# The aDORe effort :  Insights regarding repository federation

Herbert Van de Sompel

Digital Library Research & Prototyping Team

Research Library

Los Alamos National Laboratory


email: herbertv@lanl.gov

URL: http://public.lanl.gov/herbertv

**Los Alamos**
NATIONAL LABORATORY

# Context

- Fact:
  - LANL Research Library stores a significant scholarly collection locally (A&I databases, journal articles, …) and creates applications based on that collection.

- Initial aDORe motivation:
  - Undo tight integration between data and application
  - Uniform approach for ingesting, storing, and disseminating LANL RL data collections

- Bigger picture:
  - Allow for multiple, parallel applications on top of stored content
  - Create an environment that provides guarantees regarding long-term accessibility of stored content

Los Alamos
NATIONAL LABORATORY

# aDORe characteristics

- ## Standards-based:
    - XML, XML Schema, MPEG-21 Digital Item Declaration, the MPEG-21 Digital Item Identification, URI, info URI, OAI-PMH, NISO OpenURL, SRU, Information Environment Service Registry, Internet Archive ARC file format, OAIS concepts.

- ## Component-based, highly modular:
    - Multiple content repositories, Identifier Locator, Service Registry, Format Registry, Semantic Registry, Harvesting front-end, Dissemination front-end

- ## Protocol-based:
    - Modules expose (REST-based) Web services
    - Interaction between modules is protocol-driven

SOME RIGHTS RESERVED

• **Los Alamos**
NATIONAL LABORATORY
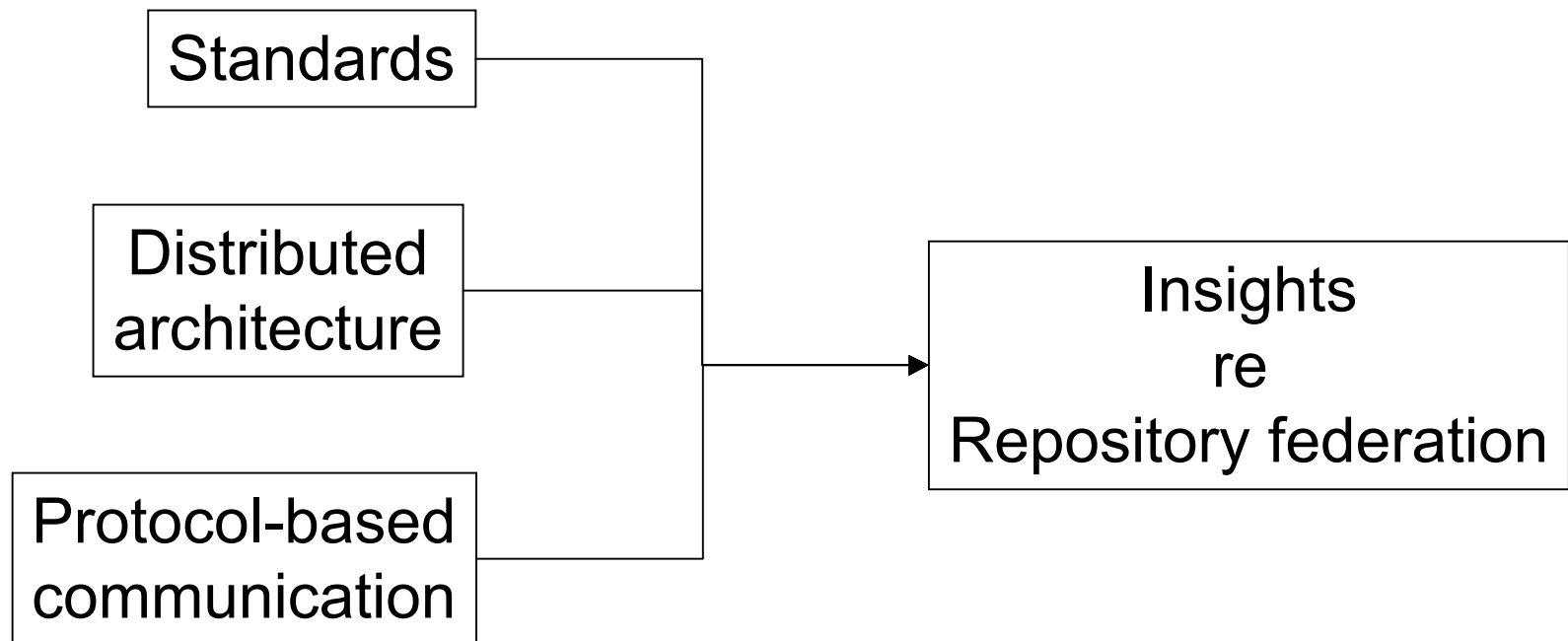
# aDORe effort

aDORe is 2 things:

- A standards-based, repository federation architecture
- Actual implementation of the architecture at LANL for local storage of digital assets (currently finalizing the 2nd version)

aDORe is not a product:

- Components of the aDORe software, usable in other environments, are being released.  aDORe Archive seems attractive for relatively static content collections.
- See http://african.lanl.gov/aDORe/

Los Alamos
NATIONAL LABORATORY

# aDORe effort

```
┌─────────────────┐
│    Standards    │───────────┐
└─────────────────┘           │
                              │
┌─────────────────┐           │        ┌──────────────────────────┐
│   Distributed   │           │        │        Insights          │
│  architecture   │───────────┼───────▶│           re             │
└─────────────────┘           │        │   Repository federation  │
                              │        └──────────────────────────┘
┌─────────────────┐           │
│ Protocol-based  │           │
│  communication  │───────────┘
└─────────────────┘
```

**Digital Library Research & Prototyping Team**
Research Library, Los Alamos National Laboratory
NISO Workshop, September 29th 2006 -Denver, CO

**Los Alamos**
NATIONAL LABORATORY

# How to create an aDORe federation?

**Digital Library Research & Prototyping Team**
Research Library, Los Alamos National Laboratory
NISO Workshop, September 29th 2006 -Denver, CO

**Los Alamos**
NATIONAL LABORATORY

# Prepare the repositories

1.  Implement common representation approach for Digital Objects

    o   Use compound object notion as the starting point; simple objects can be represented as compound objects.

    o   Pick format to represent Digital Objects using XML packages: MPEG-21 DIDL, METS, …

    o   Map internal Digital Object representations used by repositories to the chosen format; this is writing software.

    o   Think carefully about identification and versioning of Digital Objects, datastreams, and XML Packages.

Los Alamos
NATIONAL LABORATORY

# sample Digital Object

| | Type | MIME | identifier |
|---|---|---|---|
| **Digital Object** | scholarly paper | N/A | DOI |
| **Constituent Datastream 1** | metadata record | application/xml | PMID |
| **Constituent Datastream 2** | fulltext file | application/pdf | – |

Los Alamos
NATIONAL LABORATORY

# XML Packages, Digital Objects, datastreams
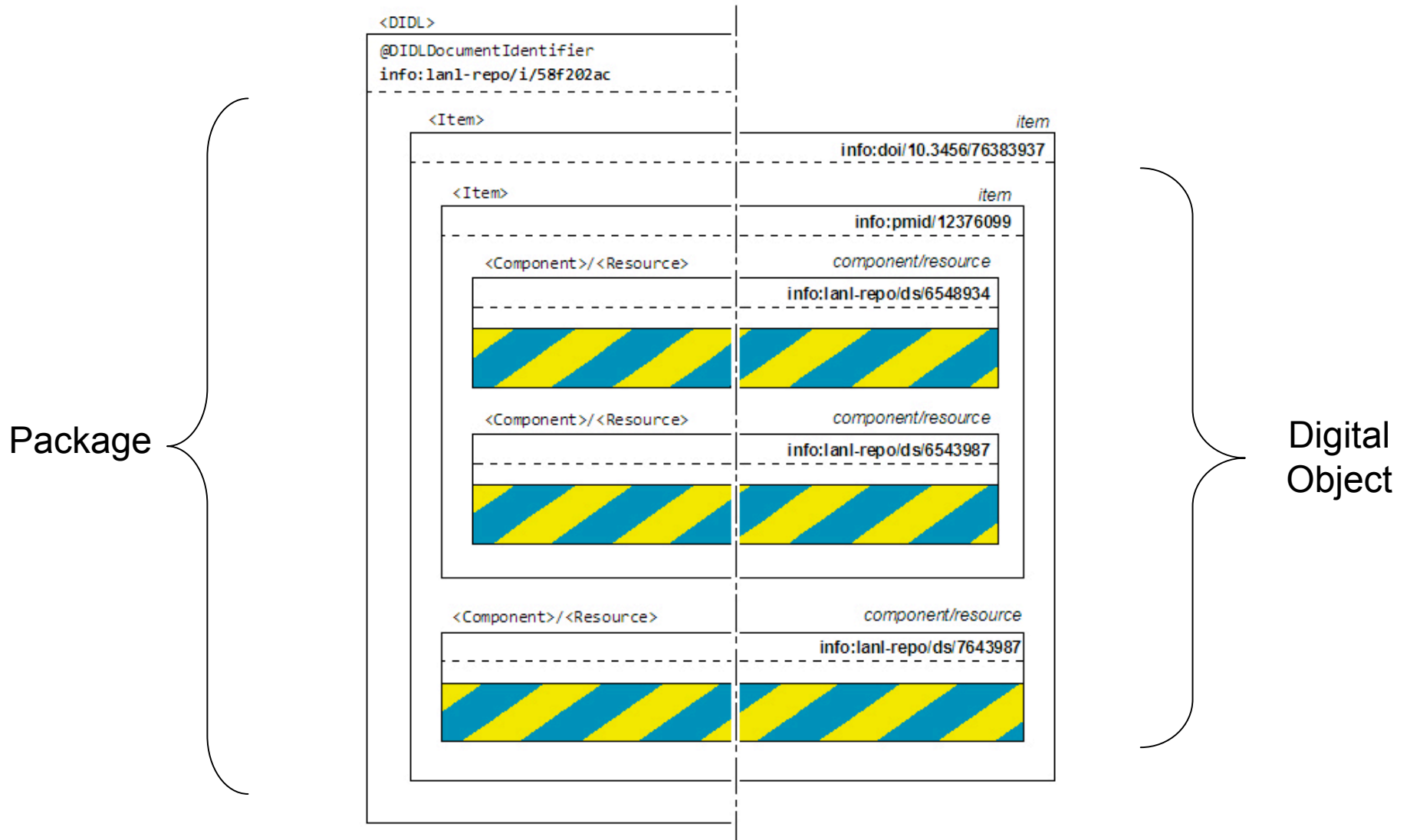
# Identification: digital objects, datastreams, XML packages



Package Identifiers

Digital Object Identifiers

Datastream Identifiers

```
<DIDL>

@DIDLDocumentIdentifier
info:lanl-repo/i/58f202ac

    <Item>                                              item
                                        info:doi/10.3456/76383937

        <Item>                                          item
                                          info:pmid/12376099

            <Component>/<Resource>              component/resource
                                        info:lanl-repo/ds/6548934

            <Component>/<Resource>              component/resource
                                        info:lanl-repo/ds/6543987

        <Component>/<Resource>                  component/resource
                                        info:lanl-repo/ds/7643987
```
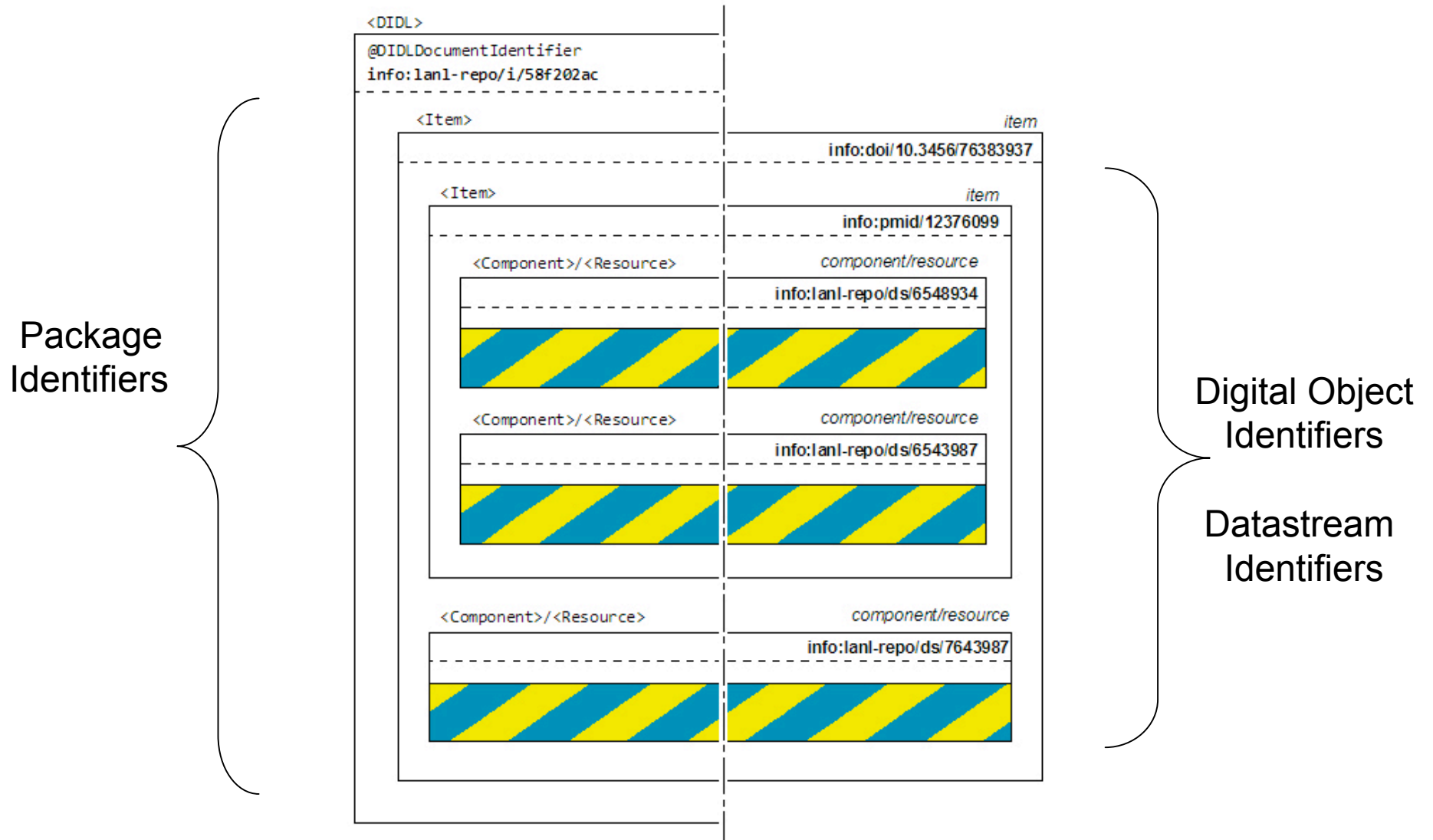
# Prepare the repositories

2. Add common, core service interfaces to the repositories
   o To access XML packages:
      o OAI-PMH for batch harvesting XML packages and identifiers; can use multiple metadata formats.
   o To access services pertaining to digital objects and datastreams:
      o OpenURL Application 1
         o Referent = digital object identifier or datastream identifier
         o ServiceType = various, including GetXMLPackage and GetDatastream
         o Response : various, including the XML Package and the datastream



**OpenURL**

**Los Alamos**
NATIONAL LABORATORY

# Create the federation

3. Introduce Identifier Locator for locating Digital Objects, datastreams, and XML Packages in the federation
    - o  Is populated through OAI-PMH harvesting of identifiers from the repositories in the federation
    - o  OpenURL Application 2 to answer the question: *Given and identifier (XML package, digital object, datastream), which repository/repositories provide(s) access to the identified thing?*
        - o  Referent = XML package, digital object or datastream identifier
        - o  ServiceType = RevealRepositoryIdentifier
        - o  Response: contains repository identifier(s); could be **SRU** response format
    - o  The Identifier Locator must be able to respond to this question irrespective the type of identifier used.  Because every realistic federation **will** deal with different types of identifiers.
    - o  Minimal stored information: (identifier, repository identifier)

**SRU**

**OpenURL**

OPEN ARCHIVES

**Digital Library Research & Prototyping Team**
Research Library, Los Alamos National Laboratory
NISO Workshop, September 29th 2006 -Denver, CO

Los Alamos
NATIONAL LABORATORY

# Create the federation

4. Introduce a Service Registry
   - o Goals:
     - o Keep track of repositories in the federation
     - o Provide information on the location of the service interfaces for each repository
     - o Allow for discovery of repositories based on a variety of metadata
   - o US IESR specification and Ockham software implementation provides an adequate platform
     - o OAI-PMH, SRU, OpenURL access
   - o Could introduce extra registries: format, semantics

**SRU**

**OPEN ARCHIVES**

**OpenURL**

**IESR** *Information Environment Service Registry*

**ockham**

**Los Alamos**
NATIONAL LABORATORY

# Pretend the federation is just a single repository

5.  Introduce service front-ends to the federation
    - o  To access XML packages:
        - o  OAI-PMH Federator for harvesting of batches of XML packages
    - o  To access services pertaining to digital objects and datastreams:
        - o  OpenURL Application 1
            - o  Referent = digital object identifier or datastream identifier
            - o  ServiceType = various, including GetXMLPackage and GetDatastream
            - o  Response : various, including the XML Package and the datastream
    - o  OAI-PMH Federator and OpenURL Resolver will interact with Service Registry, Identifier Locator and repositories.

**OpenURL**

Los Alamos
NATIONAL LABORATORY

# Summary

- These kinds of federations can actually really be built:
  - LANL aDORe implementation
  - Ghent University effort
  - Very different repository systems at the basis
- They can be built by using existing standards and technologies.
- It's nice to be able to benefit from existing technologies and standards
  - Especially if we invested a lot of energy to create them.

**Digital Library Research & Prototyping Team**
Research Library, Los Alamos National Laboratory
NISO Workshop, September 29th 2006 -Denver, CO

Los Alamos
NATIONAL LABORATORY

| Title | [Prieuré Amay] |
|---|---|
| City | Amay(4540) |
| Suburb | Amay(4540) |
| Date | [eind 19e-begin 20e eeuw]. |
| Subject | Abbeys |
| Note | Exterieur |
| Description | 1 fotografische druk : zwart/wit. |
| Collection | Topografische collectie Universiteitsbibliotheek Gent |
| Foto number | info:ugent-repo/topo/226-B-00001 |
| Catalog number | 000979444 |
| Copyright | Reproductierecht Universiteitsbibliotheek UGent |

s M L 🔍 🔎 🛒

[Technical metadata]

[URLs and code for use in websites]

| Thumbnail | http://adore.ugent.be/OpenURL/resolve?rft_id=info:ugent-repo/topo/226-B-C |
|---|---|
| Small | http://adore.ugent.be/OpenURL/resolve?rft_id=info:ugent-repo/topo/226-B-C |
| Medium | http://adore.ugent.be/OpenURL/resolve?rft_id=info:ugent-repo/topo/226-B-C |
| Large | http://adore.ugent.be/OpenURL/resolve?rft_id=info:ugent-repo/topo/226-B-C |
| Info | http://adore.ugent.be/OpenURL/resolve?rft_id=info:ugent-repo/topo/226-B-C |
| DIDL | http://adore.ugent.be/OpenURL/resolve?rft_id=info:ugent-repo/topo/226-B-C |
| Zoomer | <object width="500" height="400"><param name="movie" value="http://ador |

Ghent University aDORe effort led by Patrick Hochstenbach
http://adore.ugent.be : Step 1 and Step 2 completed

**Digital Library Research & Prototyping Team**
Research Library, Los Alamos National Laboratory
NISO Workshop, September 29th 2006 -Denver, CO

Los Alamos
NATIONAL LABORATORY