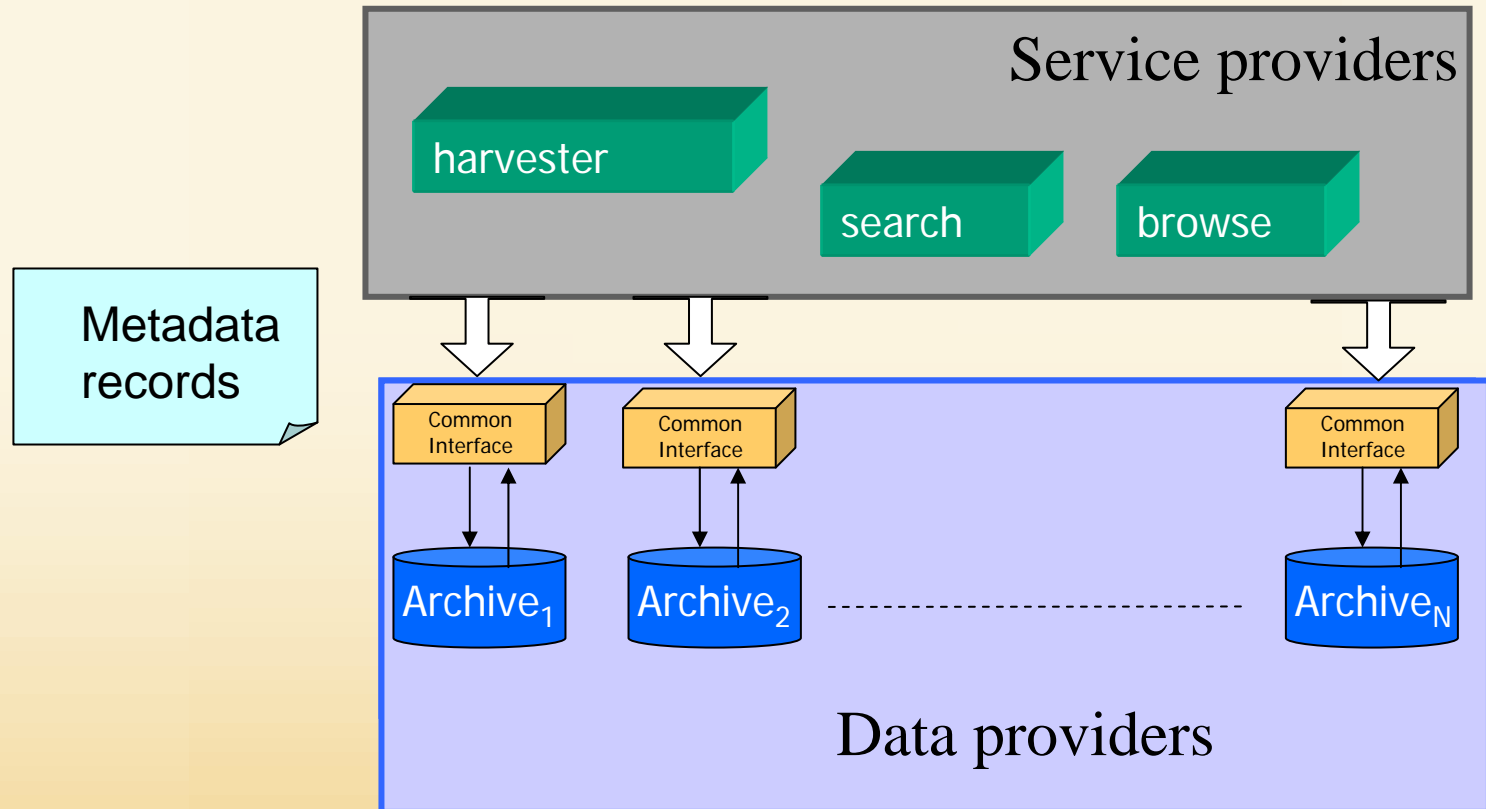




# Main problems

- the data sources must support complex Z39.50 server software
- considerable coordination is required to set up workable profiles
- Z39.50 search does not scale

# Open Archives Initiative Protocol for Metadata Harvesting



# Note: OAI terminology

- **Archive**  
repository of digital information
- **Open archive**  
repository that provides a machine interface for making its content available to external services

# History

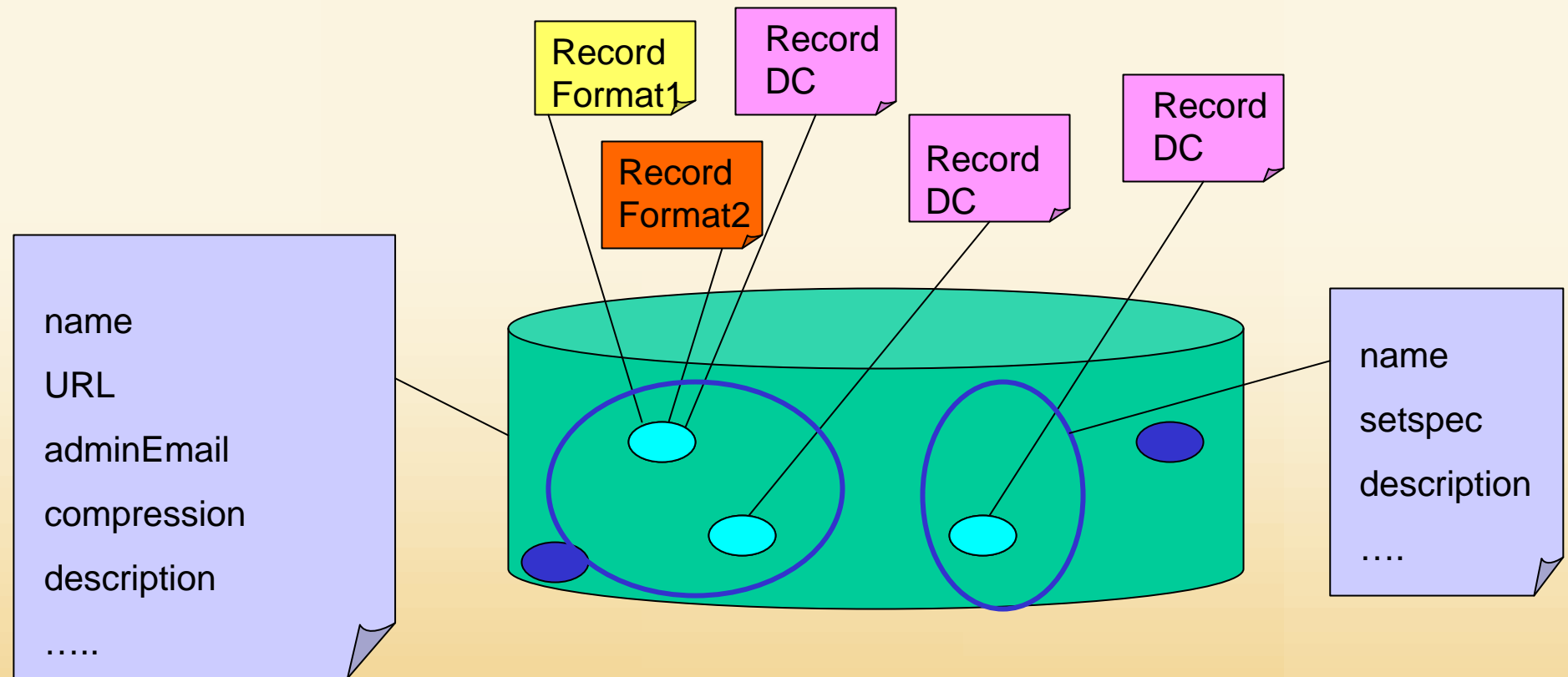
Santa Fe (New Mexico), July 1999 - Scientific communities that publish their pre-prints on electronic archives met together

- **ArXiv** (Los Alamos National Laboratory Physics Archive -100.000 papers, 50.000 user daily)
- **CogPrints** (University of Southampton - papers on Psychology)
- **NCSTRL** (University of Cornell - papers on Computer Science from 120 Organizations)
- **NDLTD** (electronic theses and dissertations)
- **RePEc** (papers on Economics)

# OAI Key Issues

- To solve the problem of interoperability among the e-prints archives
- Very simple, low-barrier to entry interface that could shift implementation complexity and operational processing load away from the archives

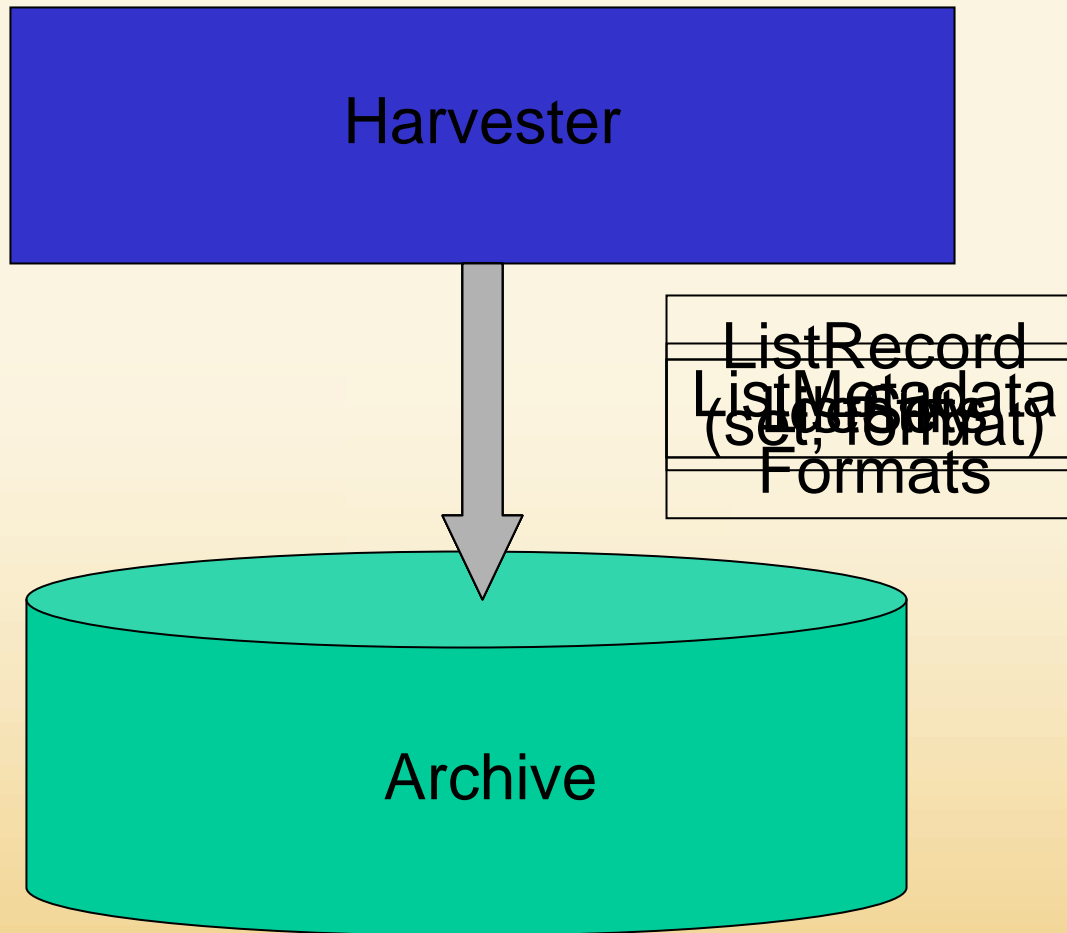
# OAI information model



# OAI Metadata Harvesting Protocol

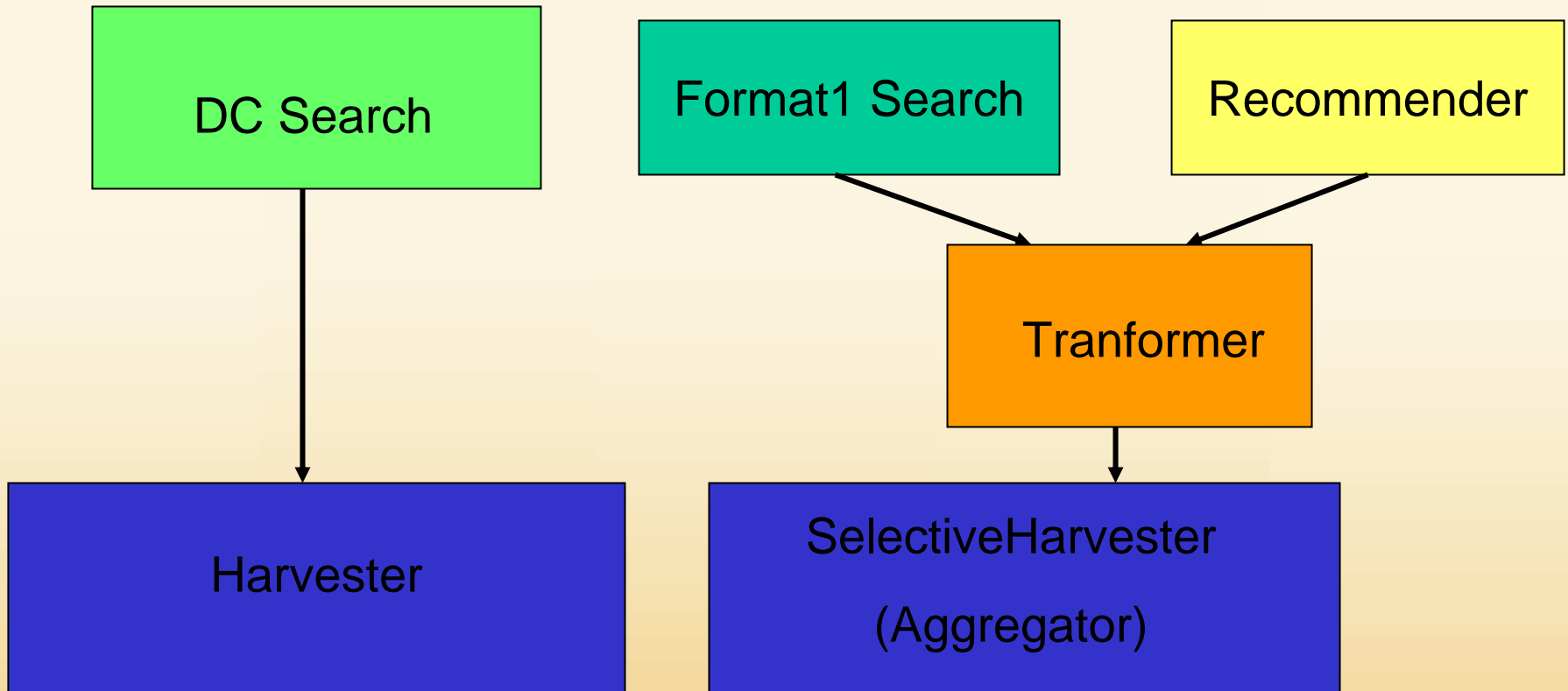
- Protocol requests
  - Identify
  - ListSets
  - ListMetadataFormats
  - ListRecords
  - ListIdentifiers
  - GetRecord
  
- HTTP-embedded
  
- XML response format

# Operational model





# Services



# OAI compliant data and service providers

- ≈ 120 archives have implemented OAI-PMH
- ≈ 20 OAI compliant services already available
  - cross-archives search
  - recommendation
  - personalized dissemination
  - etc.

<http://www.openarchives.org>

# OAI-PMH General purpose mechanism

- Agnostic to assumptions about
  - type of content
  - economic model
  - intellectual property constraints
  - system of information flow
  - ...
  
- “Any organization offering networked access to information resources may find OAI-PMH a new way to make content available to a variety of innovation service provider”

*Clyfford Lynch, ARL 217, August 2001*

# Use of OAI in the library domain

OAI-PMH approach has been perceived by the library community, especially by the academic and scholarly library communities, as a promising approach that is capable of combining the best of library and Internet techniques into a wholly new model for accessing library resources

# Experimentation

- Digital Library Federation
- Library of Congress
  - American Memory
  - Prints&Photograph Division Online Catalog
- EU-funded projects
  - The European Library
  - MALVINE
  - LEAF

# Outcomes of the experimentation

“The result of this experimentation has shown that the protocol was straightforward to implement and the harvesting traffic has no perceptible effect on the primary users of the American Memory project”

DLF Evaluation of the OAI, 2001

# Use of OAI-PMH in the conventional archive domain

- Many archive institutions have begun to aggregate their finding aids
- Behind libraries in understanding the potential and developing practical implementations for achieving interoperability
- Cross-access through portals

# Experimentation

Few conventional archives yet know about OAI-PMH and even fewer are using it

- University of Illinois at Urbana-Champaign (USA)
- AIM25 (UK)
- Bright Sparcs (Australia)
- Access to Archives - A2A (UK)



# Outcomes of the experimentation

UIUC investigated conversion from Encoded Archival Description (EAD) to Dublin Core

- difficult mainly because of inconsistency with which archivists have employed EAD
- OAI records could possibly mitigate the encoding differences found between institutions and between the finding aids of different cataloguers
- barrier in exposing EAD description into Dublin Core when attempting to transfer the different EAD hierarchies

# Outcomes of the experimentation

- UIUC suggests to use OAI-PMH for fonds or collections level descriptions only
  - More similarities of structure and approach at this level of description
  - Fonds level descriptions are not hierarchical
  - This approach may reduce concerns about possible loss of control of intellectual property rights in the catalogue data

# Lowering the barrier

- For some organizations OAI-PMH is still “a too expensive barrier”
  - lack of technical expertise
  - size of collection does not justify the investment
  
- Some organisations have technical problems
  - ISP does not allow 3rd party software
  - security considerations

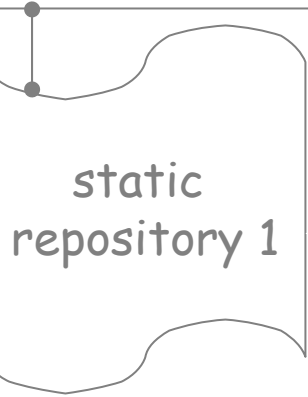
# OAI-PHM Static Repository

Patrick Hochstenback, Henry Jerez and Herbert Van de Sompel,

“The OAI-PMH Static Repository and  
Static Repository Gateway”

JCDL, Houston, May 2003

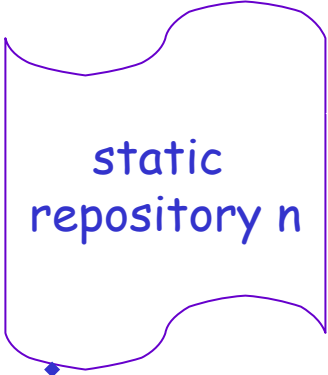
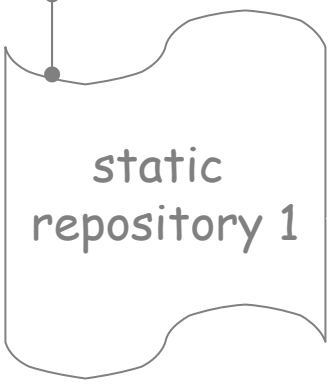
<http://an.oai.org/ma/mini.xml>



static repository n

[http:// site1.org/mini/file1](http://site1.org/mini/file1)

<http://an.oai.org/ma/mini.xml>



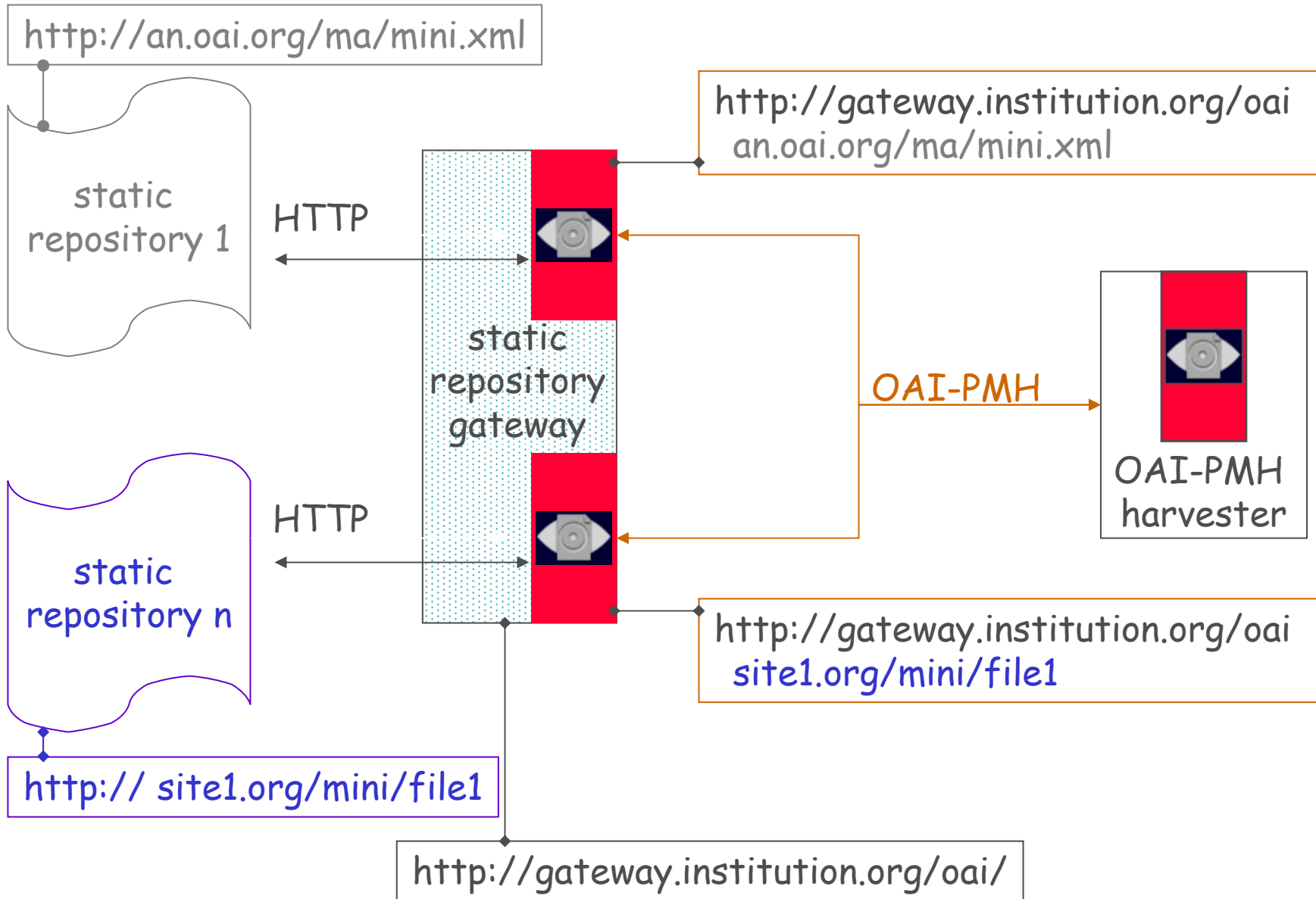
<http://site1.org/mini/file1>



<http://gateway.institution.org/oai/an.oai.org/ma/mini.xml>

<http://gateway.institution.org/oai/site1.org/mini/file1>

<http://gateway.institution.org/oai/>



# OAI Static Repository early adopters

- Open Language Archives (OLAC) Community
- Union Catalogues in Belgium, Brazil and United States
- Institutions collaborating with the Digital Library Federation (DLF) and the National Science Digital Library (NSDL)



# Organizational issues

- Intellectual Property Rights
- Metadata formats and vocabularies
- Metadata quality
- ...

# Economical issues

- Cost of adoption
  
- Business models
  - within the same domain
    - remote access to finding aids
    - access to the originals
    - digitization on demand
  
  - interoperability with other domains
    - portals to resources of different institutions

# Conclusions

- Many domains push the open archives approach as a mean for promoting their holdings
- The metadata harvesting is a scalable solution for easily implementing this approach
- The OAI-PMH is a general technical solution that applies to many different domains
- The realization of the open archive approach now depends on the solution of organizational and economical issues

# Some References

“Open Access to Hidden Resources” –  
Proc. 2nd Open Archives Forum Workshop, Lisbon, Dec.2002

“How Real Archivists can learn to love OAI”  
by George MacKenzie and Goran Kristiansson

[http:// www.oaforum.org](http://www.oaforum.org)