

The OAI Static Repository

A file-based approach to exposing metadata via the OAI-PMH

DLF Forum - New York, NY - May 15th 2003

Herbert Van de Sompel

Los Alamos National Laboratory



The OAI Static Repository

A file-based approach to exposing metadata via the OAI-PMH

- motivation
- nature of solution
- status
- URIs
- acknowledgments



motivation

- OAI-PMH is low-barrier protocol
- OAI-PMH favors to make it easy for Data Providers
 - bias has origins in the Santa Fe Convention



motivation

- nevertheless, implementation is sometimes not trivial:

- lack of technical expertise
- size of collection does not justify the investment
- security considerations re database
- ISP does not allow 3rd party software
- cf. OLAC, union catalogues, LoCKSS



motivation

- therefore: research to devise approaches to further lower the barrier to sharing metadata collections through the OAI-PMH



target group

- Data Providers that are unable to operate special OAI-PMH software
- Data Providers that have access to file services of Web server



nature of solution

- make metadata available in XML file of well-defined format (not dbase)
- put XML file on web-server
- make content of XML file OAI-PMH harvestable



nature of solution

- 2 research tracks:
 - autonomous Data Provider
 - dependent Data Provider
- => both can operate on the basis of the same XML file format



autonomous Data Provider

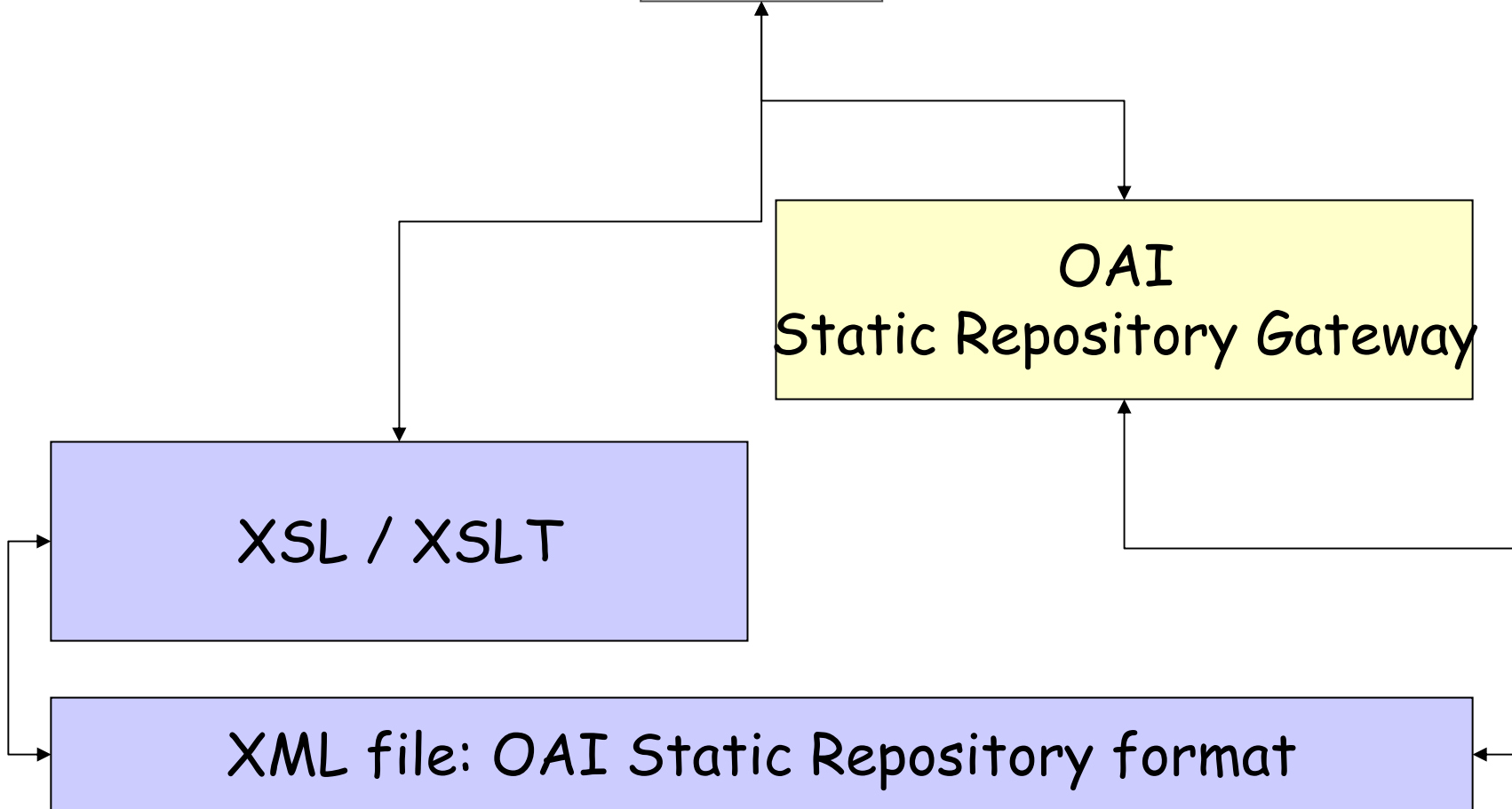
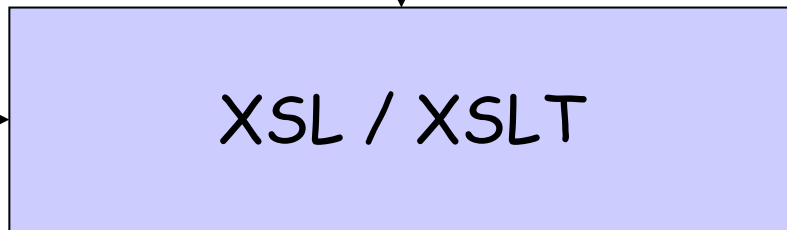
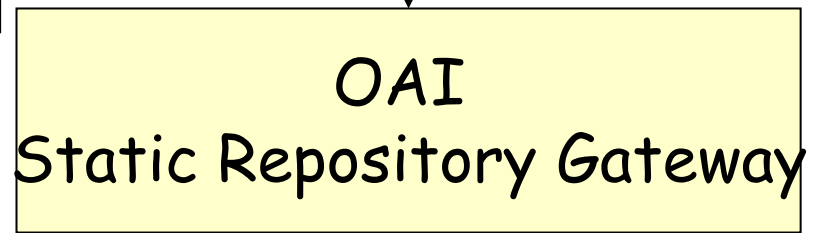
- XML file on web-server
- XSL style sheet to respond to OAI-PMH requests on web-server
- requires:
 - native XSLT support in web server
 - XSL v.2 functionality
 - ⇒ not (yet) low barrier
 - ⇒ but very promising



dependent Data Provider

- XML file on web-server
 - depend on Gateway to respond to OAI-PMH requests
 - requires:
 - *registration* with Gateway
 - Gateway implementation(s)
- => cf. OLAC ViDa





OAI Static Repository

- XML file
- format defined via W3C XML Schema
- uses elements from OAI-PMH Namespace
- sections that contain response to Identify and ListMetadataFormats verbs
- one ListRecords section per Metadata Format
- no support for "sets", "deleted records"



OAI Static Repository

- Data Provider creates/updates:
 - via XML editor
 - via recurrent downloads from database
- demo



OAI Static Repository Gateway

- makes Static Repository harvestable as an autonomous OAI-PMH Repository
- base URL of OAI-PMH Repository is concatenation of:
 - base URL of Gateway
 - HTTP address of Static Repository



OAI Static Repository Gateway

- Static Repository *registers* with Gateway by issuing an Identify against its own OAI-PMH base URL
- Gateway exposes all registered Static Repositories in "friends" container in all Identify responses

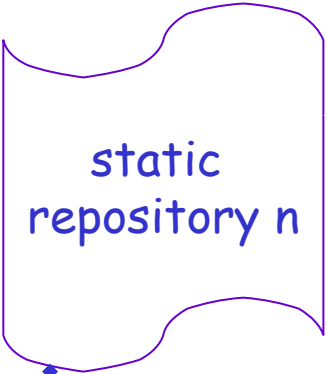
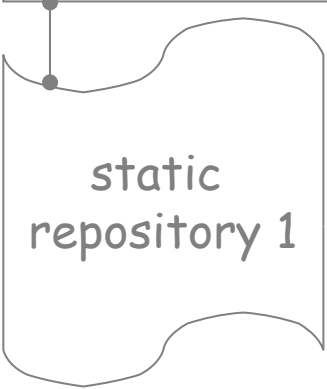


OAI Static Repository Gateway

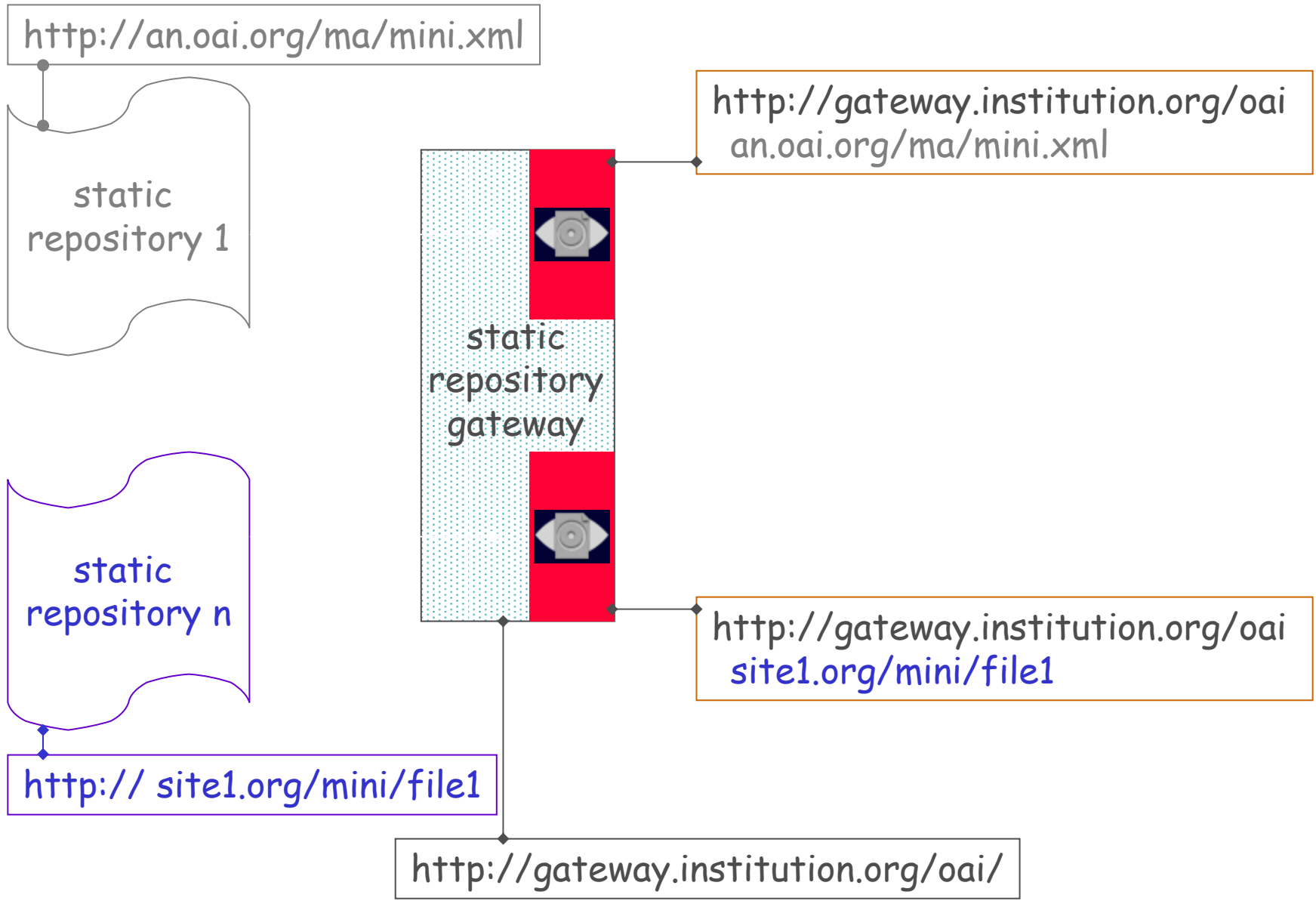
- Strictly defined behavior re guaranteeing accuracy of responses to harvesters:
 - always use most recent version of Static Repository
 - may cache Static Repositories but must perform freshness test before responding

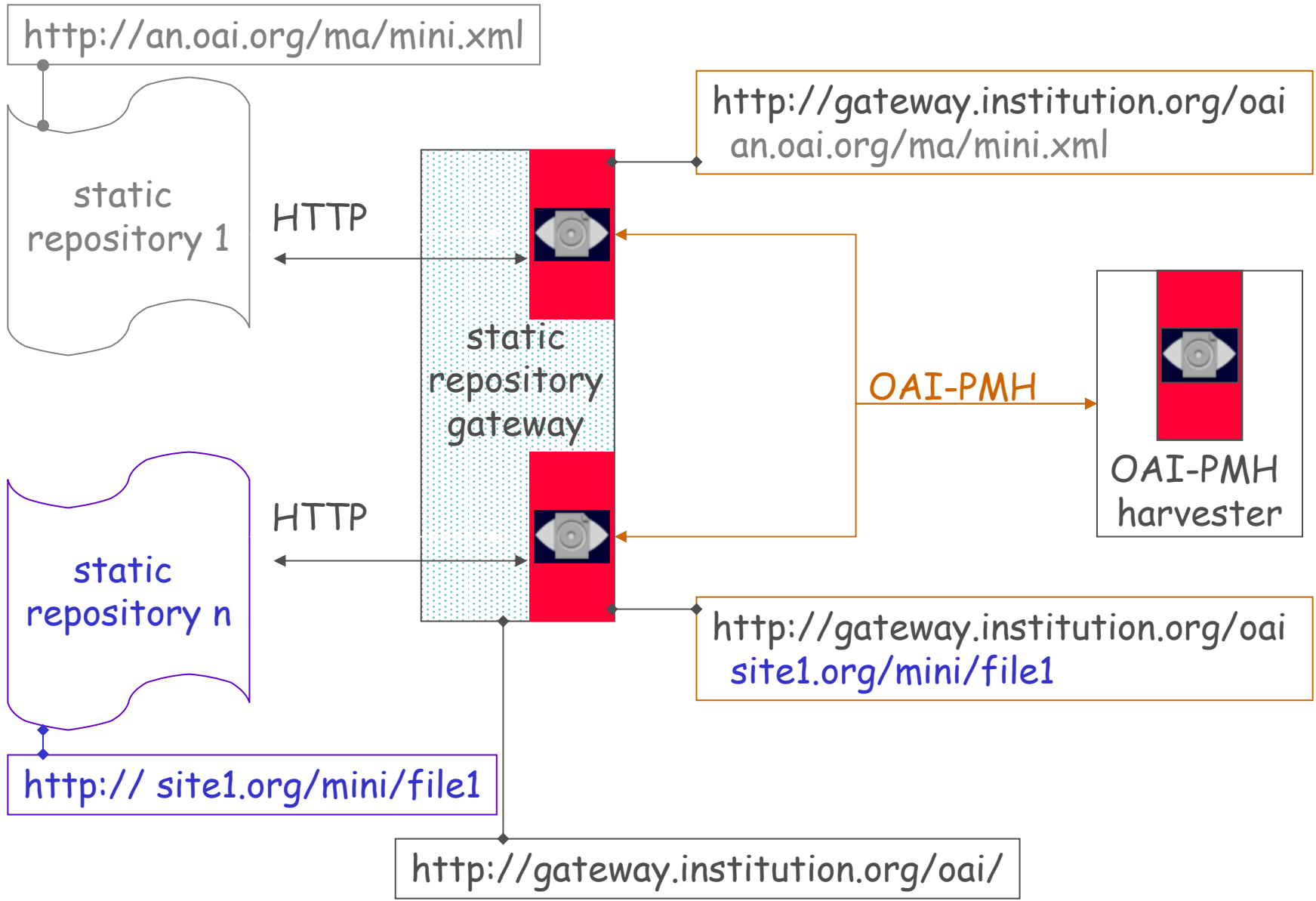


<http://an.oai.org/ma/mini.xml>



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





OAI Static Repository Gateway

- demo



status

- alpha OAI specification for Static Repository and Gateway
- initial feedback received
- ~ August 2003: OAI tech group to move spec to beta and release
- interest?



status

- strong interest from:
 - OLAC
 - NSDL
 - U Illinois Urbana Champaign
 - Belgian Union Catalogue
 - Brazilian Union Catalogue
 - Virginia Tech



status

- LANL Gateway implementation
- tested in several environments, e.g. OLAC
- demo site live
- very soon: release of software



status

- XML editor for Static Repositories
 - OLAC
 - Virginia Tech



URIs

- OAI Static Repository & Gateway spec

<http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>

- LANL demo Gateway

<http://libtest.lanl.gov/cgi-bin/gateway.cgi>

- This presentation

<http://lib-www.lanl.gov/~herbertv/presentations/DLF-200305/>

JCDL 2003 paper

<http://lib-www.lanl.gov/~herbertv/papers/jcdl2003-submitted-draft.pdf>



acknowledgments

- Patrick Hochstenbach, Henri Jerez - LANL, Research Library
- Carl Lagoze, Simeon Warner - Cornell U, CS
- Michael Nelson - Old Dominion U, CS
- Steven Bird, Gary Simmons - OLAC

