

# 3<sup>rd</sup> OAF Workshop - Berlin, March 28<sup>th</sup> 2003



the Open Archives Initiative  
<http://www.openarchives.org>

Herbert Van de Sompel

Los Alamos National Laboratory - Research Library



# outline

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000011000000



- origin & evolution & future
- innovative use of OAI-PMH



001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

origin & evolution & future



# core motivation

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- libraries are there to facilitate access to scholarly information

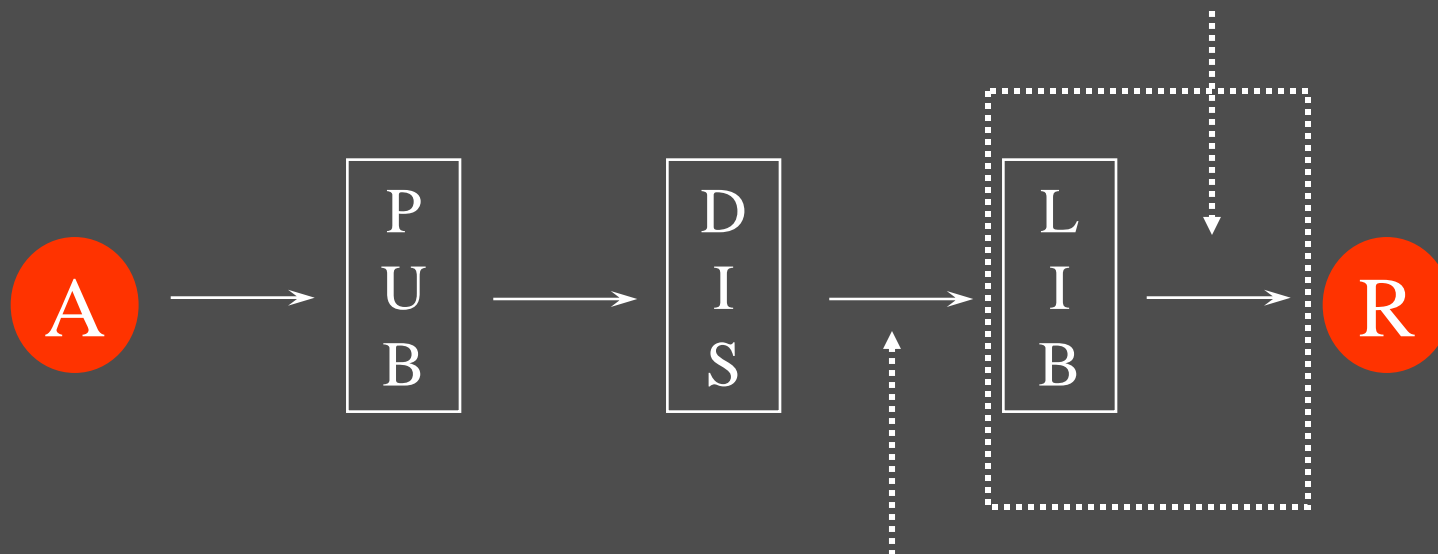
we were trying hard, but ...



# library position

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

optimizing the output



the input is far from optimal



# the journal system

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- **serials crisis:** increasing journal prices limit, rather than broaden, access to scholarly research
- **IP drain:** faculty signs away copyright
- **publication delay:** journal system can not cope with increasing volume of scholarly output
- **criticism of peer-review:** suppresses ideas, outcome criticized
- **inertia:** system is self-stabilizing



# core motivation

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- libraries are there to facilitate access to scholarly information

we were trying hard, but ...  
could there be other (better) ways?



# preprint systems

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- xxx e-print archive  
(Physics - 1991 - Los Alamos - Ginsparg)
- RePEc  
(Economy - Surrey U - Krichel)
- NCSTRL  
(Computer Science - Cornell U - Lagoze)
- NDLTD  
(Theses - Virginia Tech - Fox)
- CogPrints  
(Cognitive Sciences - Southampton U - Harnad)





# core motivation

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- libraries are there to facilitate access to scholarly information



local perspective

global perspective



- libraries as institutions are there to guarantee access to information



# alternative library position?

00111010001001111010000100010001111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

capture & share the input... using the global network



action!

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000000

economy

technology

scholarly  
communication

sociology

law



# action!

0011101000100111101000010001000111111000110011100001100011100000000010111101001110010001111001110100010011110100001000100011111100011001110000110000000

economy

establish a technological basis that

technology



sociology

allows addressing the other issues.

law



# action!

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- UPS Prototype [1999]:
  - Van de Sompel, Krichel, Nelson
  - cross-repository searching
  - SFX/OpenURL linking
  - insights regarding lack of interoperability
  - recommendation: metadata harvesting
- launch Open Archives Initiative
  - Ginsparg, Luce, Van de Sompel



# Luce \* Van de Sompel \* Ginsparg



# since then ...

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

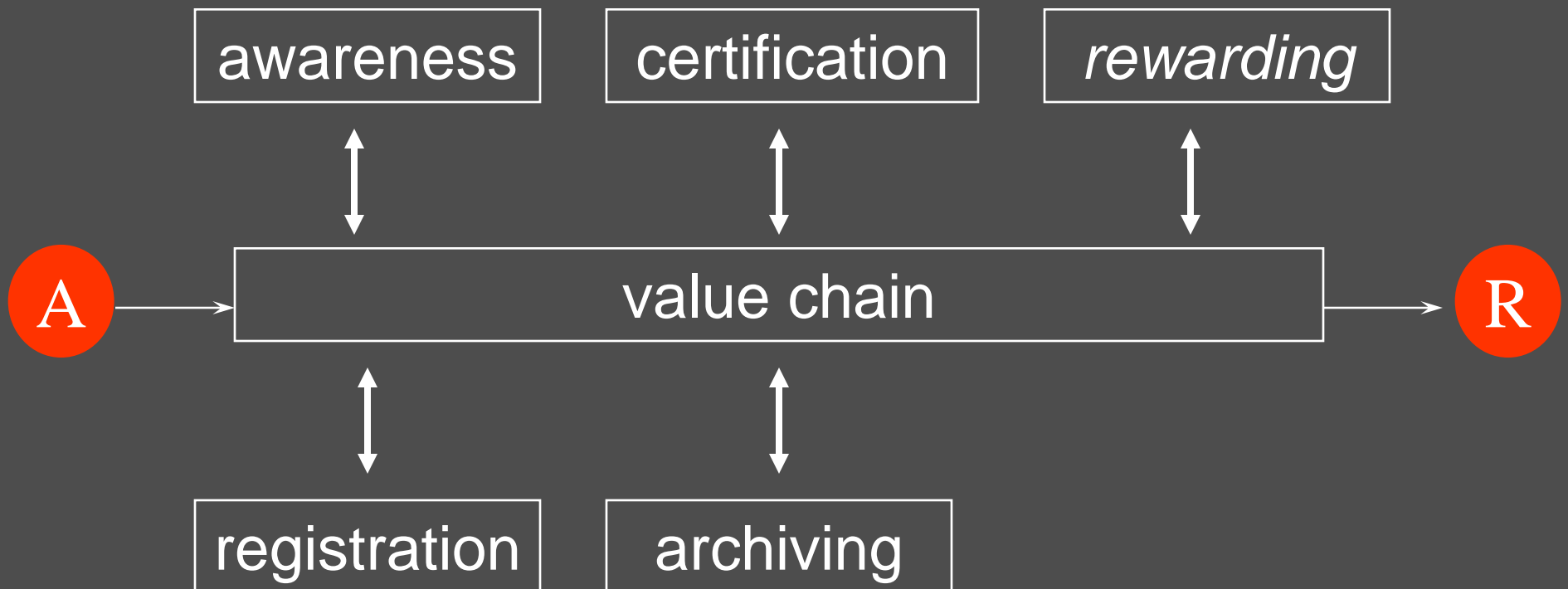
- OAI:
  - Lagoze, Van de Sompel and lots of great people
  - Santa Fe Convention [2000]
  - OAI-PMH v.1 [2001]
  - OAI-PMH v.2 [2002]
- OAI-PMH:
  - simple, generic protocol to harvest structured data
  - HTTP based
  - responses are valid XML instance documents
  - unqualified Dublin Core as mandatory metadata format
  - /use of other metadata formats encouraged



# since then ...

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- scholarly communication:
  - deconstructed communication system
  - SPARC white paper on institutional repositories





# and ... ?



great

- OAI-PMH:
  - successful
  - used in different communities (scholarly communication, library collections, museums, e-learning, ...)
  - project funding opportunities
  - fundamental piece of information infrastructure (Clifford Lynch)



# and ... ?



SO & SO

- scholarly communication:
  - OAI-PMH and associated tools make moving to action easier
  - work of the OAI has helped raise awareness:
    - some national forums discussing transformation of scholarly communication
    - OAI Workshop CERN
  - modest attempts to new communication models (Caltech, CDL, BMC, ...)
  - the big change didn't happen
  - are libraries losing focus, e.g. learning systems ?



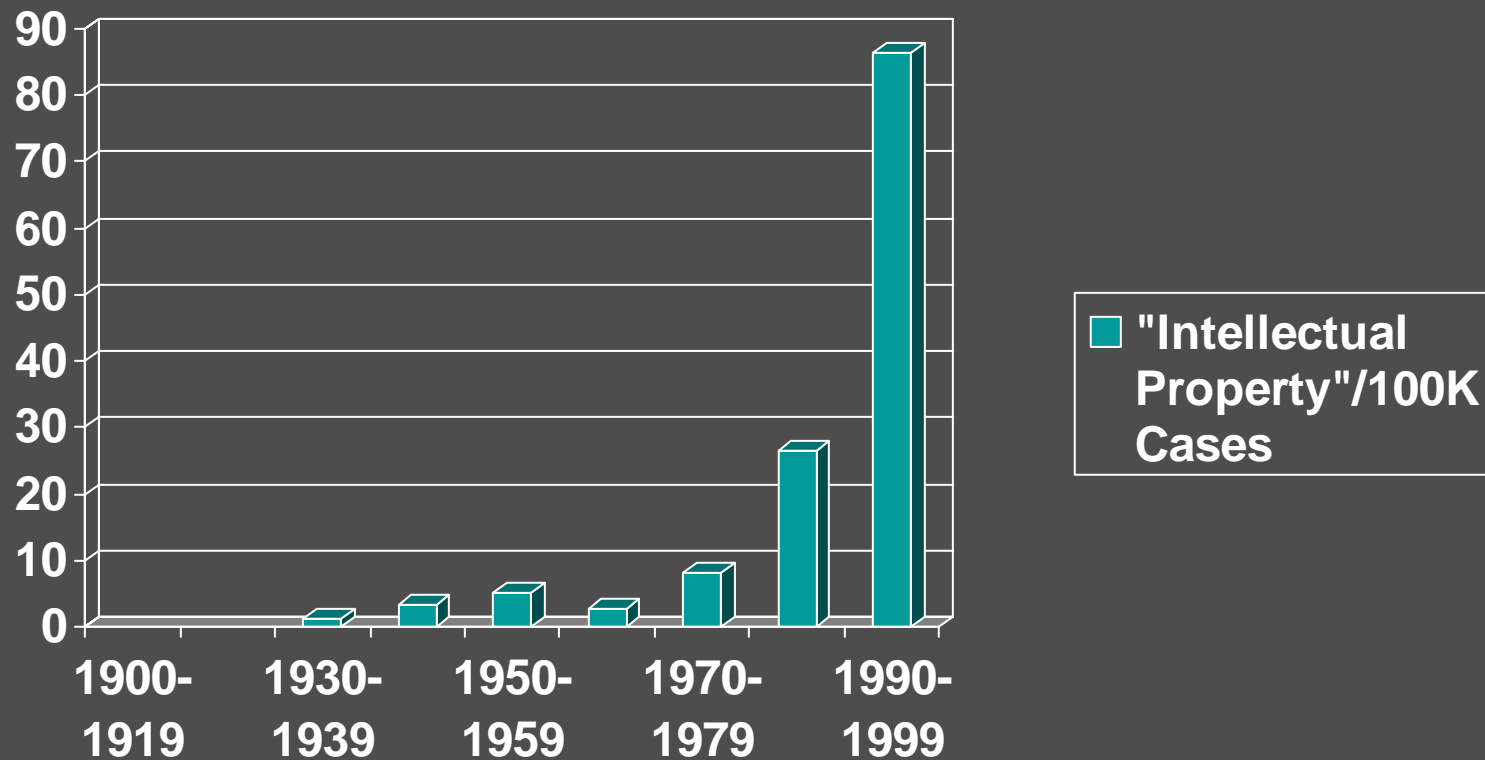
# should we still act?

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- the initial motivations remain valid
- consortium buying starts to reveal its inherent dangers
- the IP issue becomes ... scary
  - increasing amount of IP cases
  - copyright debate steered by intermediaries (Big Media), no longer something between creators and consumers







# occurrences of the phrase "intellectual property" per 100,000 U.S. Federal Cases



"Professor Hank Greely" Cited in Lessig, L. The future of ideas: the fate of the commons in a connected world. NY, Random House, 2001. P. 294.

You're probably familiar with the phrase, "All rights reserved," and the little (c) that goes along with it. Creative Commons wants to help copyright holders send a different message: "Some rights reserved."

For example, if you don't mind people copying and distributing your online image so long as they give you credit, we'll have a license that helps you say so. If you want people to copy your band's MP3 but don't want them to profit off it without your permission, use one of our licenses to express that preference. Our licensing tools will even help you mix and match such preferences from a menu of options:

-  **Attribution.** Permit others to copy, distribute, display, and perform the work and derivative works based upon it only if they give you credit.
-  **Noncommercial.** Permit others to copy, distribute, display, and perform the work and derivative works based upon it only for noncommercial purposes.
-  **No Derivative Works.** Permit others to copy, distribute, display and perform only verbatim copies of the work, not derivative works based upon it.
-  **Copyleft.** Permit others to distribute derivative works only under a license identical to the license that governs your work.

seems to resonate better with scholarly communication

# should we still act? yes!

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- the OAI plans go back to its eprint roots
- build upon the OAI-PMH to move beyond interoperability at the level of discovery:
  - references
  - usage logs
  - certification metadata
  - rights metadata
- some of the work conducted by OAI, some coordinated with other projects
- part of a project that will look at evolution of metadata in chain of OAI-PMH systems



001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

## innovative use of OAI-PMH

- OAI Static Repository
- OAI-PMH-based access to DL usage logs
- OAI-PMH-conformant OpenURL Registry



# OAI Static Repository

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- OAI-PMH is low-barrier protocol
- nevertheless, implementation is sometimes not trivial:
  - size of collection does not justify the investment
  - ISP does not allow 3rd party software
  - security considerations





# OAI Static Repository

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- research on lowering barrier even further
  - make metadata available in XML files (not dbases)
  - put XML file on web-server
  - make XML file OAI-PMH harvestable
- 2 tracks:
  - autonomous data provider
  - dependent data provider



# OAI Static Repository

00111010001001111010000100010001111110001100111000011000111000000000101111010011100100011100111010001001111010000100010001111110001100111000011000111000000

- autonomous data provider:
    - XML file on web-server
    - XSL style sheet to respond to OAI-PMH requests on web-server
    - requires:
      - native XSLT support in web server
      - XSL v.2 functionality
- => Not (yet) low barrier



# OAI Static Repository

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- dependent data provider:
  - XML file on web-server
  - depend on Gateway to respond to OAI-PMH requests
- requires:
  - *registration* with Gateway
  - Gateway implementation(s)

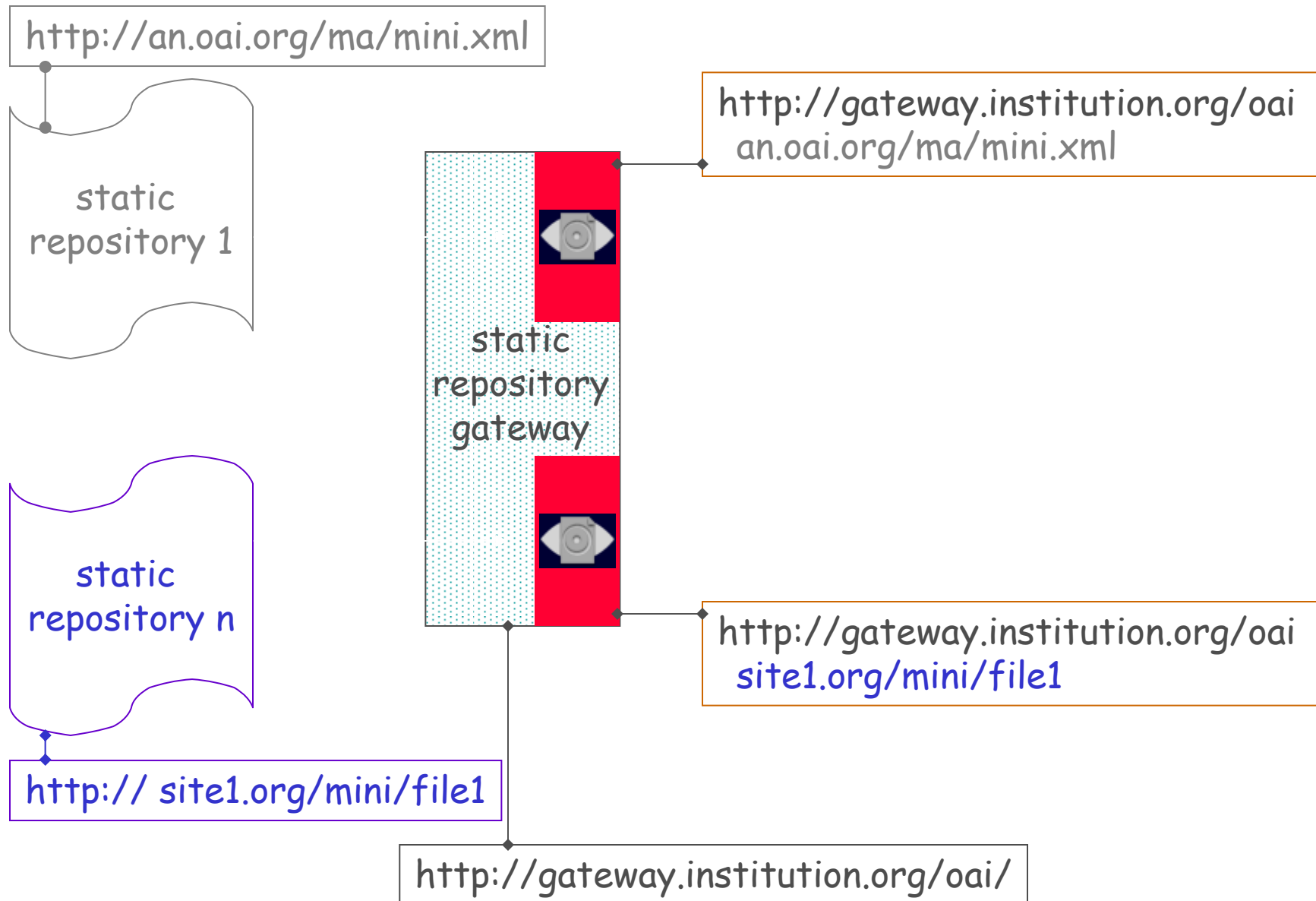


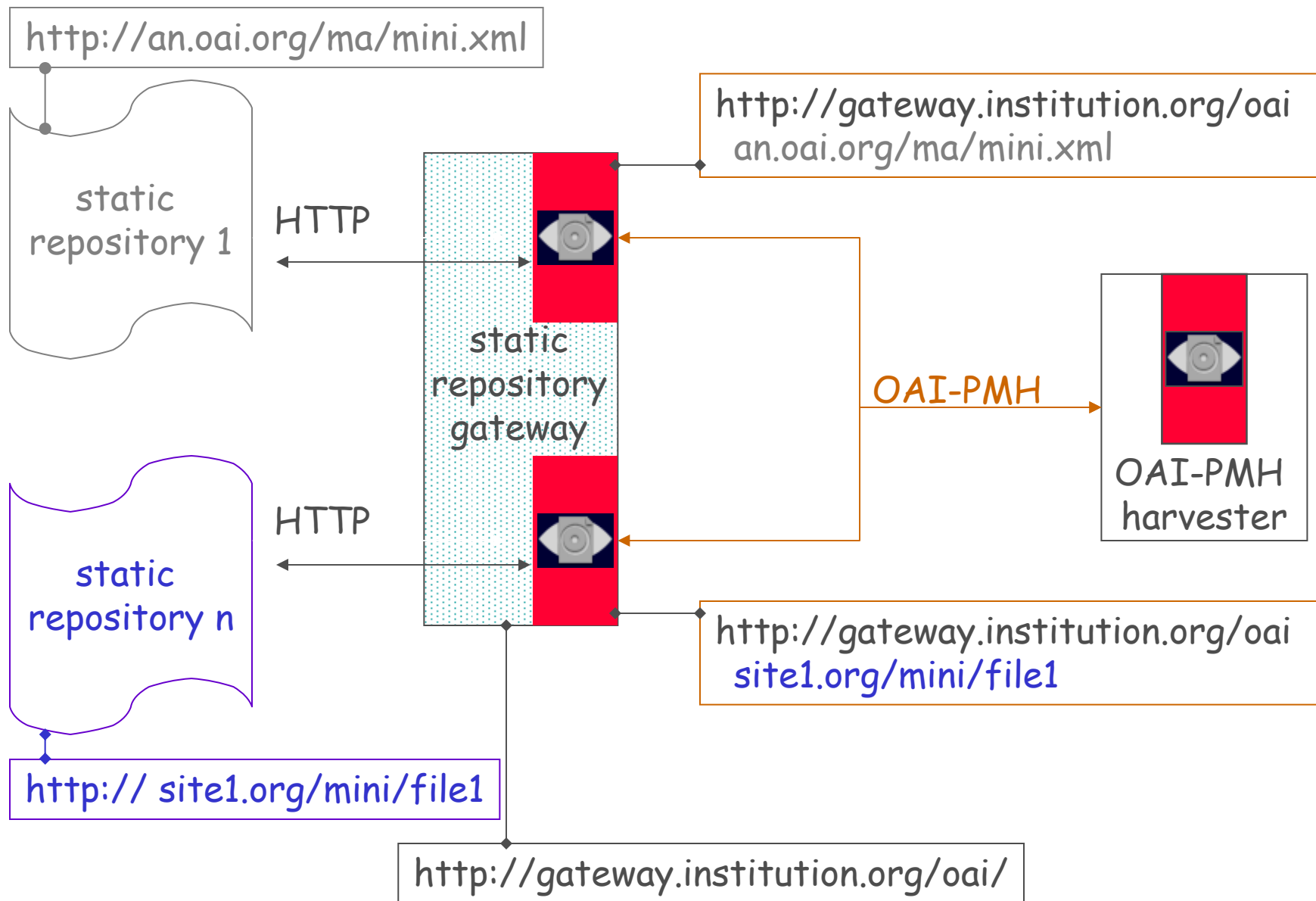
<http://an.oai.org/ma/mini.xml>

static  
repository 1

static  
repository n

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





# OAI Static Repository

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- sample OAI Static Repository
- alpha specification for OAI Static Repository
- access to demo Gateway
- draft JCDL 2003 paper



001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

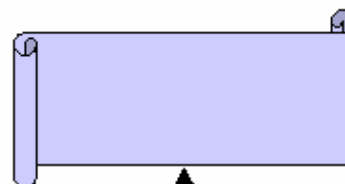
## innovative use of OAI-PMH

- OAI-PMH-based access to DL usage logs
- OAI-PMH-conformant OpenURL Registry





resource



about

item

oai:ab.org:1234

identifier

O

A

metadata  
records

I

oai\_dc

xxx

metadataPrefix

timestamp1

timestamp2

timestamp

oai\_dc  
metadata

xxx  
metadata



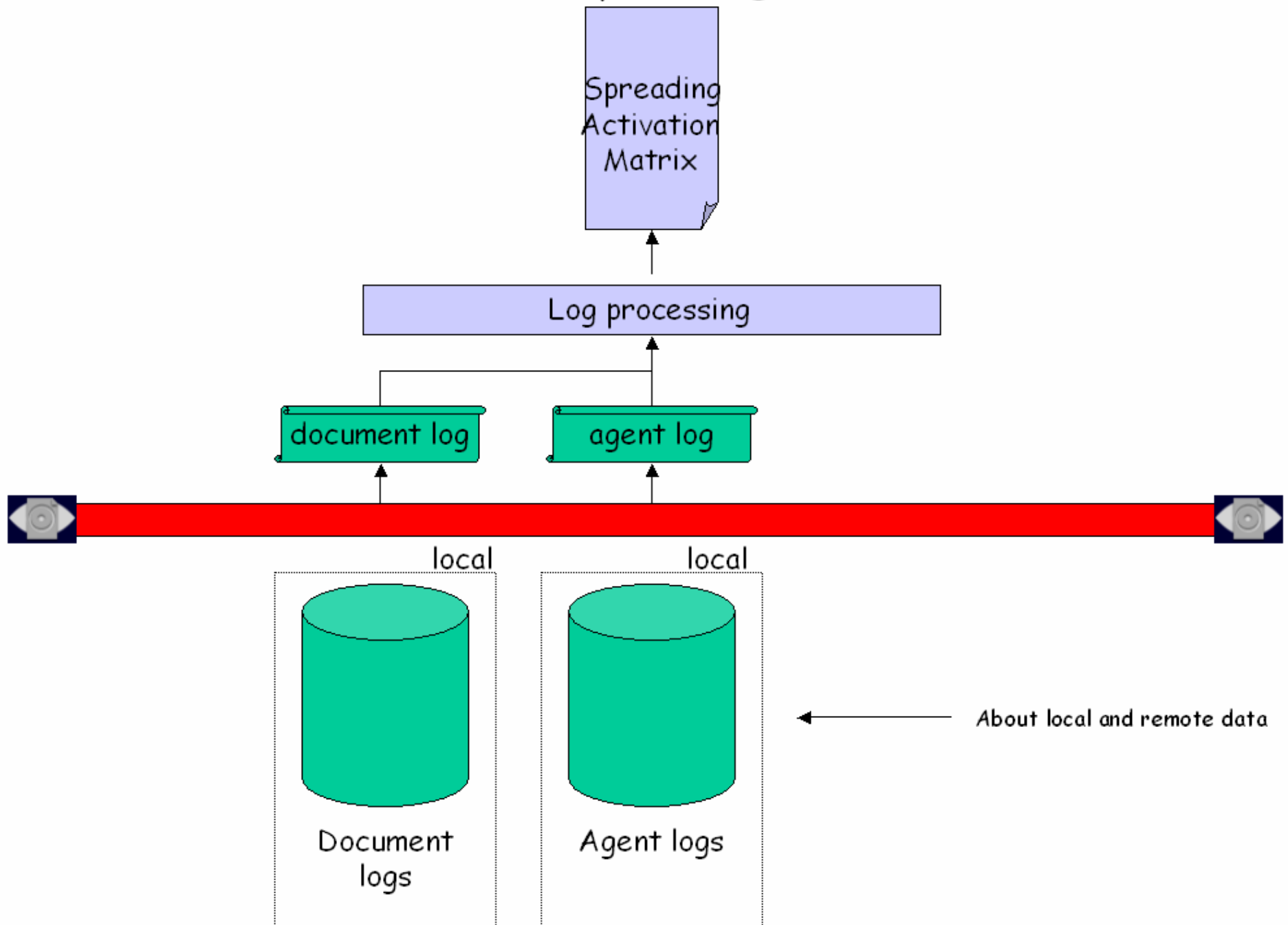
# OAI-PMH access to DL usage logs

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

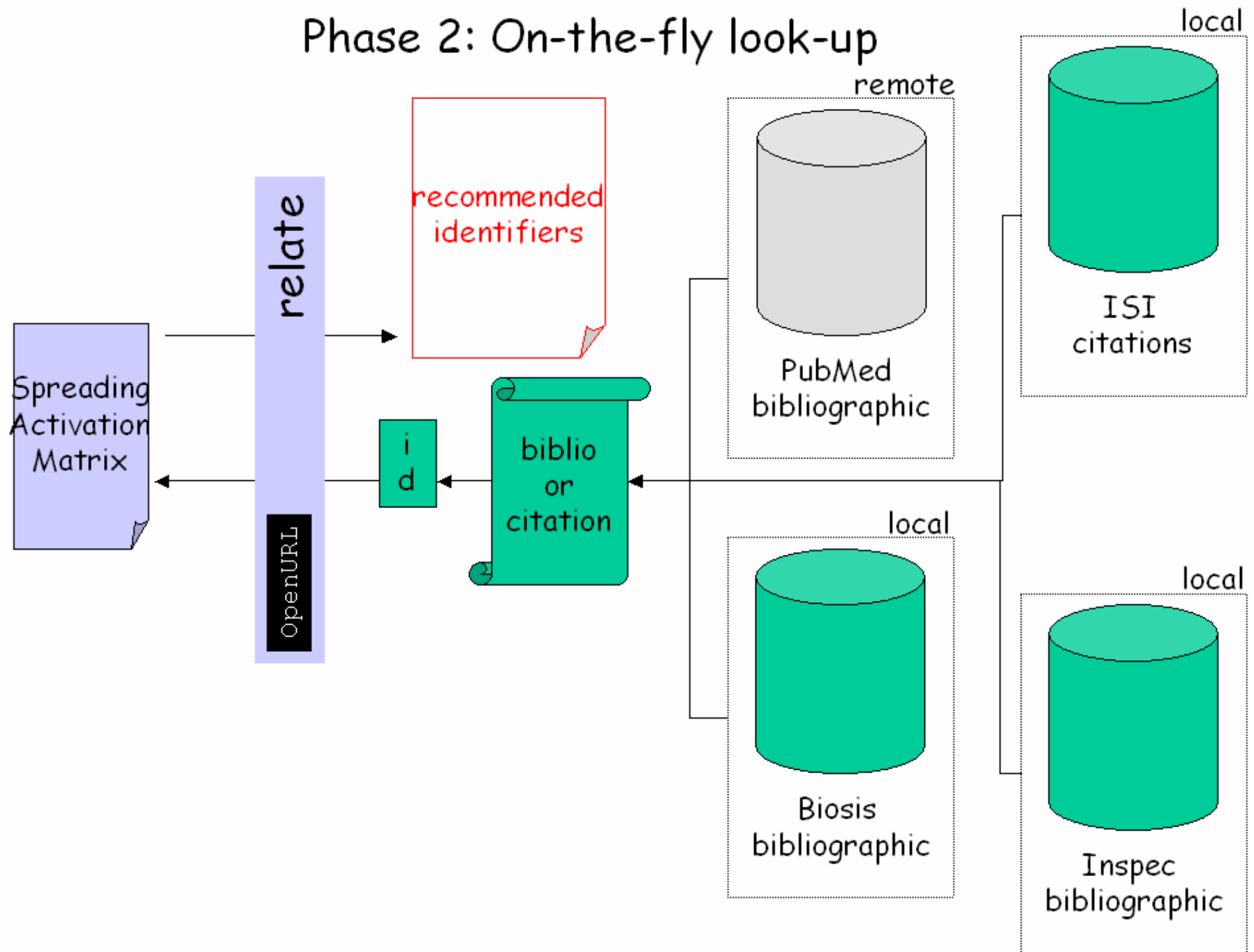
- usage logs filtered and stored in MySQL db
- accessible as 2 OAI-PMH repositories:
  - document oriented
  - agent oriented (user-proxy)
  - interlinked
- recommender system:
  - harvests logs
  - interpretes logs
  - exposes relationships (OpenURL access)



# Phase 1: creation of Spreading Activation Matrix

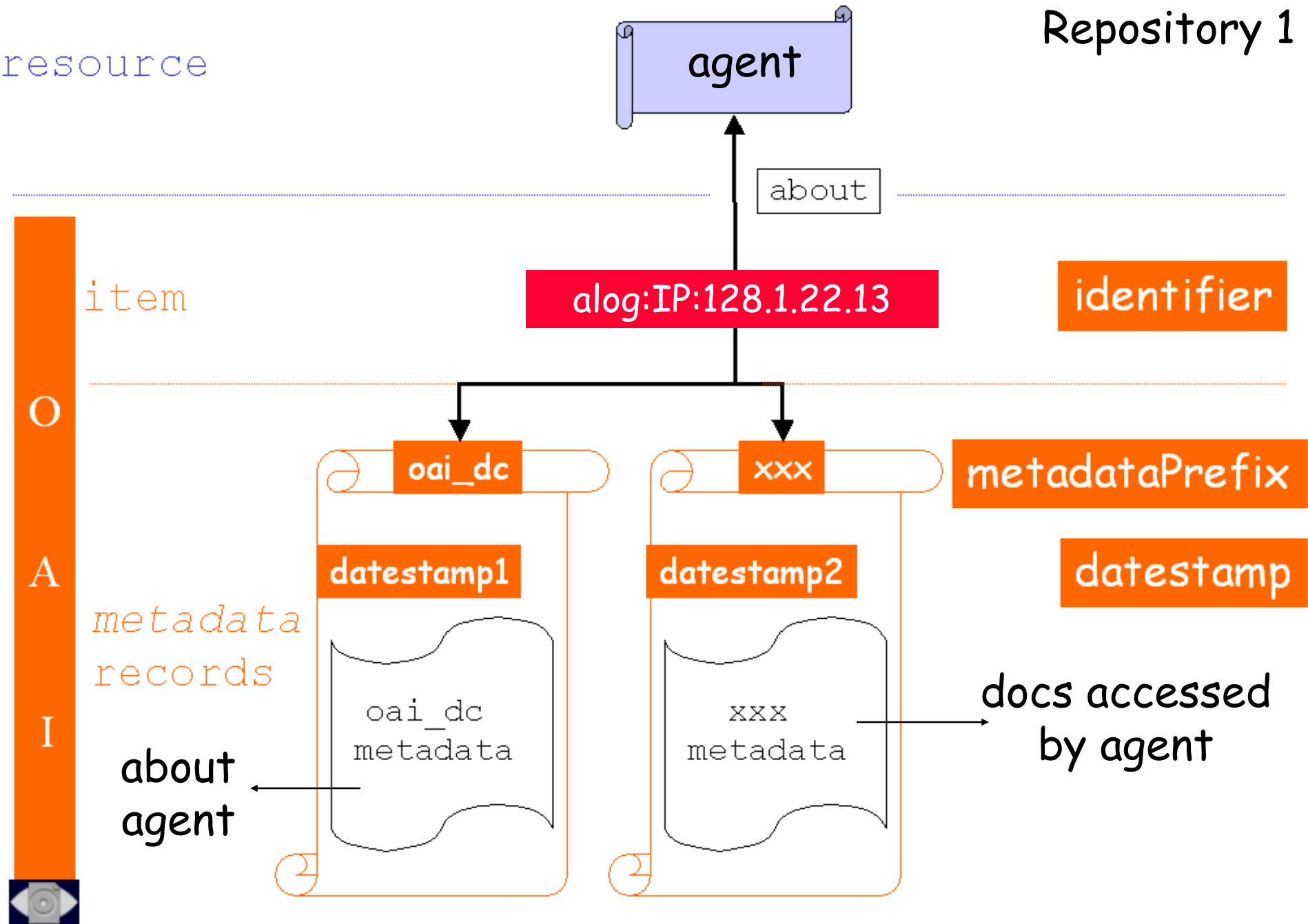


## Phase 2: On-the-fly look-up



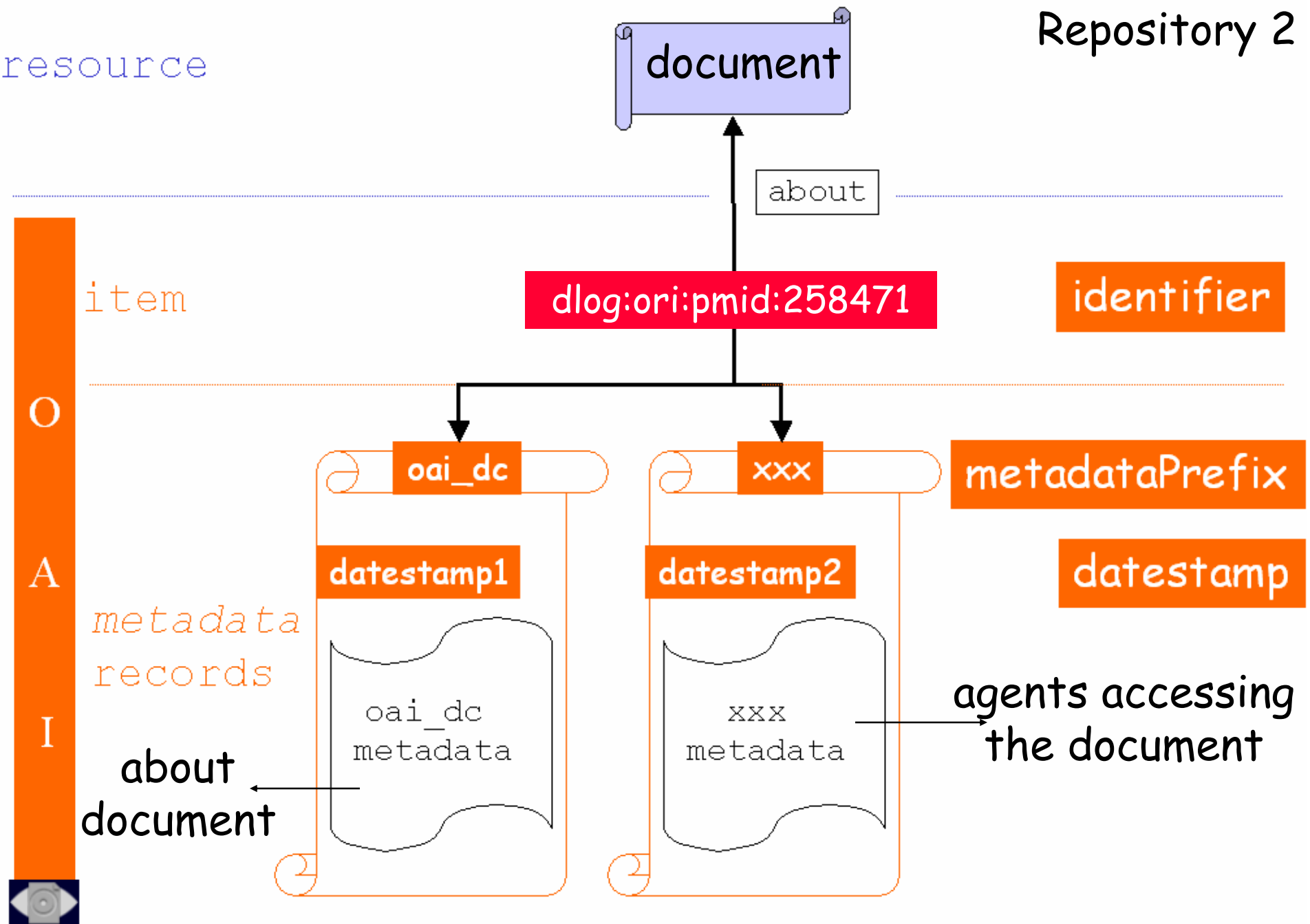
resource

Repository 1



resource

Repository 2



# OAI-PMH access to DL usage logs

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- log repository screencam



001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

innovative use of OAI-PMH

- OAI-PMH-conformant OpenURL Registry





# OAI-PMH-conformant OpenURL Registry

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- NISO OpenURL Framework builds on Registry
- Registry entry:
  - unique identifier
  - always DC record
  - sometimes XHTML or XML Schema definition



# OAI-PMH-conformant OpenURL Registry

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- Collaboration with OCLC Office of Research:
  - Registry is OAI-PMH harvestable
  - Registry is browseable through overlaying of PURL and XSLT



# OAI-PMH-conformant OpenURL Registry

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

baseURL?

verb=GetRecord &  
metadataPrefix=xsd &  
identifier=ori:fmt:xml:xsd:ctx

- access to rendered data:

[http://www.openurl.info/registry  
/xsd  
/ori:fmt:xml:xsd:ctx](http://www.openurl.info/registry/xsd/ori:fmt:xml:xsd:ctx)



# OAI-PMH-conformant OpenURL Registry

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

baseURL?

verb=GetRecord &  
metadataPrefix=xsd &  
identifier=ori:fmt:xml:xsd:ctx

- access to raw data:

[http://www.openurl.info/registry  
/docs  
/xsd  
/ori:fmt:xml:xsd:ctx](http://www.openurl.info/registry/docs/xsd/ori:fmt:xml:xsd:ctx)



# OAI-PMH-conformant OpenURL Registry

001110100010011110100001000100011111100011001110000110001110000000001011110100111001000111100111010001001111010000100010001111110001100111000011000111000000

- OpenURL Registry screencam





questions

Herbert Van de Sompel

Los Alamos National Laboratory - Research Library

