

# Untitled I challenges ahead

Herbert Van de Sompel  
Research Library  
Los Alamos National Laboratory, USA

# About?

- Original intent: talk about technical work at LANL
- But LANL is **sooooo** different:
  - Local storage of Terrabytes of content
  - Local creation of services over that content
- ⇒ Whatever LANL does, doesn't apply to other libraries

# About?

- In this keynote, I will:
  - Show that many libraries will soon be in a quite similar situation
  - Explore the characteristics and consequences of that situation
  - Focus on fundamental infrastructure
- Structure:
  - Slides that make the major arguments
  - Sidebars that illustrate (related) thoughts

# Sidebar



# A brief history of digital library collections

	Storage		Service	
	Local	Remote	Local	Remote
catalogue	X		X	
A&I	X	X	X	X
full content		X		X

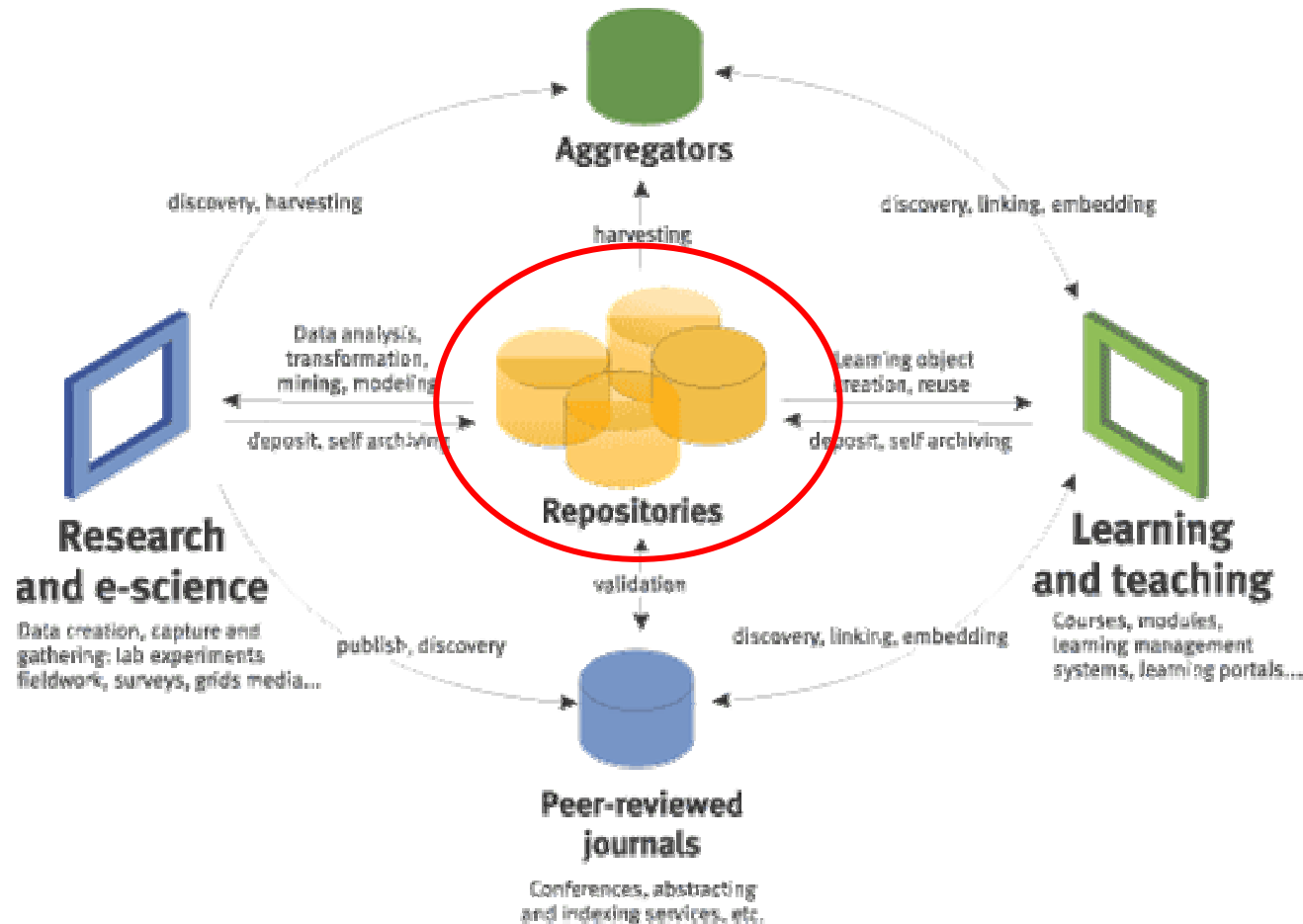
- 2 considerations:
  - Minimal locally hosted collection
  - Storage and Service are tied together
- Both will change

the

repository

model

# The repository model



"Pattern Recognition: The 2003 OCLC Environmental Scan"  
<http://www.oclc.org/membership/escan/toc.htm>

# The repository model

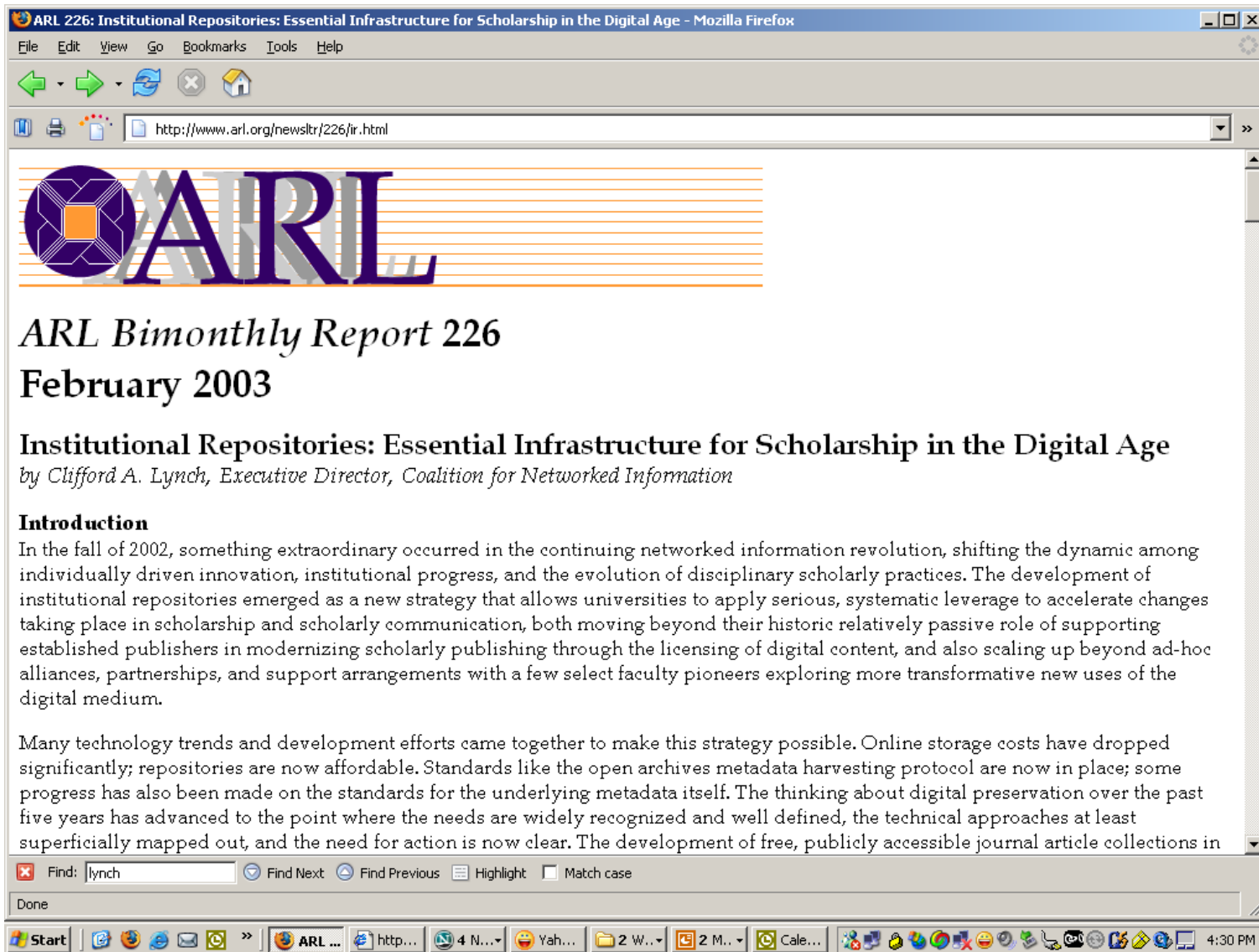


Different repository types:

- scholarly communication (preprint, postprint),
- dataset repositories,
- cultural heritage collections,
- cultural event collections,
- learning object repositories,
- teaching object repositories,
- digitized book repositories,
- ....

Can be institution-based, discipline-based, ...





<http://www.arl.org/newsltr/226/ir.html>



Herbert Van de Sompel  
Olybris 2005, Monday April 18<sup>th</sup>, Kos, Greece



# The repository model



Before they know it,  
institutions will be swamped  
with digital information of  
all kinds

Libraries seem to be the  
natural parties to take care  
of this

Vast growth of digital  
collection:

- Local repository (ies)
- Thousands of remote repositories

# The repository model

Explore (some of) the characteristics & consequences of this model:



- Value chains starting in repositories
- Local capacity
- Archiving
- Rights
- Interoperability
- Standards

# The repository model

Explore (some of) the characteristics & consequences of this model:



- Value chains starting in repositories
  - Local capacity
  - Archiving
  - Rights
  - Interoperability
  - Standards



## Value chains starting in repositories

- New knowledge is really being created when allowing for non-anticipated use of stuff.
- These repositories are *not* about creating *services for local users* (only)
- These repositories are *not* about creating *a service* (user interface) *for all users*
- These repositories are about *facilitating* the *use* of materials *in many contexts*
- *These repositories are the starting point of value chains*

Technorati: Search for <http://weblog.infoworld.com/udell/2005/04/11.html#a1212> - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.technorati.com/cosmos/search.html?rank=&url=http%3A%2F%2Fweblog.infoworld.com%2Fudell%2F2005%2F04%2F11.html#a1212

Getting Started Latest Headlines

Technorati Sign In Help

http://weblog.infoworld.com/udell/ Search  
Keyword or URL

Cosmos Search Results:

<http://weblog.infoworld.com/udell/2005/04/11.html#a1212>  
Make this a Watchlist

3 links from 3 sources Query took 0.017 seconds

Sort by: **Most Recent** | Authority

**Greasemonkey, Meme tracking, and privacy** 19 hours ago  
**meme tracking with Greasemonkey**  
The JavaScript Weblog – javascript.weblogsinc.com \_ 234 links  
from 188 sources

**techno weenie, Apr 12.** 1 day ago  
interface. He also points to a script that adds persistent searches to Gmail ("Search Folders" for you Outlook users). Finally, Jon Udell is **tracking memes**. Right now, he's adding a simple bloglines and delicious count to links, showing how hot they are currently. How about an Old  
techno weenie 10 links from 9 sources

**Flickr without the Flash** 2 days ago  
**In Meme tracking with Greasemonkey**  
Simon Willison's Weblog 1047 links from 808 sources

Ads by Google

**Start your blog now**  
Share information never made easier 5 minutes away from your own blog.  
[www.myownblog.net](http://www.myownblog.net)

**Your Own Blog for Free**  
Keep an online blog including image galleries. All for Free.  
[www.BlogsToGo.com](http://www.BlogsToGo.com)

**InfoWorld Magazine**  
Qualify for your free

Done

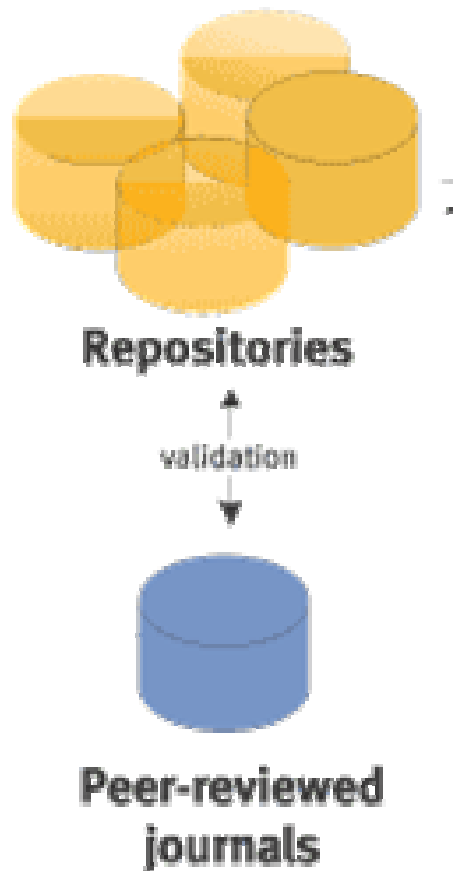
Start Netscape Microsoft Keynote Microsoft Technorati 10:55 AM

- Value chains emerging from RSS feeds

<http://www.technorati.com>

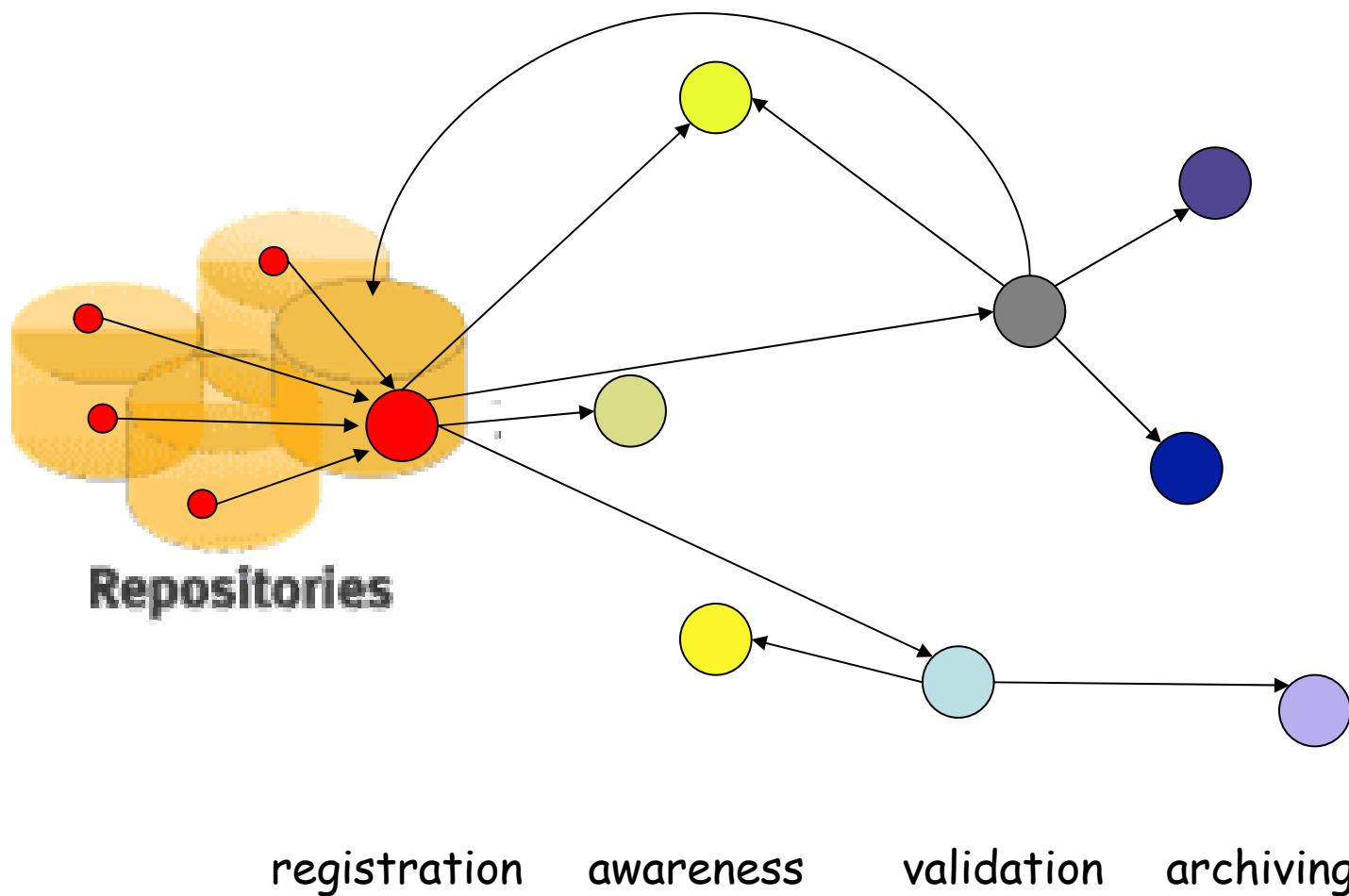


## Example: scholarly communication value chains



- Journal system is just one possible, vertically integrated value chain
- In a networked world, the functions it performs can/will be handled in a deconstructed/distributed manner:
  - registration in repository
  - validation by different nodes/parties
  - archiving by different nodes/parties
  - awareness by different nodes/parties

# Example: scholarly communication value chains





Rethinking Scholarly Communication: Building the System that Scholars Deserve - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html

Getting Started Latest Headlines

Search | Back Issues | Author Index | Title Index | Contents

---

**OPINION**

---

**D-Lib Magazine**  
September 2004

Volume 10 Number 9  
ISSN 1082-9873

**Rethinking Scholarly Communication**

**Building the System that Scholars Deserve**

[Herbert Van de Sompel](#)  
Los Alamos National Laboratory, Research Library  
<herbertv@lanl.gov>

[Sandy Payette](#)  
Cornell University, Computing and Information Science  
<payette@cs.cornell.edu>

[John Erickson](#)  
Hewlett-Packard Laboratories, Digital Media Systems Lab  
<john.erickson@hp.com>

[Carl Lagoze](#)  
Cornell University, Computing and Information Science  
<lagoze@cscornell.edu>

Done

Start | 2 Ne... | Calen... | Keynote | 2 Fir... | Firewo... | Yahoo... | 2:47 PM

<http://dx.doi.org/10.1045/september2004-vandesompel>



Herbert Van de Sompel  
Olybris 2005, Monday April 18<sup>th</sup>, Kos, Greece





# Value chains starting in repositories

- Lesson learned:

To allow value chains to emerge on the basis of materials in repositories, those repositories need a clear/clean machine interface that allows downstream applications to consume materials, aggregate them, build services, ...

- ⇒ Disconnection of repository content and service: allows for creation of both local and remote services
- ⇒ On-Web: Protocol-oriented interface
- ⇒ These value chains are about *the real stuff* not (only) about metadata

Resource Harvesting within the OAI-PMH Framework - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.dlib.org/dlib/december04/vandesompel/12vandesompel.html

Getting Started Latest Headlines

Search | Back Issues | Author Index | Title Index | Contents

**ARTICLES**

**D-Lib Magazine**  
December 2004

Volume 10 Number 12  
ISSN 1082-9873

**Resource Harvesting within the OAI-PMH Framework**

[Herbert Van de Sompel](#)  
Los Alamos National Laboratory, Research Library  
<herbertv@lanl.gov>

[Michael L. Nelson](#)  
Old Dominion University, Computer Science Department  
<mln@cs.odu.edu>

[Carl Lagoze](#)  
Cornell University, Computing and Information Science <lagoze@cs.cornell.edu>

[Simeon Warner](#)  
Cornell University, Computing and Information Science  
<simeon@cs.cornell.edu>

**Abstract**

Done

Start Resource Harvesting ... Microsoft PowerPoint - [k... 10:12 AM

- LANL aDORe
- APS/LANL
- DSpace plugin
- mod\_oai

<http://dx.doi.org/10.1045/december2004-vandesompel>



Herbert Van de Sompel  
Olybris 2005, Monday April 18<sup>th</sup>, Kos, Greece



# The repository model

Explore (some of) the characteristics & consequences of this model:

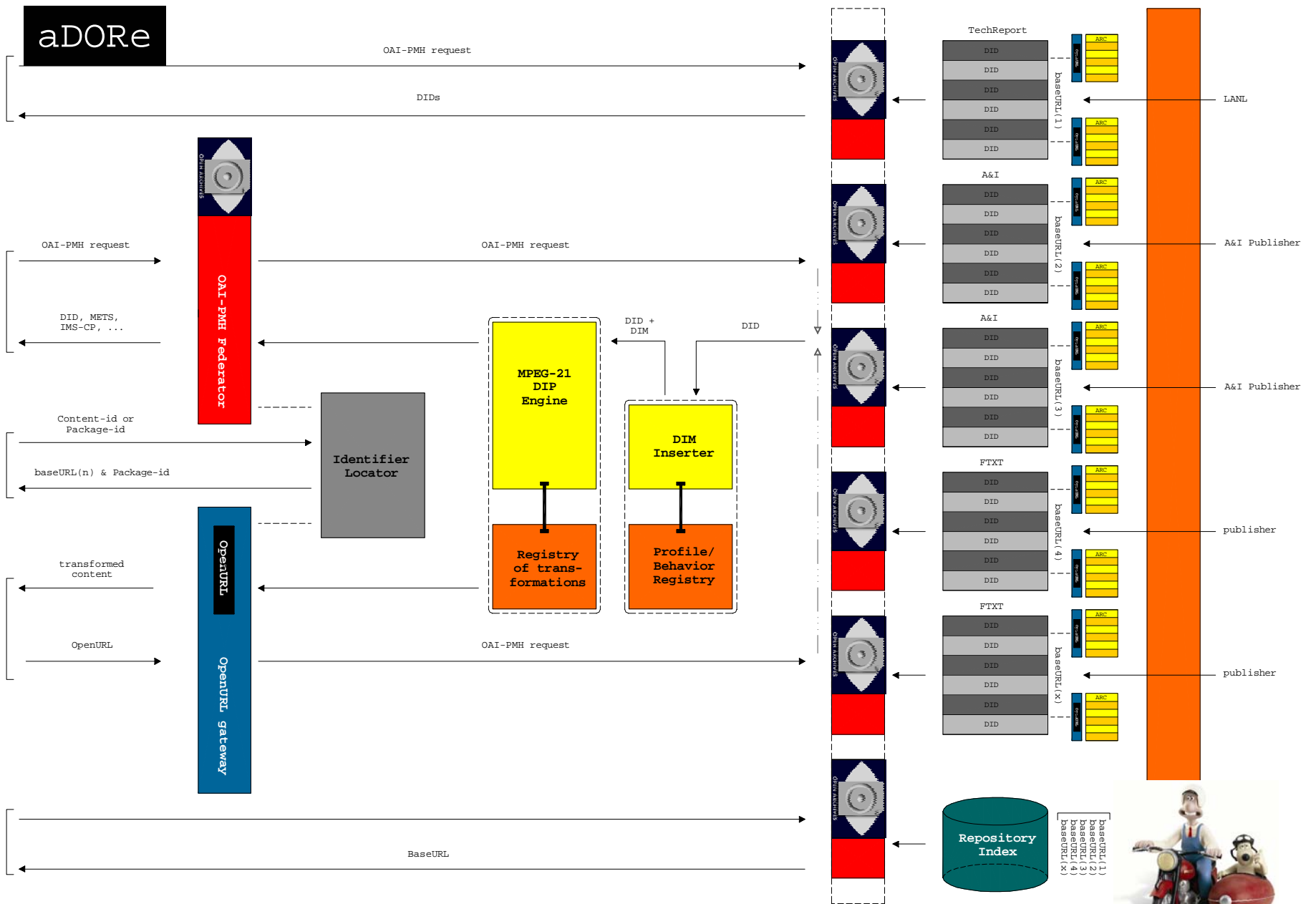


- Value chains starting in repositories
- Local capacity
  - Archiving
  - Rights
  - Interoperability
  - Standards



## Local capacity

- Need basic infrastructure to be able to deal with digital materials of all kinds
- Infrastructure has the real stuff, not metadata at its core
- DSpace, eprints.org, Fedora, ...
  - Doctypes?
  - Vertical application vs basic plumbing?
  - Service-orientation?
  - On-Web?
  - **Multiple repositories?**
  - **Scale?**



[cs/0502028] aDORe: a modular, standards-based Digital Object Repository - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://arxiv.org/abs/cs.DL/0502028

## Computer Science, abstract cs.DL/0502028

From: Herbert van de Sompel [[view email](#)]  
Date: Fri, 4 Feb 2005 22:09:14 GMT (467kb)

### aDORe: a modular, standards-based Digital Object Repository

Authors: [Herbert Van de Sompel](#), [Jeroen Bekaert](#), [Xiaoming Liu](#), [Luda Balakireva](#), [Thorsten Schwander](#)

Comments: Draft of submission to Computer Journal

Subj-class: Digital Libraries

ACM-class: H 3.7

This paper describes the aDORe repository architecture, designed and implemented for ingesting, storing, and accessing a vast collection of Digital Objects at the Research Library of the Los Alamos National Laboratory. The aDORe architecture is highly modular and standards-based. In the architecture, the MPEG-21 Digital Item Declaration Language is used as the XML-based format to represent Digital Objects that can consist of multiple datastreams as Open Archival Information System Archival Information Packages (OAIS AIPs). Through an ingestion process, these OAIS AIPs are stored in a multitude of autonomous repositories. A Repository Index keeps track of the creation and location of all the autonomous repositories, whereas an Identifier Locator registers in which autonomous repository a given Digital Object or OAIS AIP resides. A front-end to the complete environment, the OAI-PMH Federator, is introduced for requesting OAIS Dissemination Information Packages (OAIS DIPs). These OAIS DIPs can be the stored OAIS AIPs themselves, or transformations thereof. This front-end allows OAI-PMH harvesters to recurrently and selectively collect batches of OAIS DIPs from aDORe, and hence to create multiple, parallel services using the collected objects. Another front-end, the OpenURL Resolver, is introduced for requesting OAIS Result Sets. An OAIS Result Set is a dissemination of an individual Digital Object or of its constituent datastreams. Both front-ends make use of an MPEG-21 Digital Item Processing Engine to apply services to OAIS AIPs, Digital Objects, or constituent datastreams that were specified in a dissemination request.

Full-text: [PDF only](#)

Done

Start [cs/0502028]... 3 Microsoft Po... Fireworks - OAI... EAB-2005 10:54 AM

• not a product

<http://arXiv.org/abs/cs.DL/0502028>



Herbert Van de Sompel  
Olybris 2005, Monday April 18<sup>th</sup>, Kos, Greece



# The repository model

Explore (some of) the characteristics & consequences of this model:



- Value chains starting in repositories
- Local capacity
- Archiving
  - Rights
  - Interoperability
  - Standards





# Archiving

- Very early days
- Current strategies:
  - Deal with materials in a way that *supports* their preservation:
    - Be certain of what you store / Record datastream-related metadata
    - Risk detecting tools
  - Mirroring

JHOVE - JSTOR/Harvard Object Validation Environment - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://hul.harvard.edu/jhove/

JHOVE - JSTOR/Harvard Object Validation Environment

# JHOVE

Home | [News](#) | [Tutorial](#) | [Documentation](#) | [Examples](#) | [Distribution](#)

## Format-Specific Digital Object Validation

### 1 Introduction

The concept of representation format, or type, permeates all technical areas of digital repositories. Policy and processing decisions regarding object ingest, storage, access, and preservation are frequently conditioned on a per-format basis. In order to achieve necessary operational efficiencies, repositories need to be able to automate these procedures to the fullest extent possible.

[JSTOR](#) and the [Harvard University Library](#) are collaborating on a project to develop an extensible framework for format validation: JHOVE (pronounced "jove"), the JSTOR/Harvard Object Validation Environment.

JHOVE provides functions to perform format-specific identification, validation, and characterization of digital objects.

- ◆ Format *identification* is the process of determining the format to which a digital object conforms; in other words, it answers the question: "I have a digital object, what format is it?"
- ◆ Format *validation* is the process of determining the level of compliance of a digital object to the specification for its purported format, e.g.: "I have an object purportedly of format  $F$ ; is it?"

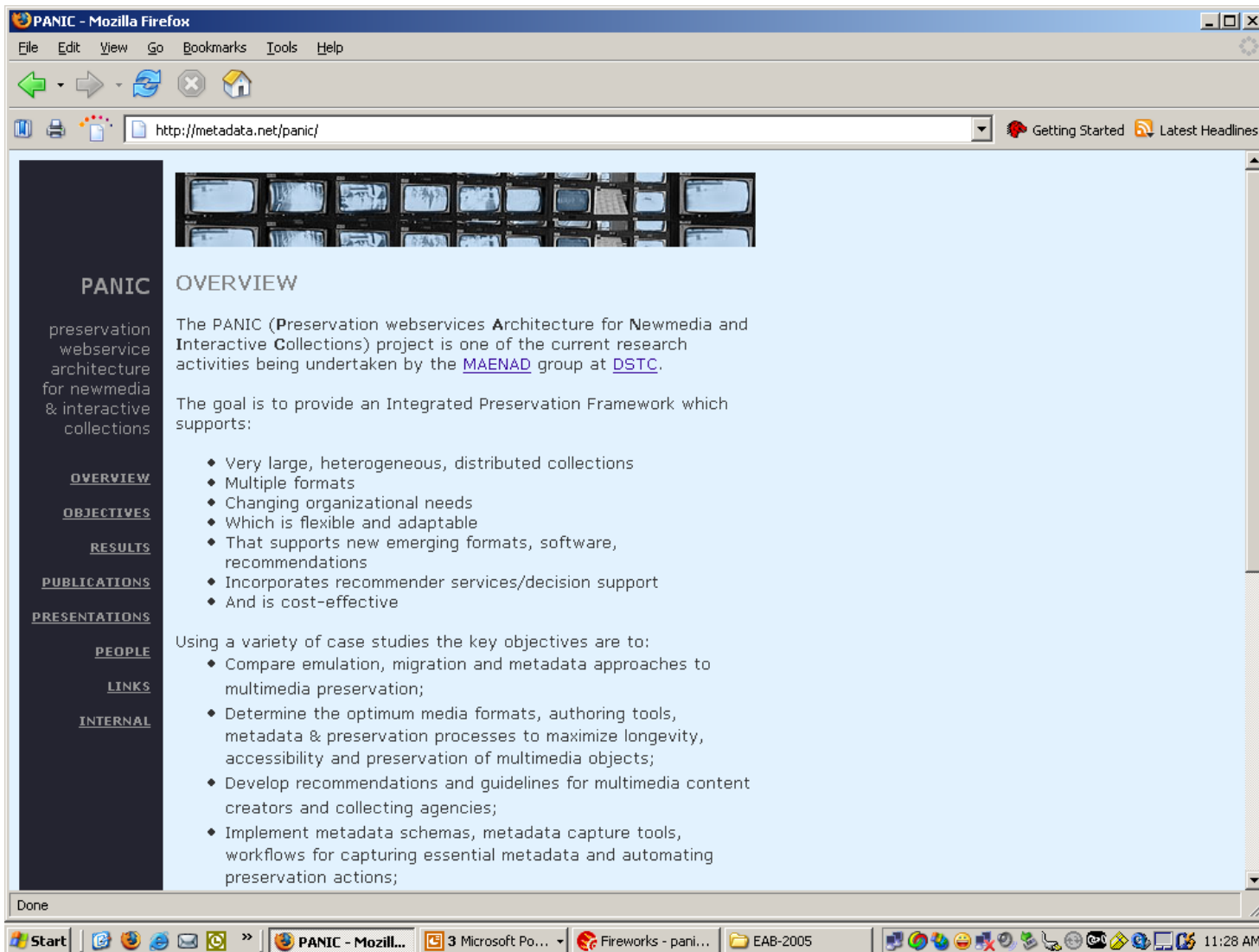
Format validation conformance is determined at two levels: *well-formedness* and *validity*.

Done

Start | JHOVE - JSTOR... | 3 Microsoft Po... | Fireworks - aDO... | EAB-2005 | 11:20 AM

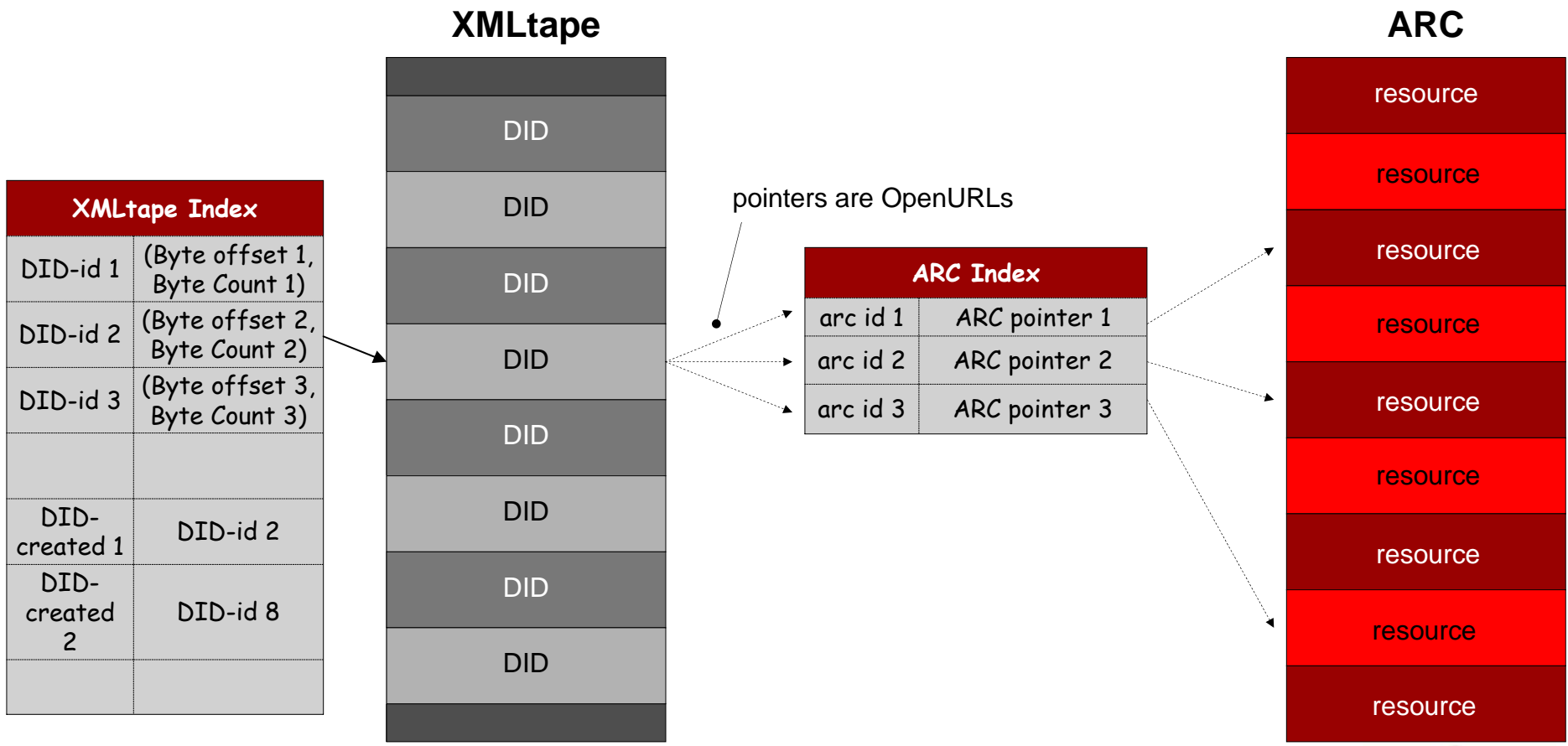
<http://hul.harvard.edu/jhove/>





<http://metadata.net/panic/>





[cs/0503016] File-based storage of Digital Objects and constituent datastreams: XMLtapes and Internet Archive ARC files - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://arxiv.org/abs/cs.DL/0503016

## Computer Science, abstract cs.DL/0503016

From: Xiaoming Liu [[view email](#)]  
Date: Mon, 7 Mar 2005 23:17:23 GMT (46kb)

### File-based storage of Digital Objects and constituent datastreams: XMLtapes and Internet Archive ARC files

Authors: [Xiaoming Liu](#), [Lyudmila Balakireva](#), [Herbert Van de Sompel](#)  
Comments: 12 pages, 1 figures  
Subj-class: Digital Libraries

This paper introduces the write-once/read-many XMLtape/ARC storage approach for Digital Objects and their constituent datastreams. The approach combines two interconnected file-based storage mechanisms that are made accessible in a protocol-based manner. First, XML-based representations of multiple Digital Objects are concatenated into a single, valid XML file named an XMLtape. The creation of indexes for both the identifier and the creation datetime of the XML-based representation of the Digital Objects, facilitates OAI-PMH-based access. Second, ARC files, as introduced by the Internet Archive, are used to contain the constituent datastreams of the Digital Objects in a concatenated manner. An index for the identifier of the datastream facilitates OpenURL-based access. The interconnection between XMLtapes and ARC files is provided by conveying the identifiers of ARC files associated with an XMLtape as administrative information in the XMLtape, and by including OpenURL references to constituent datastreams of a Digital Object in the XML-based representation of that Digital Object.

**Full-text:** [PostScript](#), [PDF](#), or [Other formats](#)

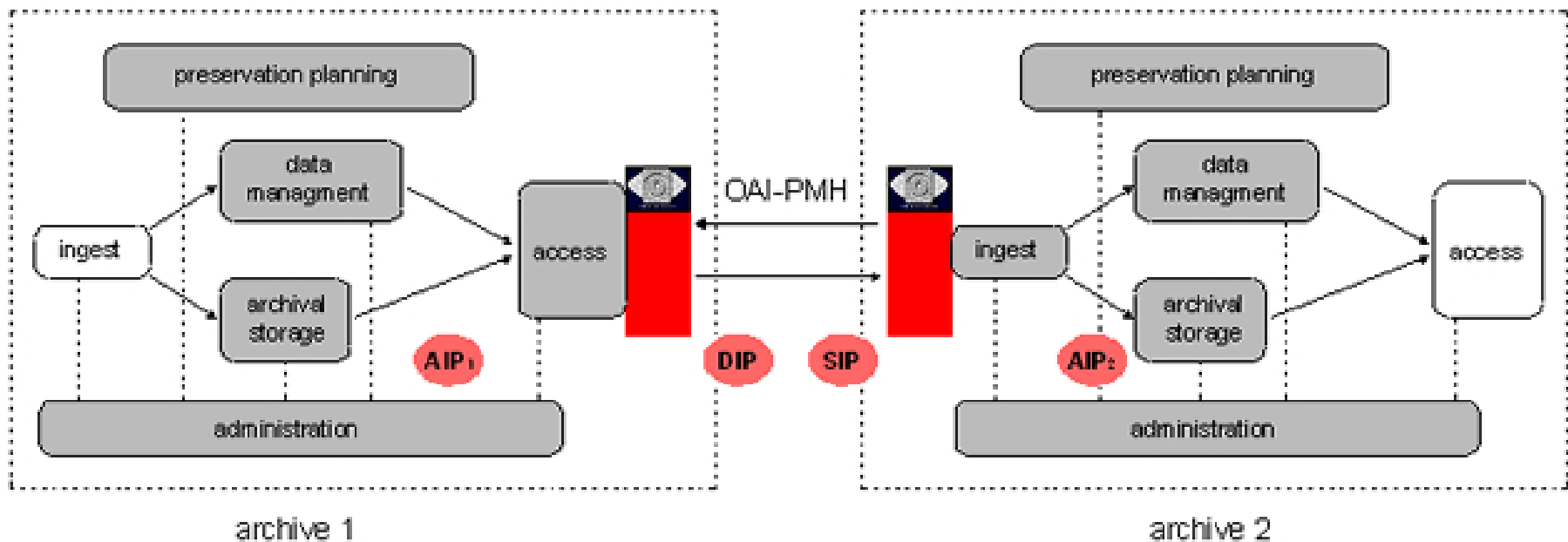
References and citations for this submission:  
[CiteBase](#) (autonomous citation navigation and analysis)

Done

Start [cs/0503016]... 3 Microsoft Po... Fireworks - pani... EAB-2005 11:41 AM

<http://arXiv.org/abs/cs.DL/0503016>





- APS/LANL mirroring:
  - Mirrors objects, not applications, not filesystems
  - Complex object format for XML-based object representation
  - OAI-PMH ~ syncing
  - XML Signatures ~ accuracy of data transfer

Paper in June 2005 D-Lib



# The repository model

Explore (some of) the characteristics & consequences of this model:



- Value chains starting in repositories
- Local capacity
- Archiving
- Rights
  - Interoperability
  - Standards



# Rights

- When facilitating the (re)use of materials (not just metadata) IP concerns increase significantly:
  - Data authenticity
  - Data integrity
  - Usage rights
- Need machine readable rights expressions:
  - Robots are the next generation readers
  - Even when materials are "free"
  - Object-level expressions
  - The world of CC, MPEG-21 REL, ODRL, XRML
  - NISO meeting to explore needs of scholarly community in this realm









# Rights

- Urgent need for an environment in which scholarly assets behave in a manner that matches the “gift exchange” spirit of scholarship.
- James Boyle: Think about what we lose by sticking with the current paradigm!
  - enormous constraints on ability to use scholarly assets: process to extract knowledge, attach knowledge, mine, evolve, build upon: **robots are the next generation readers**



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.

<http://creativecommons.org>



Science Commons | Creative Commons - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://science.creativecommons.org/

Getting Started Latest Headlines

**science commons**

## Welcome to Science Commons

Science Commons is a new project of [Creative Commons](#) and will launch early 2005.

The mission of Science Commons is to encourage scientific innovation by making it easier for scientists, universities, and industries to use literature, data, and other scientific intellectual property and to share their knowledge with others. Science Commons works within current copyright and patent law to promote legal and technical mechanisms that remove barriers to sharing.

### Background

The sciences depend on access to and use of factual data. Powered by developments in electronic storage and computational capability, scientific inquiry is becoming more data-intensive in almost every discipline. Whether the field is meteorology, genomics, medicine, or high-energy physics, research depends on the availability of multiple databases, from multiple public and private sources, and their openness to easy recombination, search and processing.

### The American Traditions

In the United States, this process has traditionally been supported by a series of policies, laws, and practices that were largely invisible even to those who worked in the sciences themselves.

First, American intellectual property law (and, until recently, the law of most developed countries) did not allow for intellectual property protection of "raw facts." One could patent the mousetrap, not the data on the behavior of mice, or the tensile strength of steel. A scientific article could be copyrighted. The data on which it rested could not be. Commercial proprietary

### Publish

license your book, essays, and papers

Creative Commons helps you publish your text online while letting others know exactly what they can and can't do with your work.

**Do you have your own web site?**

You can choose a Creative Commons License to apply to your text works:

**CC Standard License**

License your text under your terms. Our set of standard licenses will let you share text with viewers while protecting your writing from limits you put in place.

- [Guide to adding Creative Commons to your existing website](#)

### Find

Search for science-related materials

Done

Start Science Comm... 3 Microsoft Po... Fireworks - resc... EAB-2005 1:08 PM

<http://science.creativecommons.org>



Herbert Van de Sompel  
Olybris 2005, Monday April 18<sup>th</sup>, Kos, Greece




OAI-PMH Implementation Guidelines - Conveying rights expressions about metadata in the OAI-PMH framework - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.openarchives.org/OAI/2.0/guidelines-rights.htm

## Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting



### Conveying rights expressions about metadata in the OAI-PMH framework

Protocol Version 2.0 of 2002-06-14  
Document Version 2004/11/05T15:52:00Z  
<http://www.openarchives.org/OAI/2.0/guidelines-rights.htm>

*Beta release, specification subject to change*

*The XML encodings of Creative Commons licenses shown in this document may change, pending the Creative Commons' plans to define XML Schema for these expressions.*

*Comments should be sent to the [oai-implementers list](#).*

**Editors**

[Carl Lagoze](#) (OAI Executive; [Cornell University - Computer Science](#))  
[Herbert Van de Sompel](#) (OAI Executive; [Los Alamos National Laboratory - Research Library](#))  
[Michael Nelson](#) (Old Dominion University - Computer Science)  
[Simeon Warner](#) (Cornell University - Computer Science)

This document is one part of the [Implementation Guidelines](#) that accompany the [Open Archives Initiative Protocol for Metadata Harvesting](#)

Done

Start | OAI-PMH Impl... | 3 Microsoft Po... | Fireworks - scien... | EAB-2005 | 1:11 PM

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



# The repository model

Explore (some of) the characteristics & consequences of this model:



- Value chains starting in repositories
- Local capacity
- Archiving
- Rights
- Interoperability
  - Standards



# Interoperability

- Use and re-use of materials in global context
  - Clean/clear machine interface is not enough.
  - Need cross-repository content-level interoperability
  - Interoperable, global federation of repositories



# Interoperability

- Architectural issues include:
  - Object representation (MPEG-21 DIDL, IMS/CP, METS, .)
  - Object identification
  - Object harvesting
  - Object disseminations
  - Object relationships
  - Discovery of object repositories
  - ...

**CORDRA™** Content Object Repository Discovery and Registration/Resolution Architecture

**CORDRA™**

**CORDRA** (Content Object Repository Discovery and Registration/Resolution Architecture):  
An open, standards-based model for how to design and implement software systems for the purposes of discovery, sharing and reuse of learning content through the establishment of interoperable federations of learning content repositories.

CORDRA is designed to be an enabling model to bridge the worlds of learning content management and delivery, and content repositories and digital libraries. CORDRA aims to identify and specify (not develop) appropriate technologies and existing interoperability standards that can be combined into a reference model used to enable a learning content infrastructure.

CORDRA is:

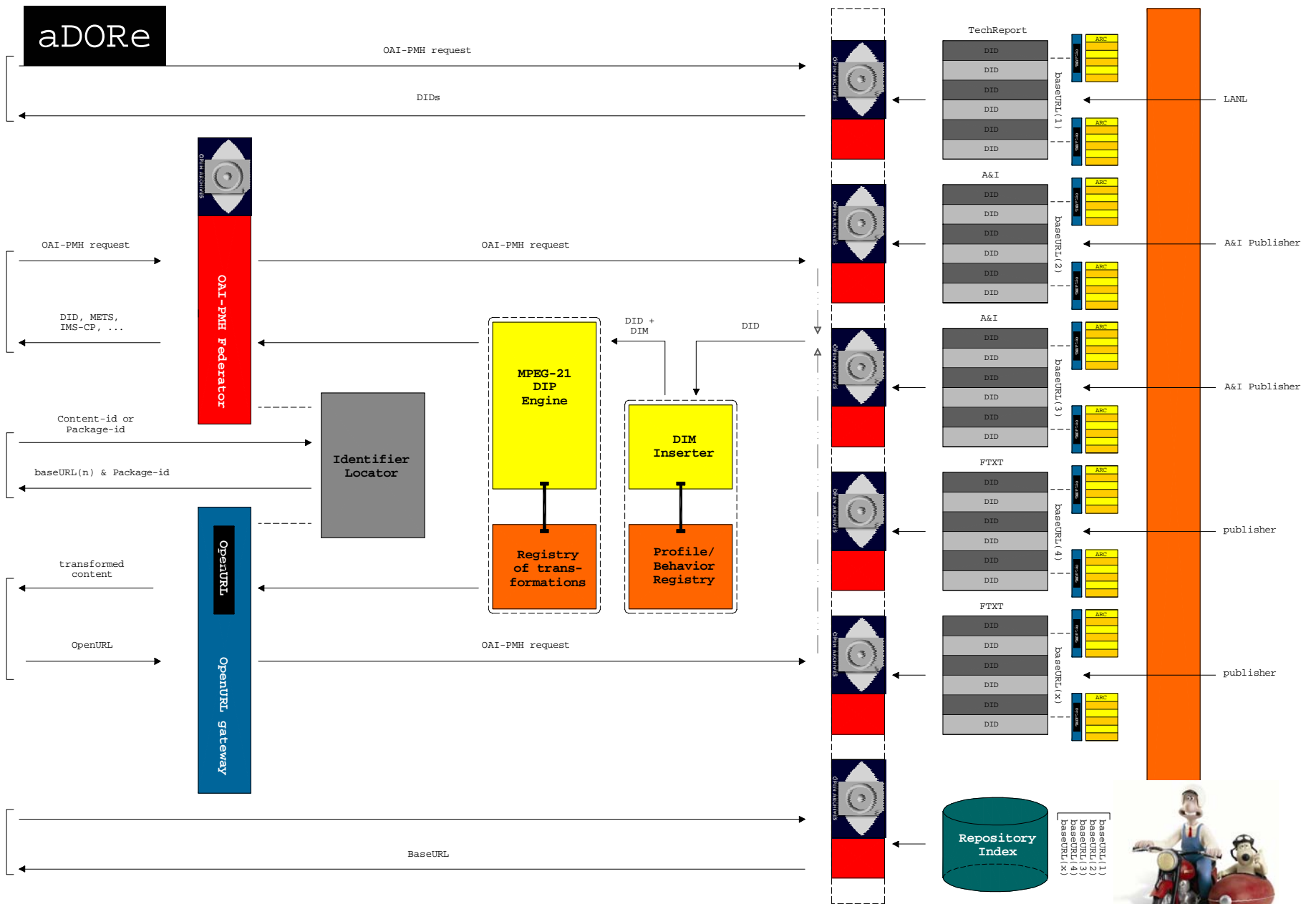
- A formal model that can be used to design federations of repositories (the CORDRA reference model).
- A collection of operational systems built from the CORDRA model, including:
  - a prototype implementation of a repository federation; and
  - an operational *federation of federations* used to combined different CORDRA federations.
- The activities and projects surrounding the definition of the CORDRA model and creation of the operational systems.

The CORDRA activities are being coordinated by the [Advanced Distributed Learning Initiative](#), (ADL) the [Corporation for National Research Initiatives](#), (CNRI) and the [Learning Systems Architecture Lab](#) (LSAL).

<http://cordra.net>







# The repository model

Explore (some of) the characteristics & consequences of this model:



- Value chains starting in repositories
- Local capacity
- Archiving
- Rights
- Interoperability
- Standards



# Standards

- Standards are the glue that holds the networked information environment together.
- Standards are crucial to facilitate the emergence of improved and integrated services across repositories.
- As the information environment becomes more complex, and as we move towards new levels of services, we will need more, not less standards.



# Standards

- Standardization efforts/bodies in our community are seriously challenged:
  - Many standards defined outside our community.
  - Lack of impact on major standardization bodies of the networked world (W3C, IETF, IANA, ...)

Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://info-uri.info/registry/OAIHandler?verb=ListRecords&metadataPrefix=oai\_dc

## "info" URI Scheme

[About "info"](#) | 
 [Internet-Draft](#) | 
 [Registry ID](#) | 
 [Registry Entries](#) | 
 [Registry Policy](#) | 
 [Register Namespace](#) | 
 [Feedback](#)

**RSS**

### Registry Entries

<a href="#">info:bibcode/</a>	Namespace of Astrophysics Data System bibcodes
<a href="#">info:ddbj-embl-genbank/</a>	Namespace of identifiers for sequence records in DDBJ/EMBL/GenBank
<a href="#">info:doi/</a>	Namespace of Digital Object Identifiers
<a href="#">info:fedora/</a>	Namespace of Fedora Digital Objects and Disseminations
<a href="#">info:lanl-repo/</a>	Namespace of identifiers used in the Repository of the LANL Research Library
<a href="#">info:lccn/</a>	Namespace of Library of Congress Control Numbers
<a href="#">info:netref/</a>	Namespace of NISO Standard for Network Reference Services
<a href="#">info:nla/</a>	Namespace of identifiers for the National Library of Australia's digital & digitised collections
<a href="#">info:oclcnum/</a>	Namespace of OCLC Worldcat Control Numbers
<a href="#">info:ofi/</a>	Namespace of Registry Identifiers used by the NISO OpenURL Framework Registry

Done

Start | 3 N... | Cale... | Key... | 2 F... | Fire... | Yah... | 2 M... | 3:01 PM

http://info-uri.info



## Standards/Interoperability context

- Standardization efforts/bodies in our community are seriously challenged:
  - Many standards defined outside our community.
  - Lack of impact on major standardization bodies of the networked world (W3C, IETF, IANA, ...)
  - Problems to interconnect within and amongst related efforts in our community: digital library, grid computing, e-learning, library automation, ...
  - Operational models/processes not adequately adapted to the realities of the networked world (cf. patent challenges OpenURL, MetaSearch)
  - Funding for standardization efforts and related infrastructure is very hard to find (cf. OAI, CIMI, info URI Registry, OpenURL Registry, ...)

Dr. William E. Moen: Luminary Lectures @ Your Library - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.loc.gov/rr/program/lectures/moen.html>

Google bill moen standards Search Web 6 blocked AutoFill Options bil Links Net Snippets Add Selection



The Library of Congress >> Virtual Programs & Services >> Luminary Lectures

**Luminary Lectures**  
@ your library™

**Dr. William E. Moen**

**No Longer Under Our Control: The Nature and Role of Standards in the 21st Century Library**

Wednesday, December 3rd, 10:30-12:00 noon  
@ the West Dining Room (James Madison Building, [Sixth Floor](#))

**ABOUT THE CYBERCAST:**

[View Cybercast](#) (1 hour, 28 minutes - requires Real Player to view).

No registration or password required to access this cybercast lecture. The video of the lecture will be presented in RealPlayer format. To view it, you must have the Real Player installed and at least a 28 K-bps (kilobits per second) Internet connection for your computer. The RealPlayer software may be downloaded, free of charge, from the [RealNetwork Web site](#).

**ABOUT THE LECTURE:**

A MS Word version of the [lecture script](#) is available.

Libraries have prospered because they developed and adopted of a wide range of standards, guidelines, and community practices to serve their needs. Broadly speaking, standards reflect vital community agreements on problems and their solution. The networked environment, however, is changing the context and use of standards in libraries. Librarians no longer create and define the terms, concepts, standards, and technologies that drive library services and practices.

Start VALA2004 CSTB Public... Inbox for h... Dr. Willia... C:\herbert... VALA - Not... 3:03 PM

<http://www.loc.gov/rr/program/lectures/moen.html>



Herbert Van de Sompel  
Olybris 2005, Monday April 18<sup>th</sup>, Kos, Greece



NSF Post Digital Library Futures Workshop - Papers / Roadblocks - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.sis.pitt.edu/~dlwshop/paper\_sompel.html

## Wave Of The Future: NSF Post Digital Library Futures Workshop

June 15 - 17, 2003  
Weguassett Inn, Cape Cod  
Chatham, Massachusetts

**NSF/JISC WORKSHOP**

**GENERAL**

- Welcome
- Background
- Agenda
- References
- Important Dates
- Participants List

**OUTREACH**

- China 2004

**FOR CONTRIBUTORS**

- Call For Papers
- Papers
- Breakout Reports
- Final Report
- Opening Plenary Session
- Supplementary Contributions

**FOR PARTICIPANTS**

**Papers**

### Roadblocks

Herbert Van de Sompel, Los Alamos National Laboratory, Research Library  
*Many thanks to Rick Luce for sharing his insights*  
Download: [PDF Version](#) [WORD Version](#)

**Introduction**

I have no problems with admitting that thinking about a 10-year timeframe poses significant challenges for me. Going through the simple thought experiment of moving myself back in time to 1993, and using that perspective to try and figure which components of the current information infrastructure I would have been able to predict, or which level of global penetration each of those would be able to achieve, leave me rather speechless. Yes, after having seen an early glimpse of the Web through the lens of a Mosaic browser made me understand intuitively that this was going to be important. Real important, as opposed to how, for instance, Gopher was "interesting". But predicting that the Web along with then unknown related technologies would transform our world and world-view in general, as well as everything we did in libraries both traditional, digital and hybrid specifically was way beyond my capabilities as a futurist. Therefore, I have chosen not to directly address the future DL research agenda, but rather, I have concentrated on trying to identify roadblocks in the DL arena that I feel need to be abolished in order to pave the way for the vision expressed in the Cyber infrastructure Report. The items I identify are situated at the level of basic plumbing - things that need to be in place so that other great things can happen. Therefore, they may not come across as all that exciting or innovative, however, I feel that failing to address them may seriously impede efforts to move towards an integrated scholarly knowledge environment, as envisioned in the Cyber infrastructure Report. Most issues I identify pose enormous challenges, but they are not merely technical research challenges. Rather, they cover a broad range of research and organizational challenges at both the technical, legal,

Transferring data from www.sis.pitt.edu...

Start | 3 N... | Cale... | Key... | 2 F... | Fire... | Yah... | Micr... | 2:54 PM

[http://www.sis.pitt.edu/~dlwshop/paper\\_sompel.html](http://www.sis.pitt.edu/~dlwshop/paper_sompel.html)





A9.com > OpenSearch > OpenSearch Query Syntax 1.0 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://opensearch.a9.com/spec/opensearchquerysyntax/1.0/

OpenSearch™ Docs

- OpenSearch Home
- OpenSearch RSS Spec
- OpenSearch Description Document Spec
- OpenSearch Query Spec
- FAQs
- How To's
- See Submitted Columns
- Submit Your Search

Home > OpenSearch > OpenSearch Query Syntax 1.0

### Overview

The OpenSearch™ Query Syntax 1.0 consists of the the minimal set of arguments that are needed to generate search results. The OpenSearch Query Syntax is used in an OpenSearch Description document to describe a search URL to a search aggregator (such as A9.com).

A search URL is defined by substituting the macro expansion tags in place of particular arguments.

All tags are optional, and some tags, such as `{startIndex}` and `{startPage}` are usually mutually exclusive, depending on the search content provider.

Search aggregators should make no assumptions that the particular arguments are honored by the search content provider, though the search content provider should make every effort to do so if possible.

### Search Arguments

Term	Definition	Comments
<code>{searchTerms}</code>	The terms to be searched for	URL-encoded search terms
<code>{count}</code>	The desired number of results.	A positive integer
<code>{startIndex}</code>	The offset, in individual search results, of the first search result	A positive (non-zero) integer
<code>{startPage}</code>	The offset, in pages of search results, of the first search result	A positive (non-zero) integer

### Examples

The following is a valid search URL for <http://koders.com>:

```
http://www.koders.com/?q={searchTerms}&p={startPage}&sort=qw&w=
```

The following is a valid search URL for <http://indeed.com>:

```
http://beta.indeed.com/opensearch?o={searchTerms}&start={startIndex}&limit={count}
```

The following is a valid search URL for Unto.net AWS Electronics:

```
http://www.unto.net/aws?o={searchTerms}&start={startIndex}&limit={count}&language={startPage}
```

### Results

Search results are expected to be returned in the OpenSearch RSS format. For search content providers that do not offer search results in OpenSearch RSS format, a wrapper must be authored that can perform the transformation.

Done

Start | A9.com... | Microsoft... | Calenda... | 3 Nets... | VALA2004 | 2 Micr... | 11:54 AM

• there is something about simplicity

<http://opensearch.a9.com/spec/opensearchquerysyntax/1.0/>



http://opensearch.a9.com/spec/opensearchdescription/1.0/

Getting Started Latest Headlines

**OpenSearch™ Docs**

- OpenSearch Home
- OpenSearch RSS Spec
- OpenSearch Description Document Spec
- OpenSearch Query Spec
- FAQs
- How To's
- See Submitted Columns
- Submit Your Search

Home > OpenSearch > OpenSearch Description Document 1.0

## Overview

Search engines that export OpenSearch RSS should identify their content and search syntax via an OpenSearch Description document.

An example OpenSearch™ Description document:

```
<?xml version="1.0" encoding="UTF-8"?>
<OpenSearchDescription xmlns="http://a9.com/~spec/opensearchdescription/1.0/">
  <Url>http://www.unto.net/aws?q=(searchTerms)&amp;searchindex=Electronics
    &amp;flavor=osrss&amp;itempage=(startPage)</Url>
  <Format>http://a9.com/~spec/opensearchrss/1.0/</Format>
  <ShortName>Electronics</ShortName>
  <LongName>Amazon Electronics</LongName>
  <Description>Search for electronics on Amazon.com.</Description>
  <Tags>amazon electronics</Tags>
  <Image>http://www.unto.net/search/amazon_electronics.gif</Image>
  <SampleSearch>ipod</SampleSearch>
  <Developer>DeWitt Clinton</Developer>
  <Contact>dewitt@unto.net</Contact>
  <Attribution>Product and search data &amp;copy; 2005, Amazon, Inc.,
    All Rights Reserved</Attribution>
  <SyndicationRight>open</SyndicationRight>
  <AdultContent>false</AdultContent>
</OpenSearchDescription>
```

Note: This document should be made available via HTTP GET at a publicly accessible URL.  
 Note: OpenSearch Description documents can not contain HTML.  
 Note: The Url and Attribution tags above are shown with line breaks for formatting purposes only.

Start | A9.co... | Micro... | Calen... | 3 N... | VALA... | 2 Mi... | Firew... | 11:59 AM

• there is another page: more complex than you thought!

http://opensearch.a9.com/spec/opensearchdescription/1.0/



# The repository model

Explore (some of) the characteristics & consequences of this model:



- Value chains starting in repositories
- Local capacity
- Archiving
- Rights
- Interoperability
- Standards

so

what?

can we conclude

# The future of digital library collections?

	Storage		Service	
	Local	Remote	Local	Remote
catalogue	X		X	
A&I	X	X	X	X
full content		X		X
repositories	X	X	X	X

- Important locally hosted collection
- Storage and Service disconnected
- Important challenges

# A content-node & service-node ecology?

- Content nodes:
  - Libraries become content-nodes, capturing the intellectual output of their parent institutions and “exposing” it.
  - Vision: A network of federated repositories that makes available the collective intellectual output of faculty and researchers of the world's research institutions
  - Ongoing with the Institutional Repository movement
  - Libraries must act in this realm

# A content-node & service-node ecology?

- Service nodes:
  - Need services (value chains) to emerge on top of that content
  - "If the content is on-Web, the services will bloom"
  - Can not solely rely on ... euh .. Google Scholar
  - Service node tasks include:
    - indexing, searching, recommendation, linking, data-mining, visualization, ... nodes
    - annotation, certification, metric-collecting, rewarding, ... nodes
    - archiving, normalization/transformation, ... nodes
  - Vision: A federation of networked services - in which Libraries take on specific service tasks - that turns into a global scholarly value chain

# The repository model

## Physical libraries:

- Local storage of content originating with 3<sup>rd</sup> parties
- Facilitate use of that content by local user base

## Current libraries:

- Remote storage of content originating with 3<sup>rd</sup> parties
- Facilitate use of that content by local user base

## Repository model libraries:

- Local storage of content that originates in-house
- Facilitating its use by remote and local users **by facilitating the emergence of services**

Emergence of a quite fundamental new library model



but really, dude,

how?

let's call upon



**RESCUE  
HEROES**

© 2003