

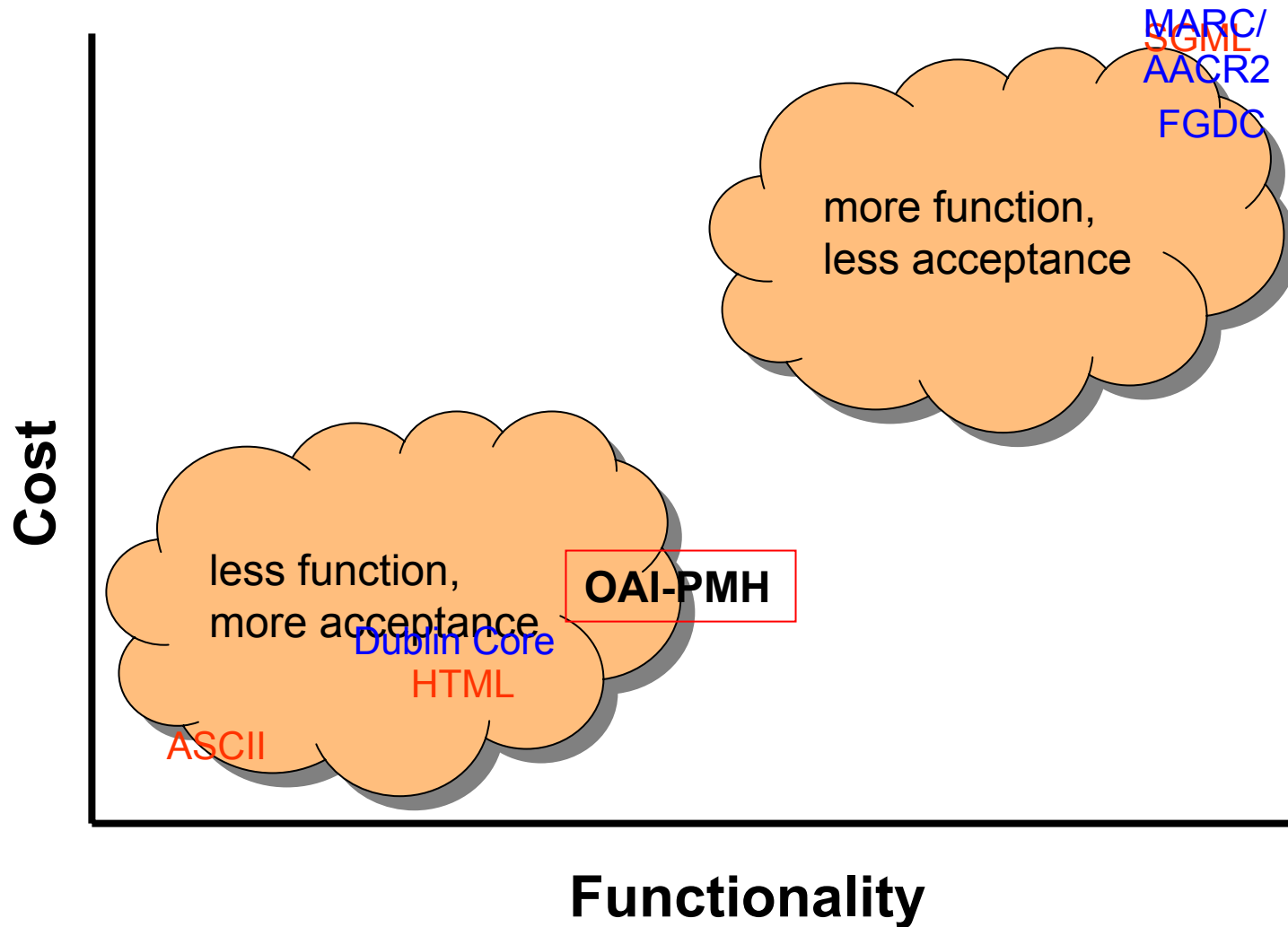


[www.openarchives.org](http://www.openarchives.org)

# The Open Archives Initiative: a low-barrier framework for interoperability

Carl Lagoze  
Computing and Information Science  
Cornell University  
[lagoze@cs.cornell.edu](mailto:lagoze@cs.cornell.edu)

# Interoperability Trade-offs



# The Open Archives Initiative

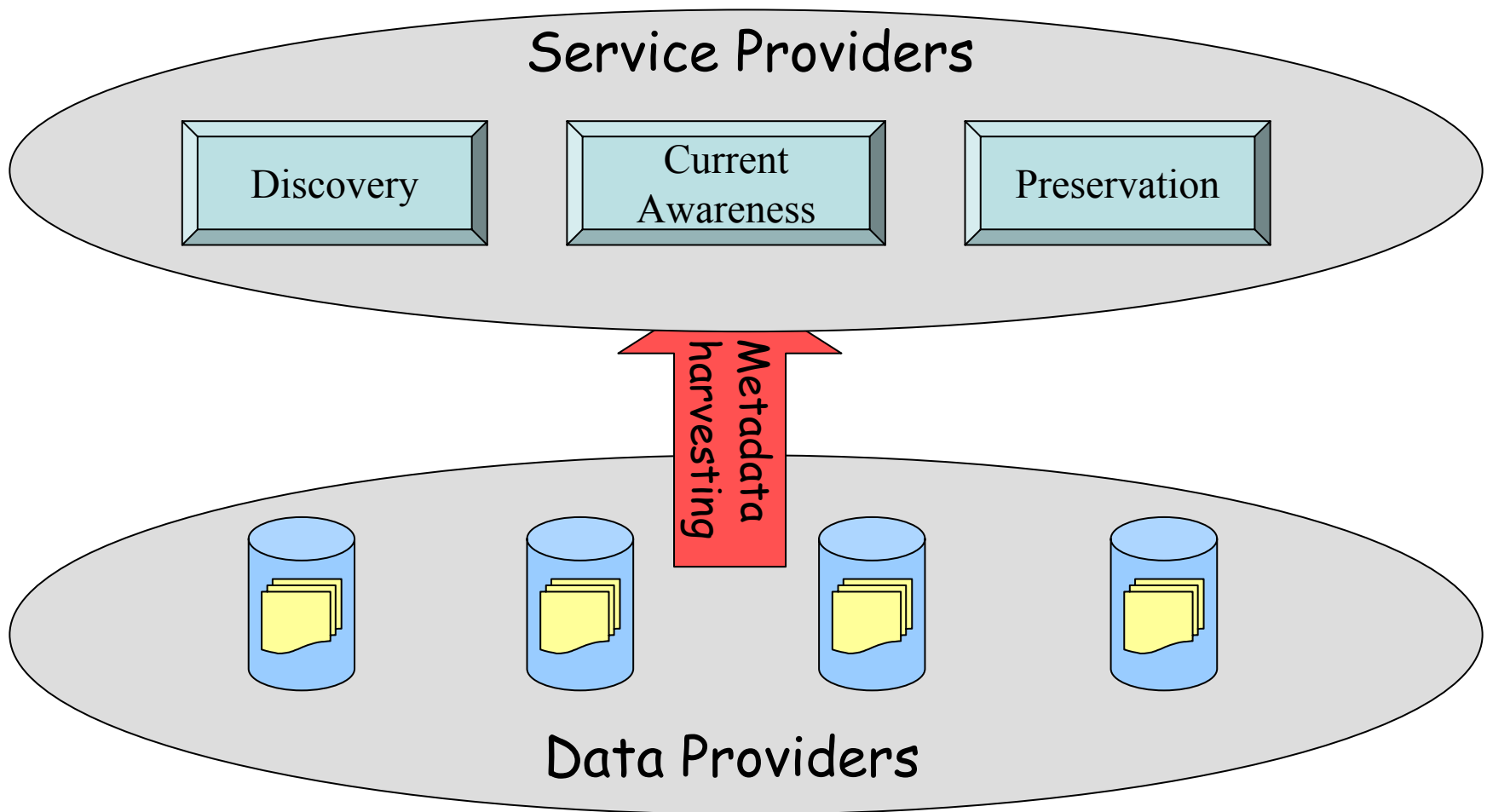
The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the **efficient dissemination of content**. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. ... The fundamental technological framework and standards that are developing to support this work are, however, independent of the both the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in **opening up access to a range of digital materials**.

OAI Mission Statement

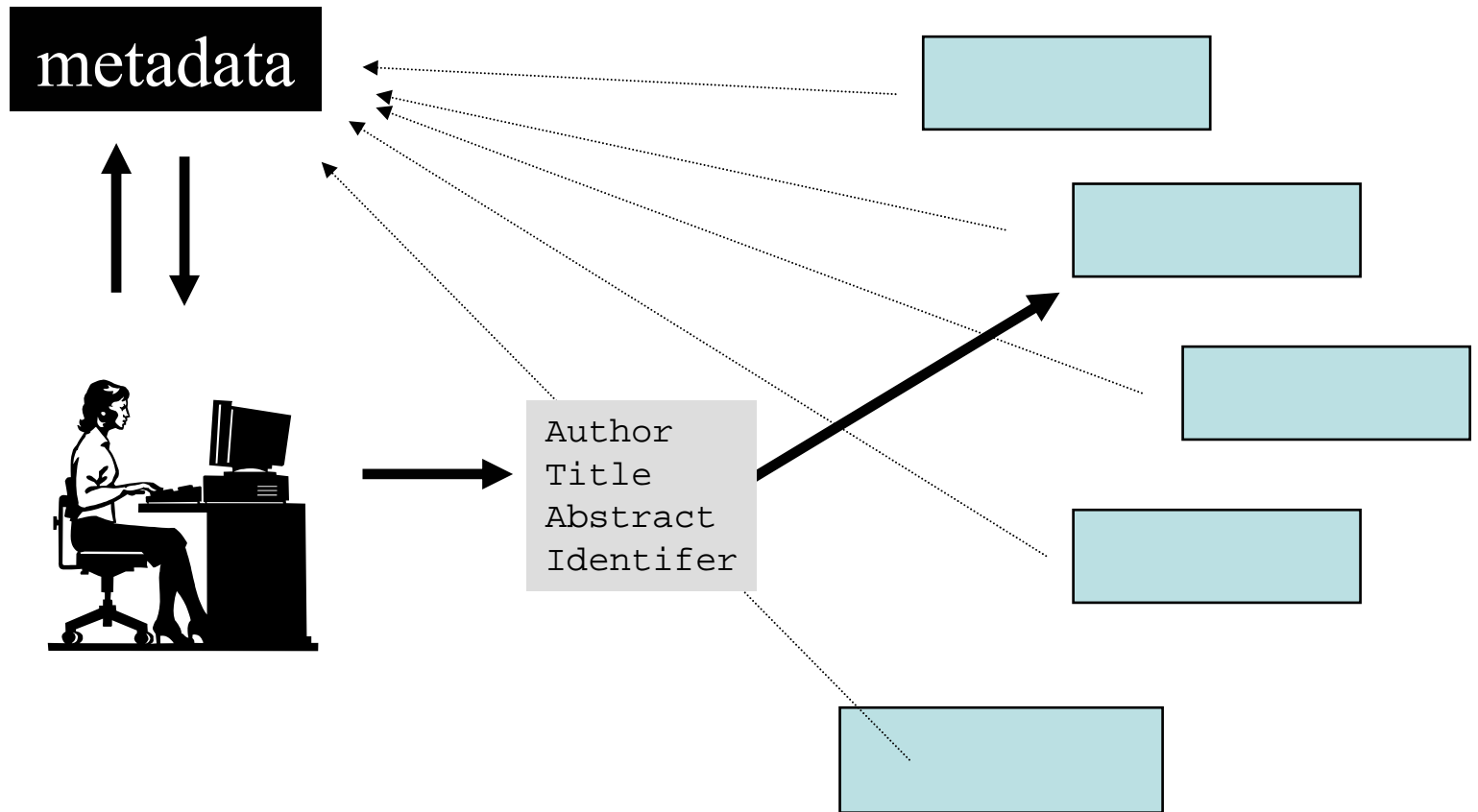
# OAI Protocol for Metadata Harvesting (OAI-PMH)

The goal of the Open Archives Initiative Protocol for Metadata Harvesting ... is to supply and promote an **application-independent interoperability framework** that can be used by a variety of communities who are engaged in publishing content on the Web. The OAI protocol ... permits *metadata harvesting*.

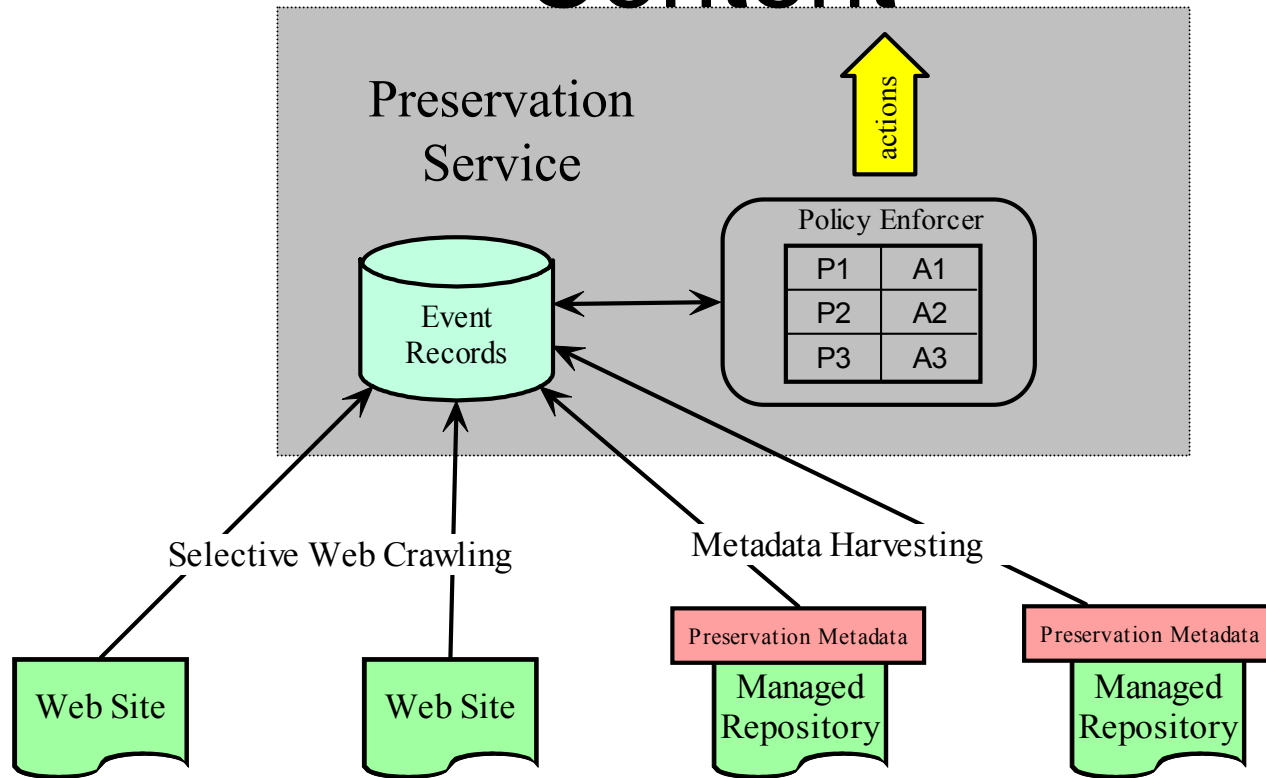
# OAI-PMH: A simple two party model for sharing structured information



# Yes, its about resource discovery over distributed collections



# Facilitating/Monitoring Longevity of Distributed Content



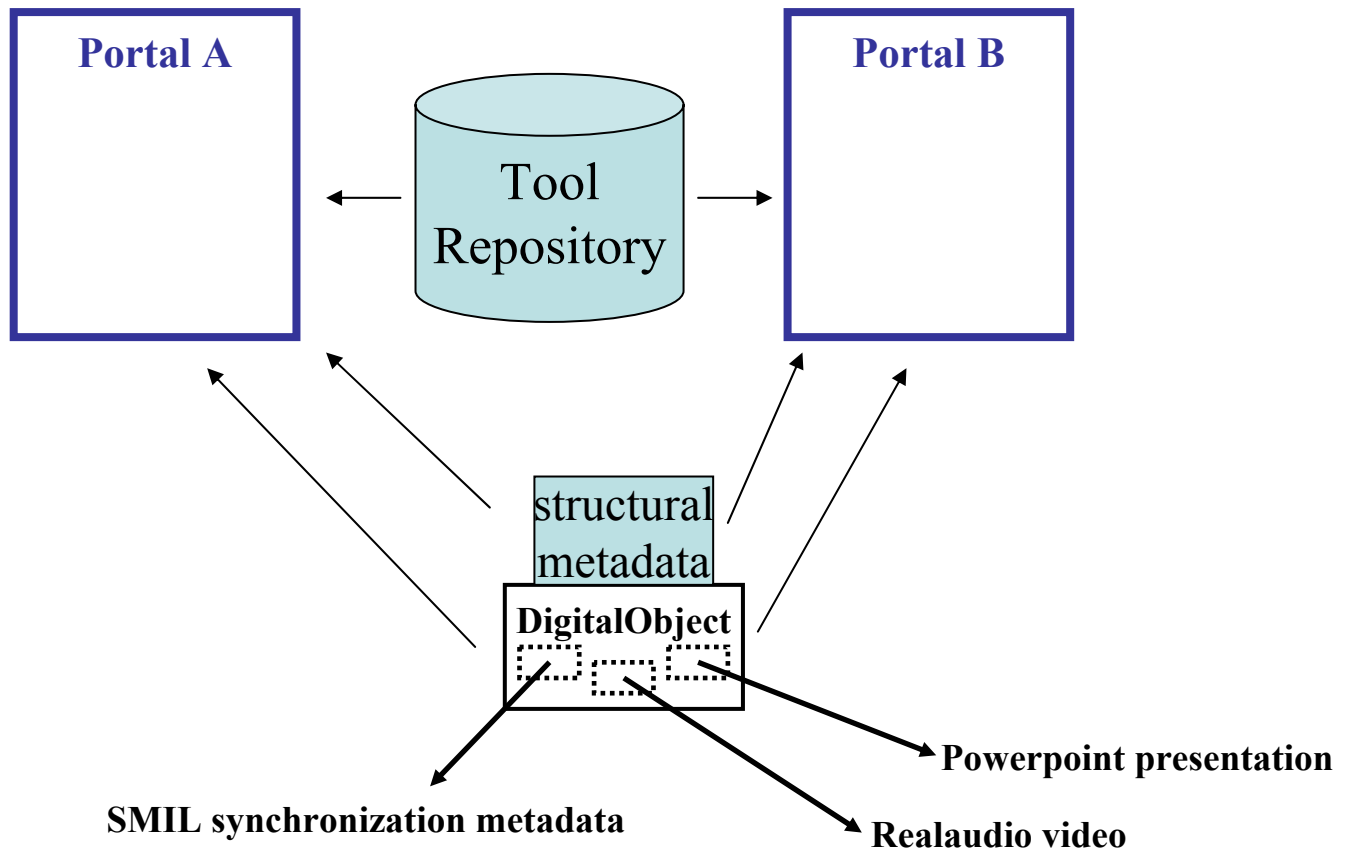
# Personalization of Content

## View A:

- View Slides
- View Video
- View synchronized presentation using applet

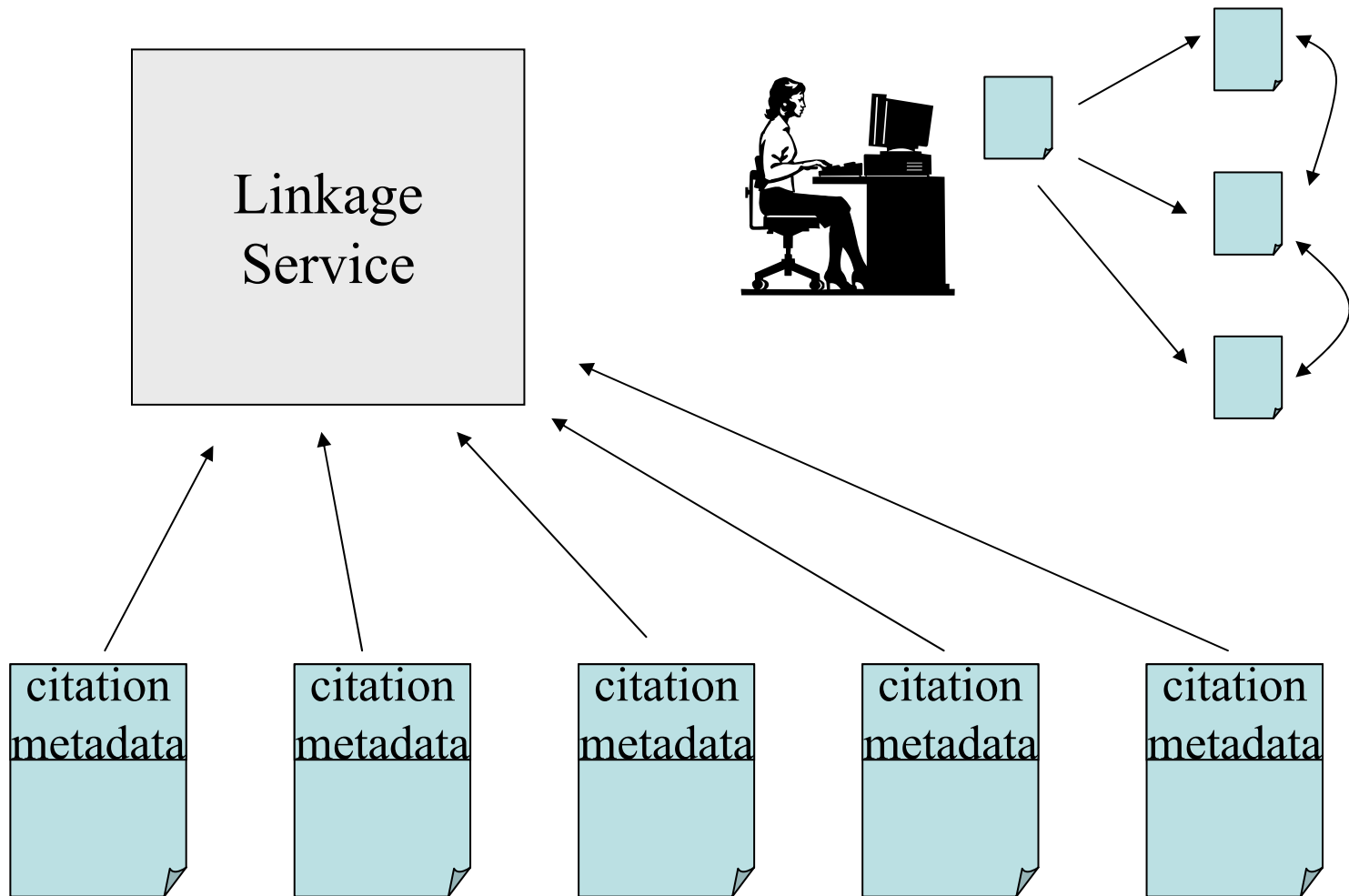
## View B:

- Get Transcript of Audio
- Search for keyword
- Get Slides translated to French





# Cross-Repository Reference Linking



# Brief History of the OAI

- Motivation: expand impact of ePrint archives through federation
- 1999: Santa Fe Meeting and convention
- 2000: OAI-PMH formation
  - Scope broadens
  - OAI steering committee
- 2001 OAI-PMH v. 1.0 “experimental” protocol
- 2002 OAI-PMH v. 2.0 “stable” protocol

# OAI-PMH Key technical features

- Deploy now technology – 80/20 rule
- Simple HTTP encoding
- Foundation of established XML standards
- Multiple metadata formats
- Repository partitioning (sets)
- Selective harvesting (sets and dates)
- Clean partition between core and implementation-specific extensions
  - Multiple item-level metadata
  - Collection level metadata

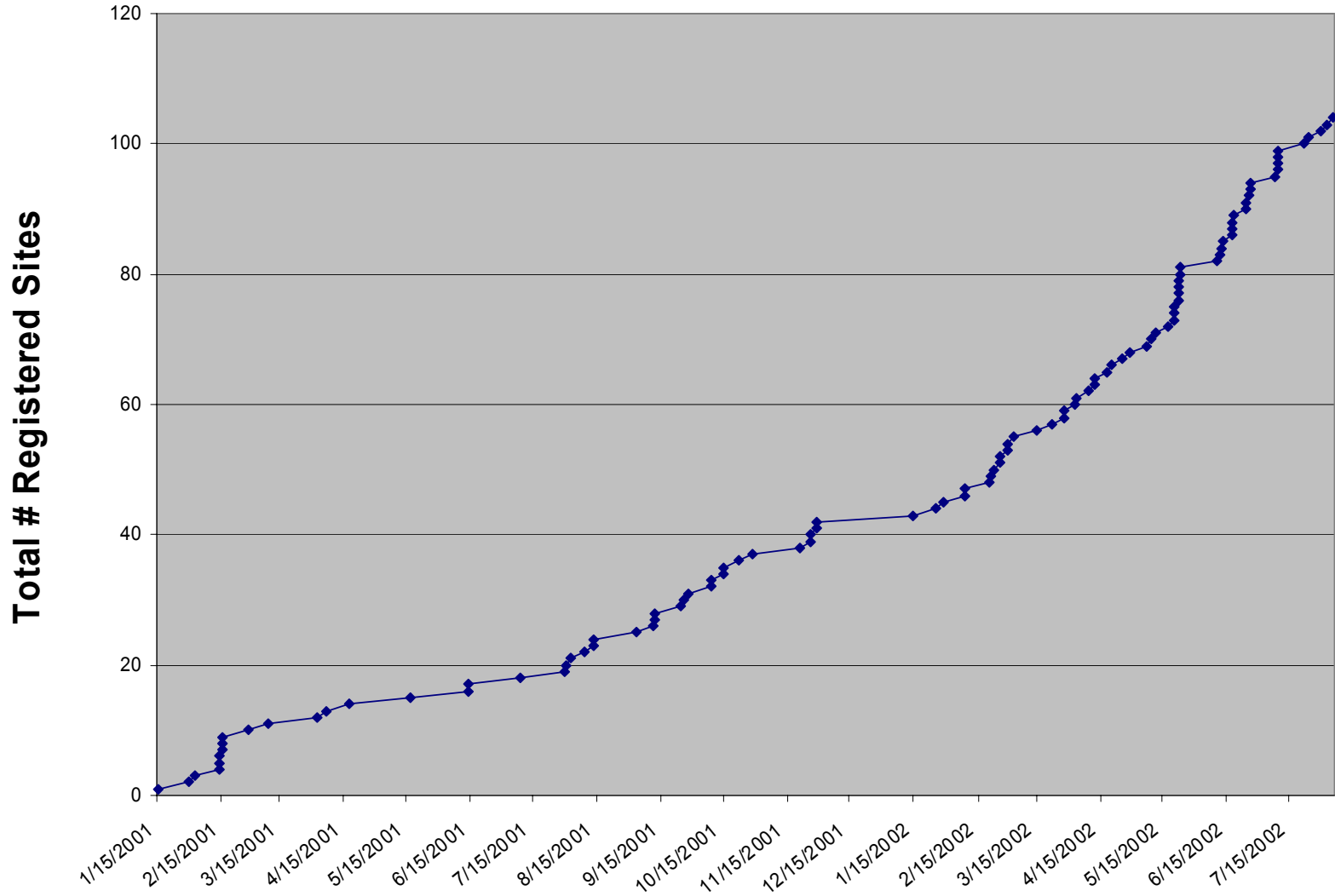
# OAI Verbs

- **Identify** – repository characteristics
- **ListMetadataFormats** – DC required
- **ListSets** – repository partitioning
- **ListRecords** – (selectively) harvest metadata
- **ListIdentifiers** – (selectively) harvest metadata identifiers
- **GetRecord** – known item retrieval

# Measures of Success

- Registered data providers
- Adoption by major projects
- Acceptance as 'fundamental infrastructure' for research and implementation

# OAI Registered Data Providers



# National Science Digital Library (NSDL)

- Very large scale distributed digital library
  - 1,000,000 users
  - 10,000,000 items
  - 100,000 collections
- Large institutional and funding commitment
  - \$25M+ funding
  - Over 80 collaborating institutions
- Technical infrastructure builds on OAI-PMH foundation
  - Aggregation and dissemination of metadata
- <http://www.nsdsl.org>

# Fundamental Infrastructure

- Eprints.org servers
  - e.g., Cal Tech ePrint framework
- Open language archives community
- JISC FAIR awards
- Mellon OAI service providers
- ECDL , DCADL, JCDL research papers



# Some questions remain

- Is OAI-PMH really low-barrier infrastructure?
  - NSDL experience indicates that significant barriers remain
- Utility of core metadata (unqualified DC)
  - NSDL and other experience raises doubts
- Utility outside of resource discovery
  - Certification, Reference linking, etc.

# Future Questions and Directions

- “Standardization”?
  - De-facto?
  - Maintenance agency?
  - Formal standards agency?
- Future OAI-PMH versions?
  - Expanded functionality?
- Targeted ‘application profiles’?
  - ePrints community?