



# Organising the Digital Library

Donatella Castelli  
CNR-ISTI

Dean Krafft  
Cornell University

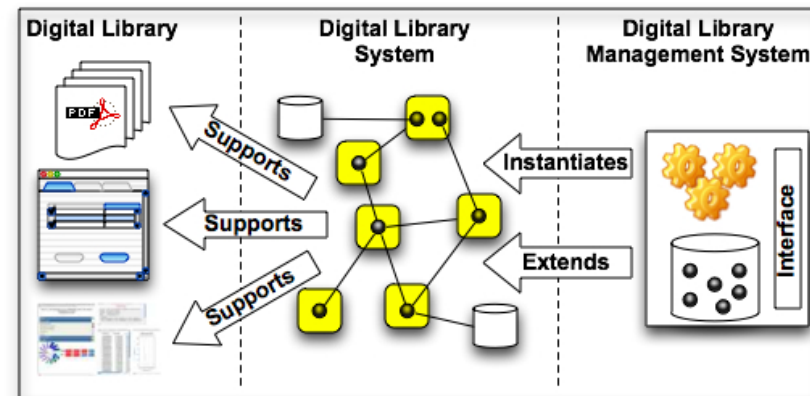
- Introductory concepts:
  - DL “systems”
  - DL Users
  - DL system from a functional point of view
- NSDL mediation & aggregation
  - Union Catalog
  - Moving beyond the union catalog
  - Semantic Digital and Collaborative DLs
- DLMSs and infrastructures
- Demos

# PART I

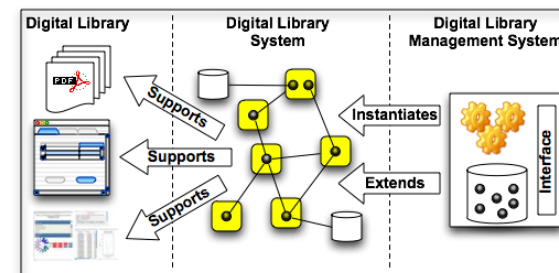
## Introductory Concepts

# Introductory concepts: DL “systems”

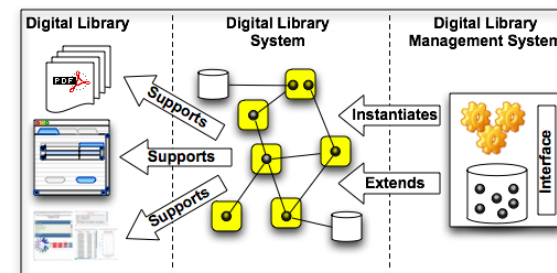
- Digital Library
- Digital Library System
- Digital Library Management System



*A (potentially virtual) organization that comprehensively collects, manages, and preserves for the long term rich **digital content** and offers to its **user** communities specialized **functionality** on that content, of measurable **quality**, and according to prescribed **policies**.*

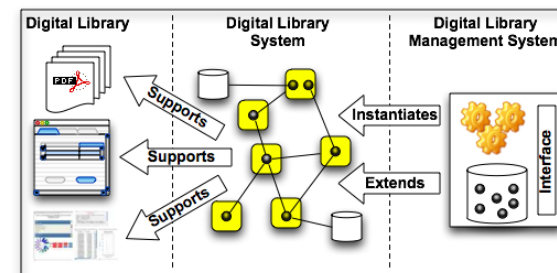


*A software system that is based on a (potentially distributed) **architecture** and provides all functionality that is required by a particular Digital Library. Users interact with a Digital Library through the corresponding Digital Library System.*



# Digital Library Management System

*A generic software system that provides the appropriate software infrastructure to both (i) produce and administer a Digital Library System that incorporates all functionality that is considered foundational for Digital Libraries and (ii) integrate additional software offering more refined, specialized, or advanced functionality.*



# Introductory concepts: DL actors

- End-user
- DL designer
- DL system administrator
- DL application developer



**DL  
End-Users**



**DL System  
Administrators**



**DL Designers**



**DL Application  
Developers**





- Exploit the DL functionality for providing, consuming, and managing the DL Content as well as some of its other constituents. They perceive the DL as a stateful<sup>(\*)</sup> entity that serves their functional needs. DL end-users may be partitioned into:
  1. *Content Creator*
  2. *Content Consumer*
  3. *Librarian*

(\*)The state of the DL corresponds to the state of its resources, i.e., it consists of the collections of information objects managed by the DL, its set of authorized users, its functionality, and its set of policies. This state changes during the Digital Library lifetime according to the functionality activated by the users and their inputs.



- Exploit their knowledge of the application semantic domain to define, customize, and maintain the Digital Library so that it is aligned with the information and functional needs of its end-users. They provide:
  - Functional configuration parameters:  
e.g. result set format, query language, user profile formats, document model
  - Content configuration parameters:  
e.g., repositories of content, ontologies, classification schemas, authority files, and gazetteers

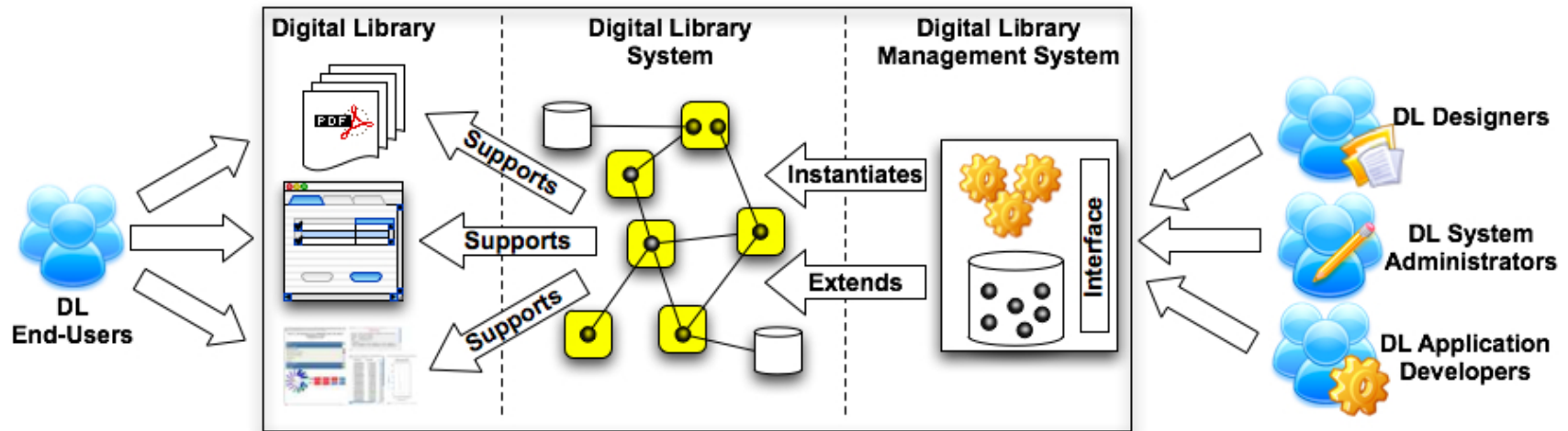
- Select the software components necessary to create the Digital Library System needed to serve the required DL and decide where and how to deploy them. They identify the architectural configuration that better fits the DLS in target ensuring the appropriate level of quality. They also provide architectural configuration parameters:  
  
e.g. selected software components, hosting nodes, components allocation



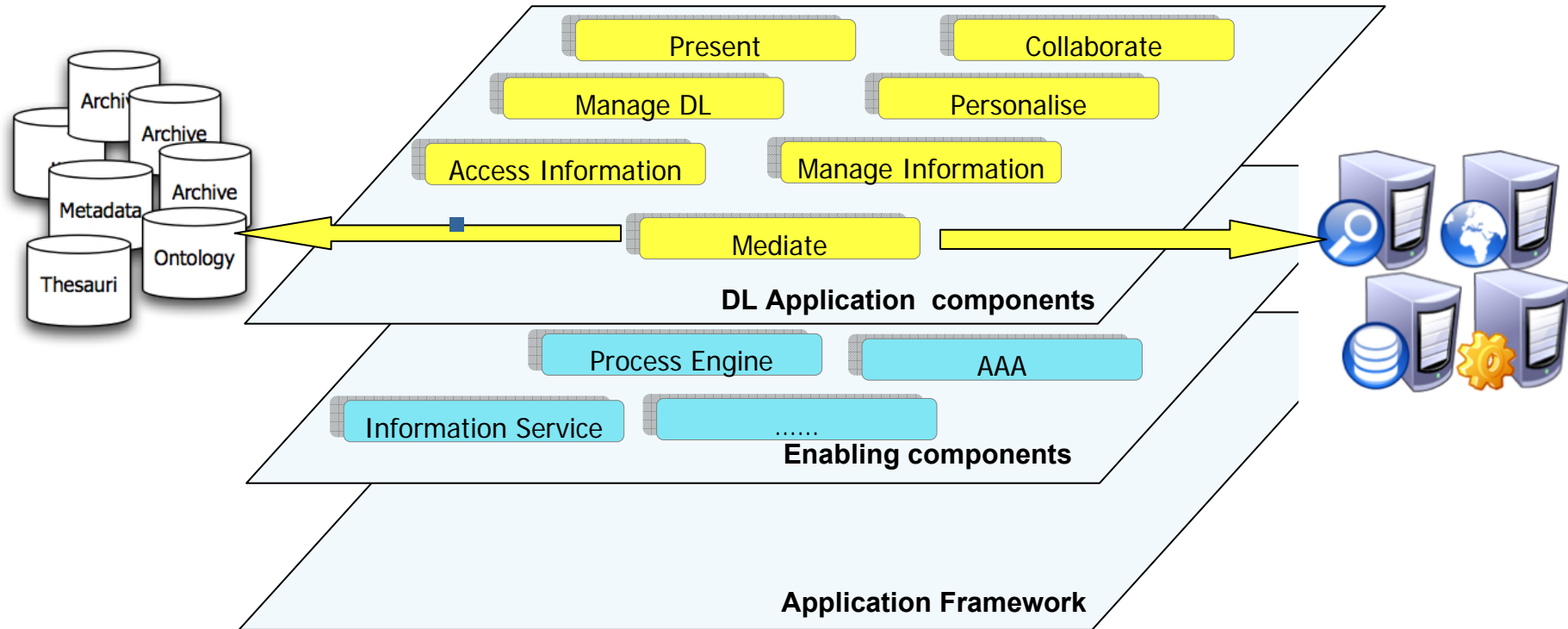
**DL**  
**End-Users**

- Develop the software components of DLMSs and DLSs, realizing the necessary functionality

# Actors – systems interaction summary

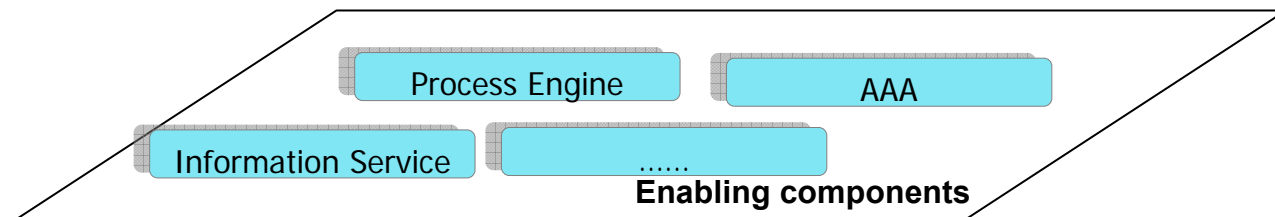


# Introductory concepts: Functions implemented by a DLS



- The middleware available at the hosting nodes which provides the run-time environment for the DL and enabling components

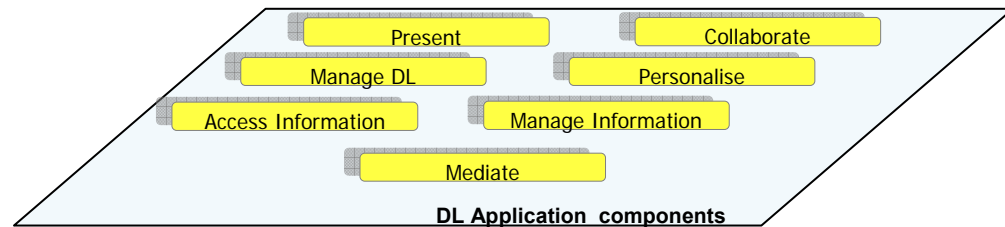
- Provide functionality supporting the operation of the components in order to guarantee DL quality of service and sustainability  
e.g. discovery and monitoring of resources, notification, workflow support, authentication and authorization





# DL Application Components

- Groups together those components that implement the DL application logic

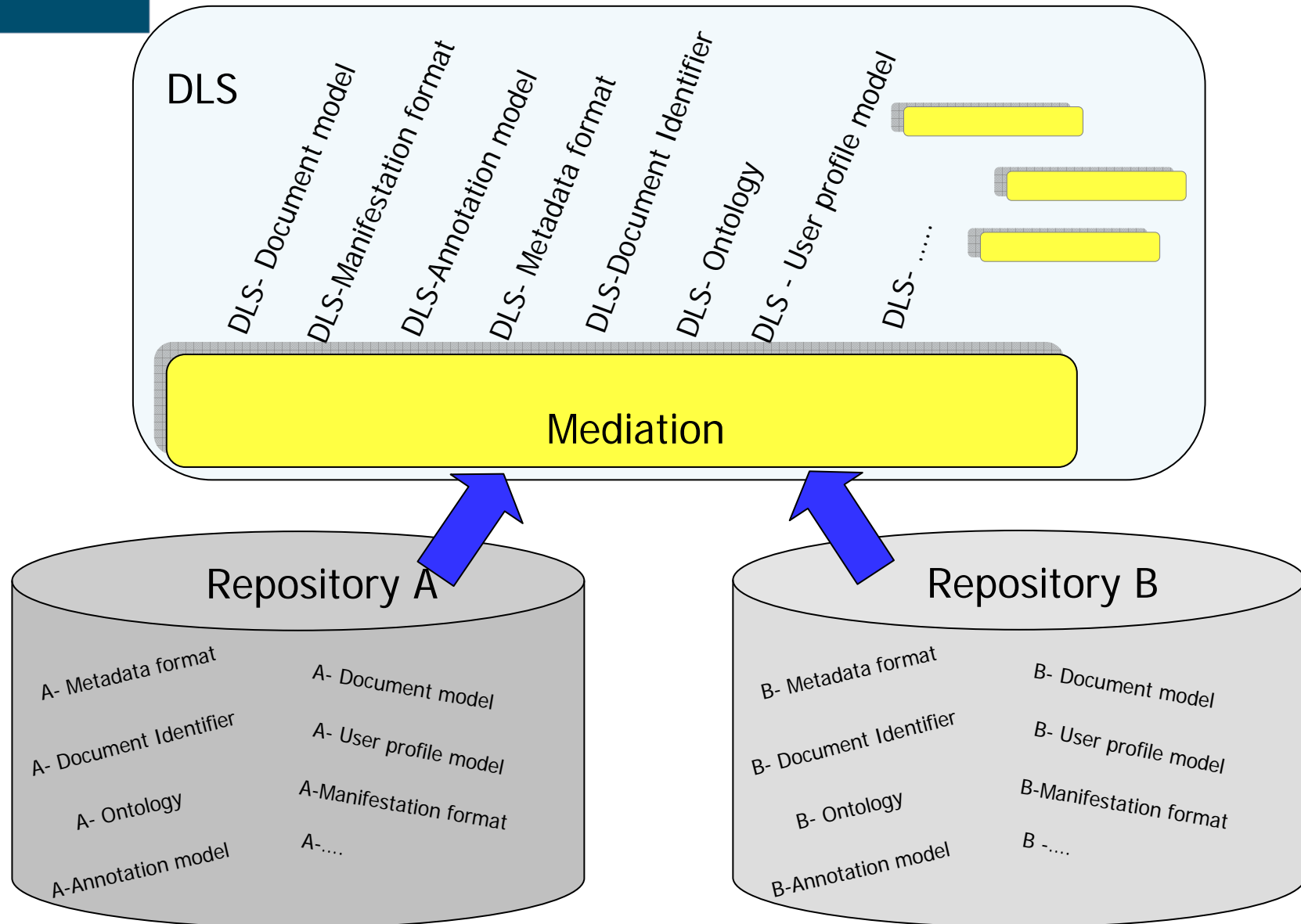


## Mediation components

- Mediate between the external heterogeneous space models of re-used resources and the resource model shared among the components implementing the DLS functionality

Mediate

# Content Mediation components



- Implement the functions that provide mechanisms for discovering and consuming information objects  
e.g. search, browse, visualise, translate

Access Information

- Implement the functions which support the production of new information objects or the update of existing information objects  
e.g. author, annotate, compose, submit

Manage Information

## Personalise components

- Implement the functions which support the content and behavior personalisation (specified by the end-user and automatically inferred by the system)  
e.g. sign-up, subscribe, apply profile, customise

Personalise

## Collaborate components

- Support actors in sharing of information, working and communicating effectively and efficiently with peers  
e.g. sharing workspace, calendar

Collaborate

- Manage the content and users resources in order to achieve the desired quality in agreement with established policies  
e.g. publish, preserve, withdraw, manage policies, manage users, manage collections

DL Management



- Deliver processes and tools for the users (both humans and services) that interact with the DL e.g. user interface, OAI-PMH publisher

Present

# PART II

## NSDL mediation & aggregation

## Outline

- From a digital collection to a digital library
- Creating a union catalog
  - OAI-PMH
  - Metadata aggregation, the NSDL experience
- Moving beyond the union catalog
  - A DL based on a digital object repository and semantic technologies: NSDL 2.0
  - Web services and information overlays
- Semantic Digital Libraries
- Digital Libraries as collaborative social spaces



Photo by Jon Crispin

# Standalone Collections

**MatDL**

**DSpace**

**EPrints**

**Fez**

**DLESE**

**arXiv**

**CWIS**

# Problem of creating a Digital Library

- Getting heterogeneous systems to work together
- Providing the user with a seamless information experience
- What services do you want to provide?
  - Search and access?
  - More (authentication/authorization, archiving)?
- How much human intervention?
- Level of perfection?

- Bring together many collections in a uniform way
- A “Union Catalog”
- Central Dublin Core Metadata Database
- Search index of metadata/content
- *Needed a protocol and mechanism to combine metadata from many collections*

“The Open Archives Initiative has been set up to create a forum to discuss and solve matters of interoperability between electronic preprint solutions, as a way to promote their global acceptance. ”

(Paul Ginsparg, Rick Luce & Herbert Van de Sompel - 1999)

Online at <http://openarchives.org>



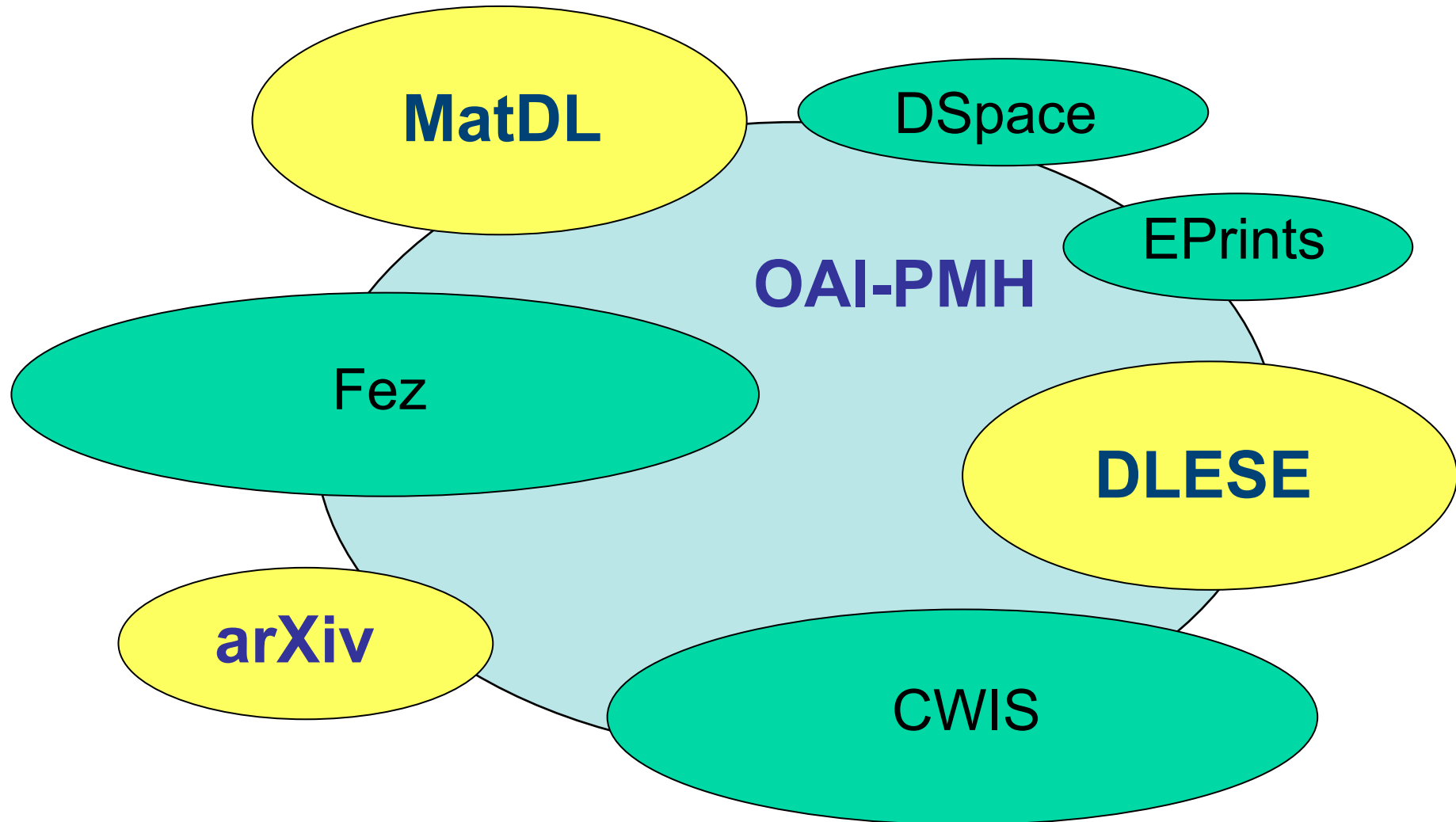
## What is the OAI now?

“The OAI develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content.” (from OAI mission statement)

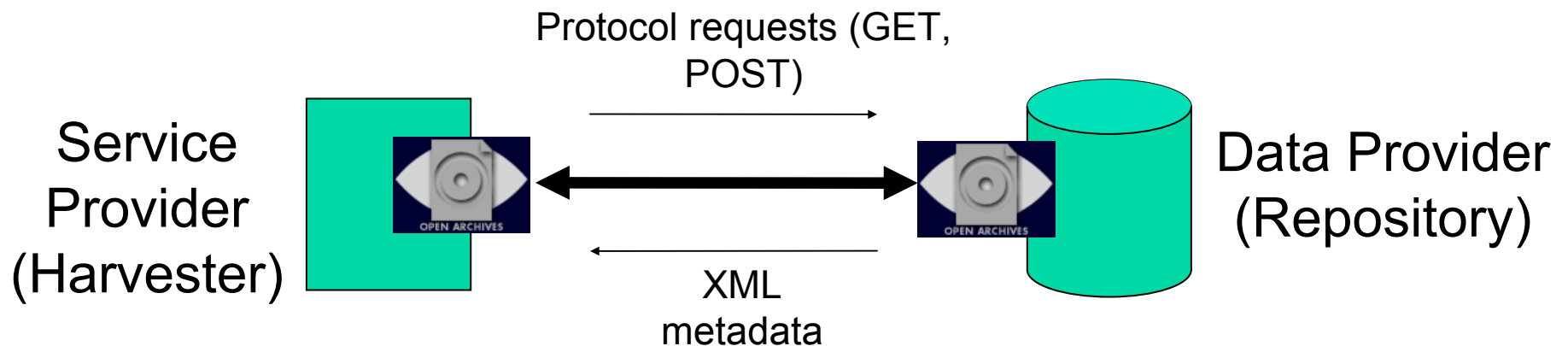
- Technological framework around OAI-PMH protocol
- Application independent
- Independent of economic model for content
- Also ... a community and a “brand”
- Something you need to complete your project 1!

# The Open Archives Initiative (OAI) Protocol for Metadata Harvesting (OAI- PMH)

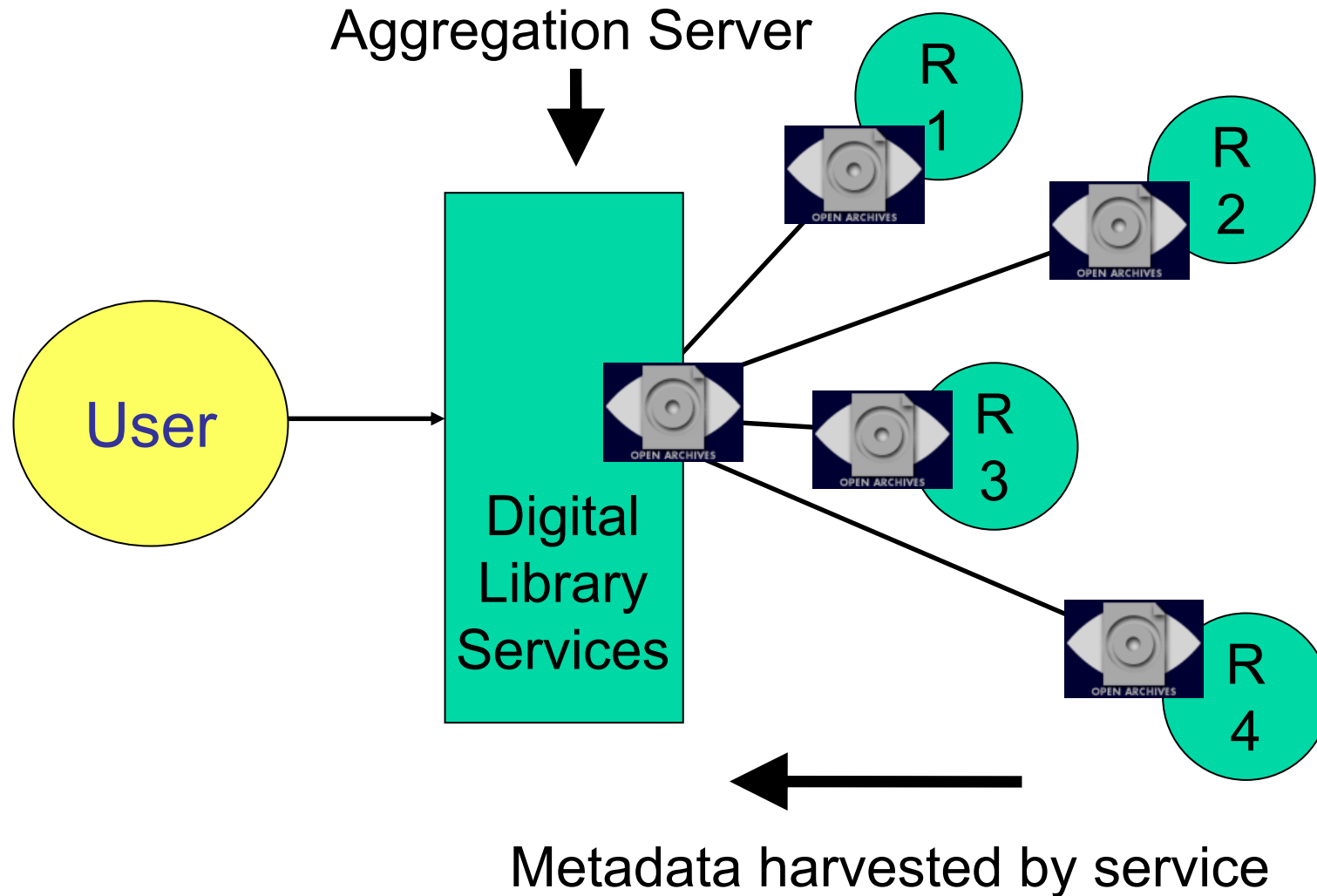
# Aggregating Collection Metadata



- PMH -> Protocol for Metadata Harvesting  
<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- Designed to allow harvesting of any XML (meta)data (schema described)
- For batch-mode not interactive use



# OAI for discovery



# OAI-PMH Data Model



← resource

item has  
identifier

all available metadata  
about this sculpture

← item

Dublin Core  
metadata

MARC21  
metadata

branding  
metadata

← records

record has identifier + metadata format + datestamp

- OAI-PMH allows record to be described in multiple metadata formats
- Dublin Core is required for “interoperability”
- Extended to include XML compound object formats: e.g., METS, DIDL
  - <http://www.dlib.org/dlib/december04/vandesompe/12vandesompe.html>

# Incremental Harvesting

- OAI-PMH records must support modification datestamps
  - YYYY-MM-DD or
  - YYYY-MM-DDThh:mm:ssZ
- Harvests allow `from` and `until` arguments to select modified records
- Deleted records cause problems (to persist, or not to persist)



- Simple notion of grouping at the item level to support selective harvesting
  - Hierarchical set structure
  - Multiple set membership permitted
  - E.g: repo has sets A, A:B, A:B:C, D, D:E, D:F
    - If item1 is in A:B then it is in A
    - If item2 is in D:E then it is in D, may also be in D:F
    - Item3 may be in no sets at all

# OAI-PMH Harvesting strategy

- Choose the sets and metadata formats to harvest
- Do initial complete harvest with no `from` and `to` parameters
- Subsequent incremental harvests start from datastamp of last response from previous harvest
- When harvesting from repositories that do not persist deleted records, must do occasional full reharvests

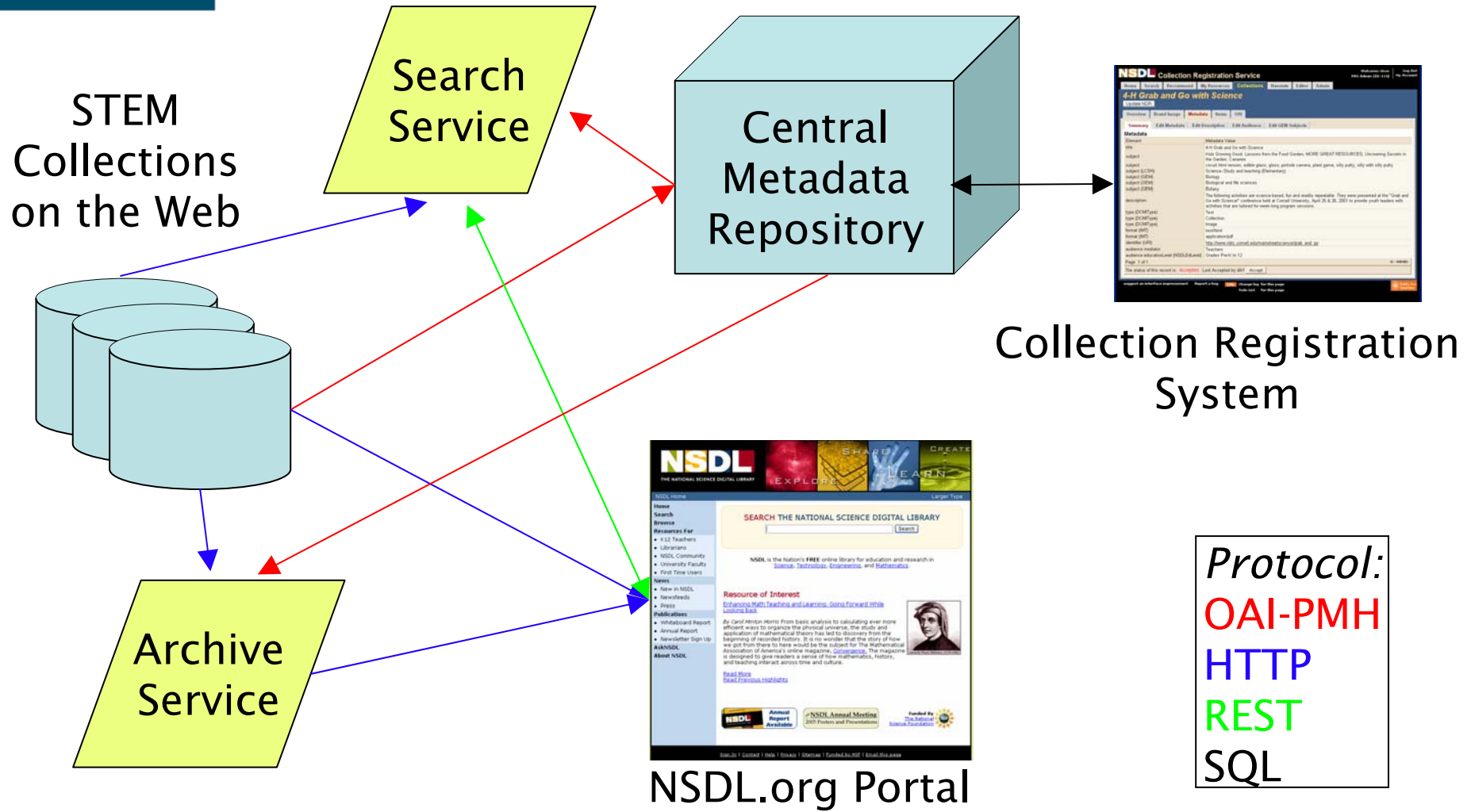
## OAI-PMH – Has it worked?

- Of course, yes...
  - Very wide deployment
  - “millions and millions of records served”
  - Incorporated into commercial systems
- But....
  - NSDL 1.0 experience illustrates some of the problems in automating metadata harvesting



# Metadata Aggregation, the NSDL experience

# NSDL 1.0: Low-Barrier Architecture

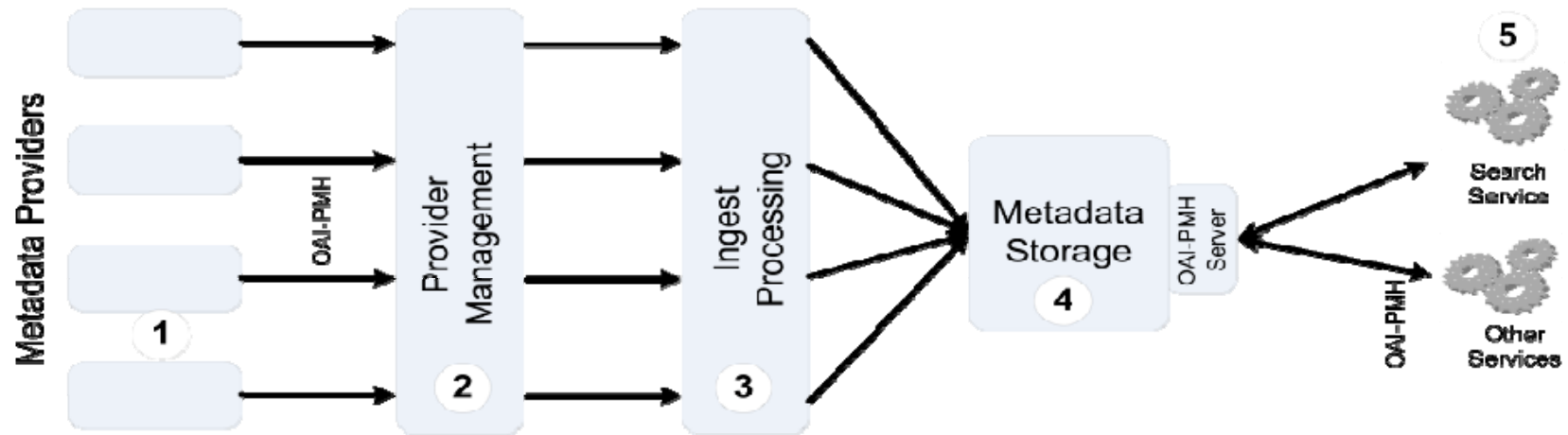


*Protocol:*  
**OAI-PMH**  
**HTTP**  
**REST**  
**SQL**

## NSDL 1.0 Choices

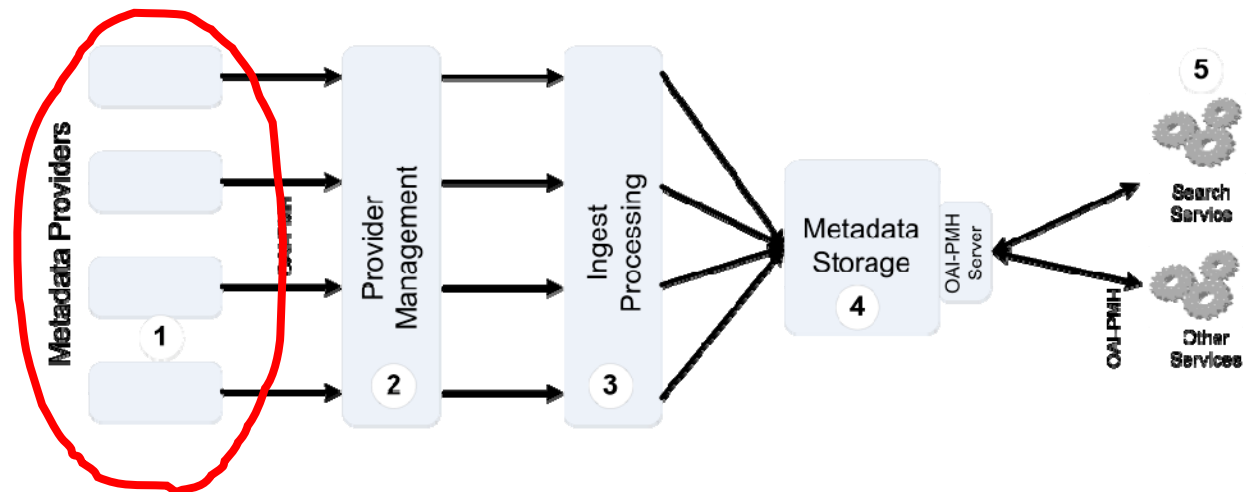
- Aggregating heterogeneous collections: lowest common denominator metadata: Dublin Core
- Automating as much as possible: OAI-PMH with automated harvesting/serving
- Resource selection delegated to collection providers
- Effort is by both central staff and collection providers

# NSDL Metadata Flow



# Metadata Provider Model

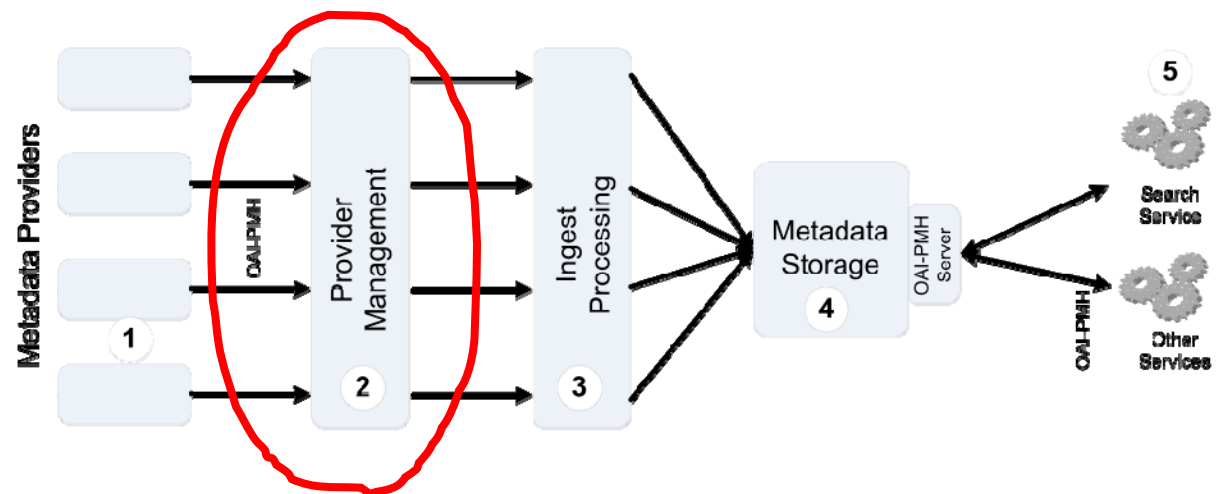
- Core integration (CI) collects, normalizes, manages, and re-distributes metadata
- Similar to shared cataloging paradigm (e.g., WordCat, UC SCP, etc.), but:
  - Dublin Core based
  - OAI-PMH harvesting





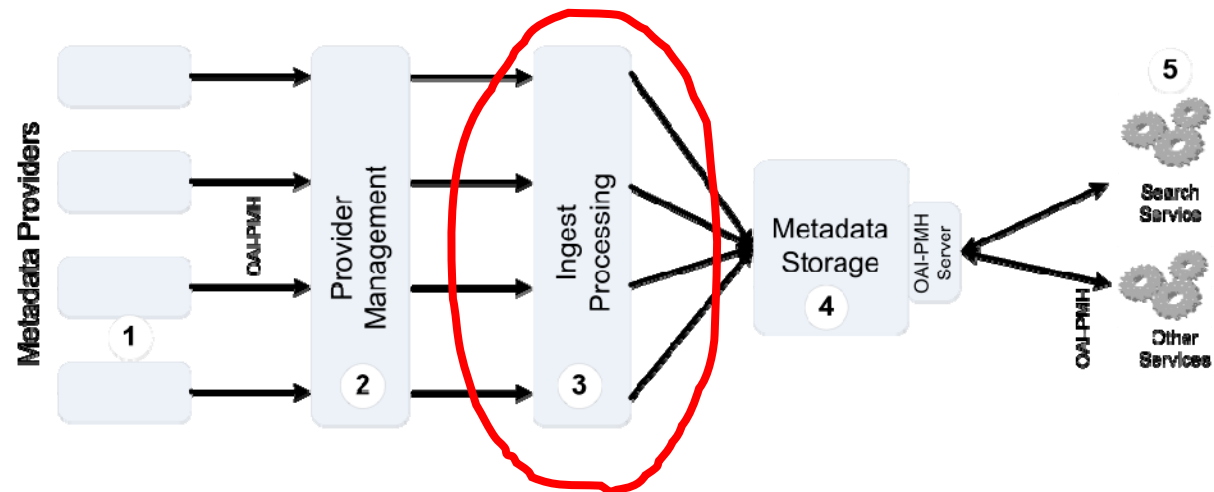
# Automated harvesting

- Collections validate their OAI-PMH server
- CI registers collection (CRS)
  - harvest schedule, baseURL, set information...
- Full harvest initiated
- Subsequent incremental harvests according to schedule
  - automated emails if problems



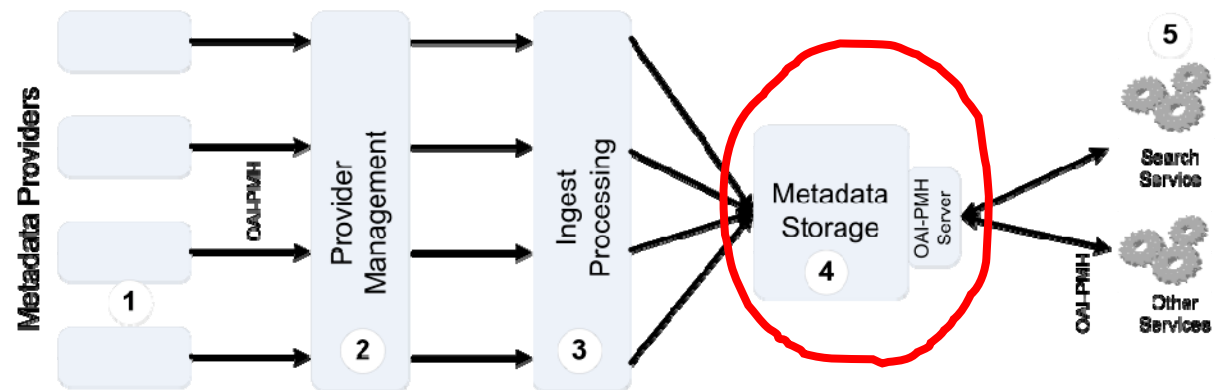
# Ingest and Aggregation

- Transform raw harvested metadata to normalized form
- Many problems are idiosyncratic to collections or even individual harvests
- Apply some general “safe transforms” (only DC)
  - fix encodings, URIs, known DCMI types



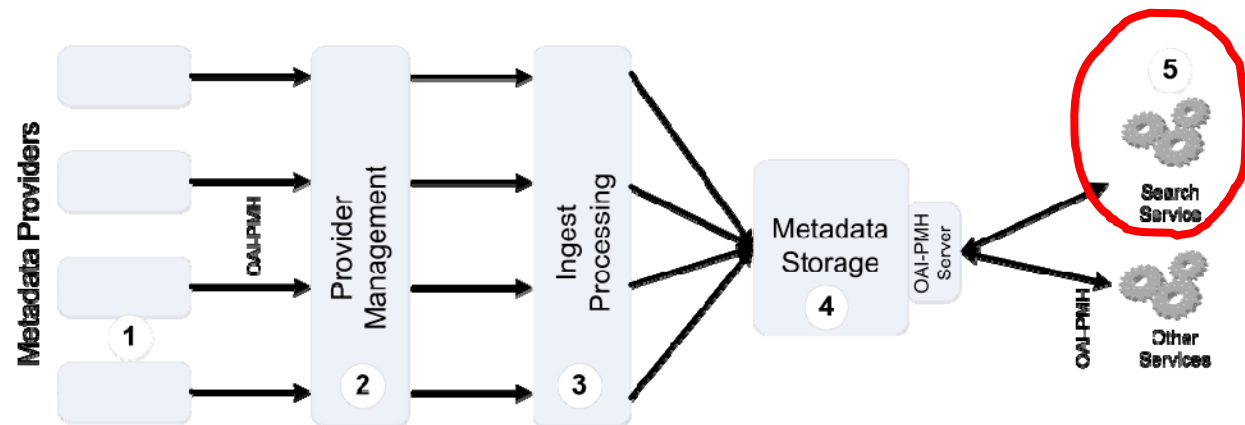
# Aggregated Metadata Storage and Exposure

- Oracle database
- Two server instances
  - ingest and storage
  - stage for oai-pmh server



# Metadata Search

- Lucene-based
  - harvests and indexes metadata consumed via OAI-PMH
- Also fetch and index text resource “referenced by metadata”
  - Nutch-based
- Metadata-centric search
  - Issues of resource equivalence



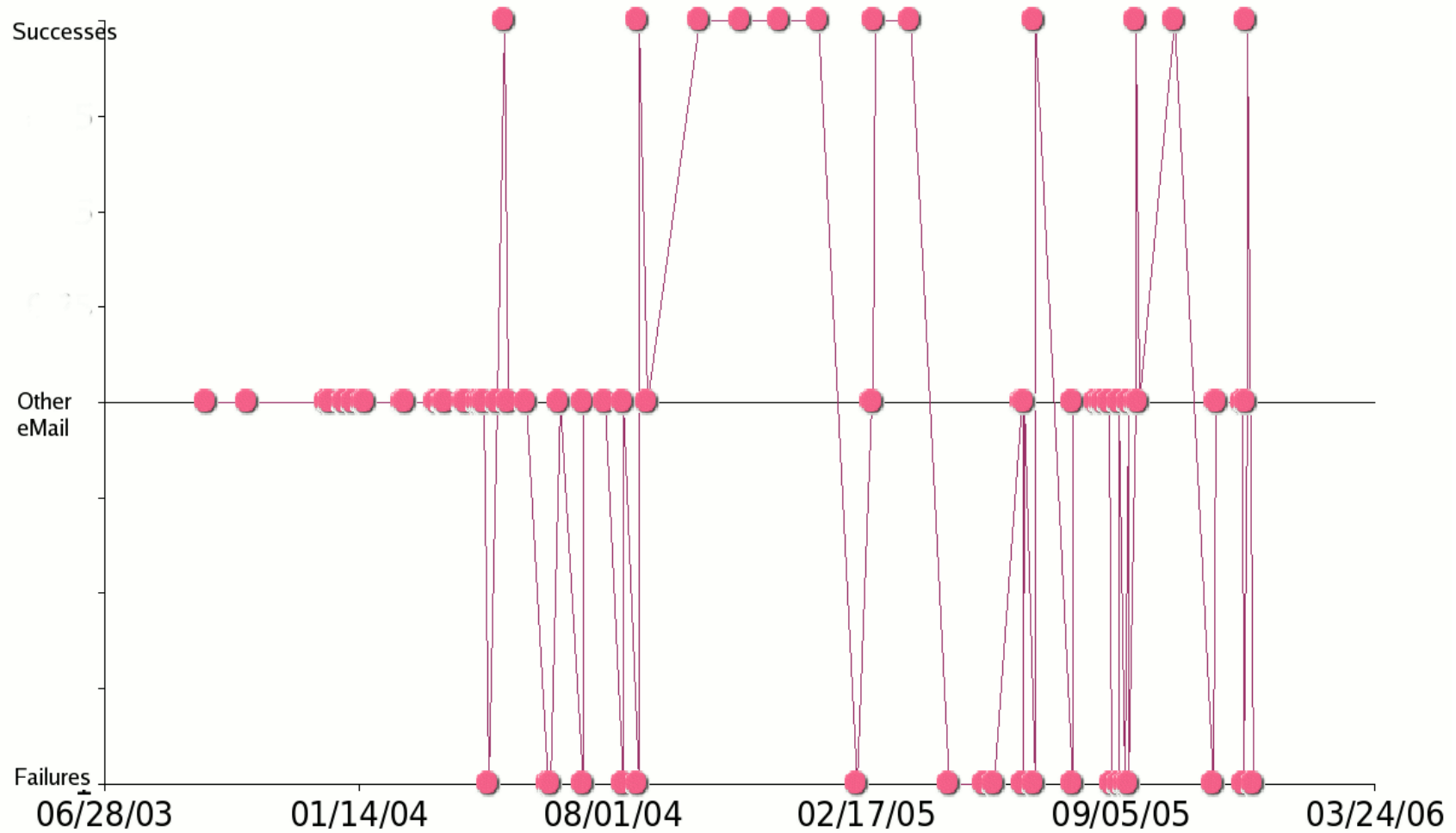
**People and organizations don't want to take the time and expense to supply metadata**

- Mandate from NSF covered only funded collections
- Commercial providers of STEM resources slow to realize potential for increased revenue

**Participating as a metadata provider is complicated  
by a “*knowledge gap*”**

- Three skill sets that are frequently distinct:
  - Domain expertise
  - Metadata expertise
  - Technical expertise
- Things that have helped:
  - Documentation somewhat
  - Lots of hand-holding
    - Many months between first contact and successful harvest
  - Collection Workflow Integration System (CWIS)

# Also's Harvest History



## Harvested metadata is not necessarily useful metadata

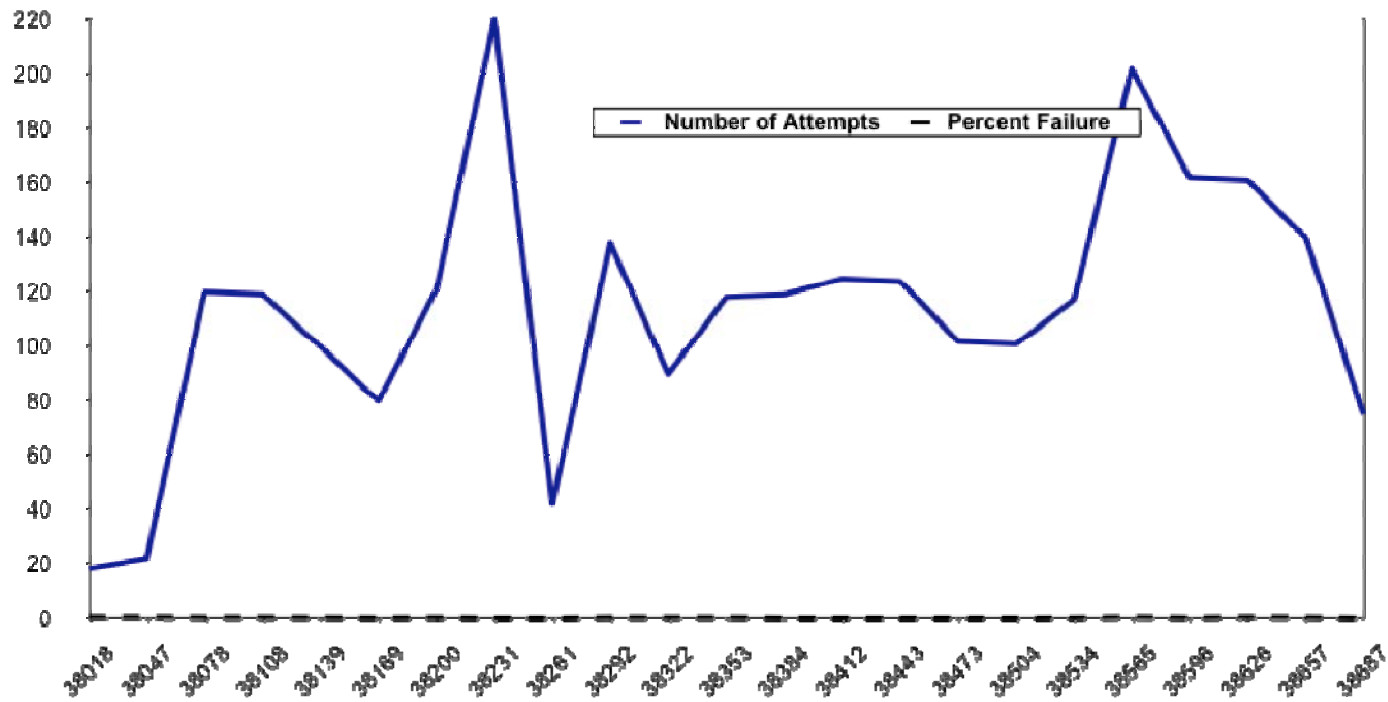
- Generally applied “safe transforms” are limited in scope
- The general problem of metadata quality remains unsolved (and maybe unsolvable without lots of human intervention)
- Heterogeneous collections resulted in extreme variability of metadata, exacerbated by mapping to Dublin Core



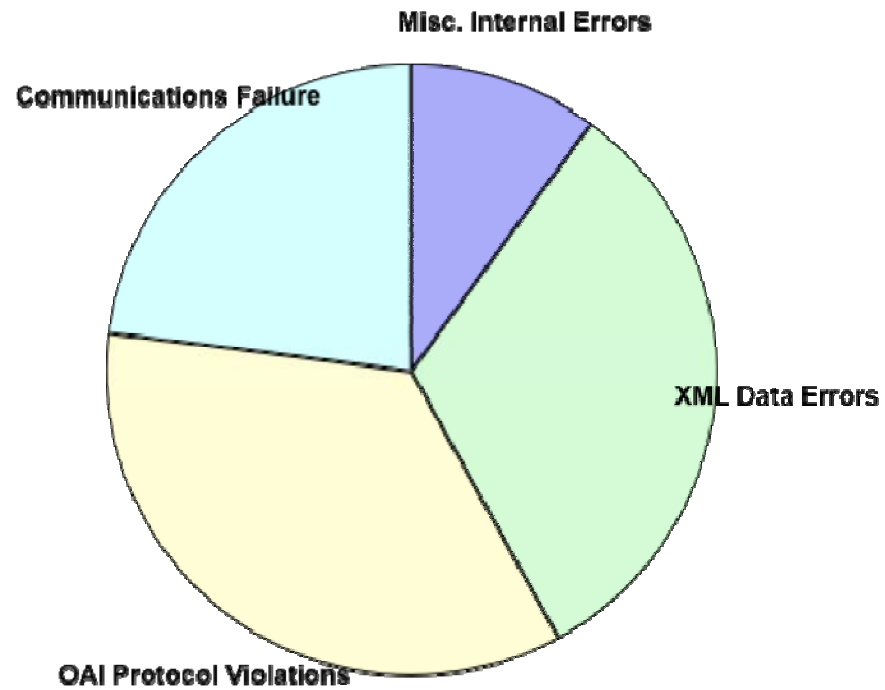
## OAI-PMH is not necessarily low-barrier and automatic

- It incorporates lots of details and assumed knowledge
  - UTF-8, XML schema validation, URL encoding, date stamping, resumption, ...
- Initial success does not persist in too many cases
  - Failure rate of subsequent harvests is high
- Incremental harvest is a nice concept but...
  - Support for “deleted” records is inconsistent
    - less than 50% of providers claim to persist deletions
    - many persistent claims are false
  - Server failures, harvest failures require full harvest “resync”

# NSDL Harvest Failure Rate



# Harvest Failure Categories



## Human cost of large-scale harvesting is high

- Considerable email support, much of it human mediated
  - ~170 messages per year per provider
- Average 98 messages exchanged to arrange first harvest
  - Extended to order 100s messages
- In many cases months of hand-holding for first successful harvest
  - which then had subsequent failures
- **THIS ALL ADDS UP!!**

## NSDL 1.0 Summary

- Metadata Repository was quick to implement using known technologies, but
- Limited model
  - Metadata-centric orientation
  - No content – only metadata
  - Resource selection delegated to collection providers
  - Limited relationships – collection/item
  - Limits on context, structure, and access
  - Severe limits on contribution and collaboration
  - One-way data flow: NSDL → Users

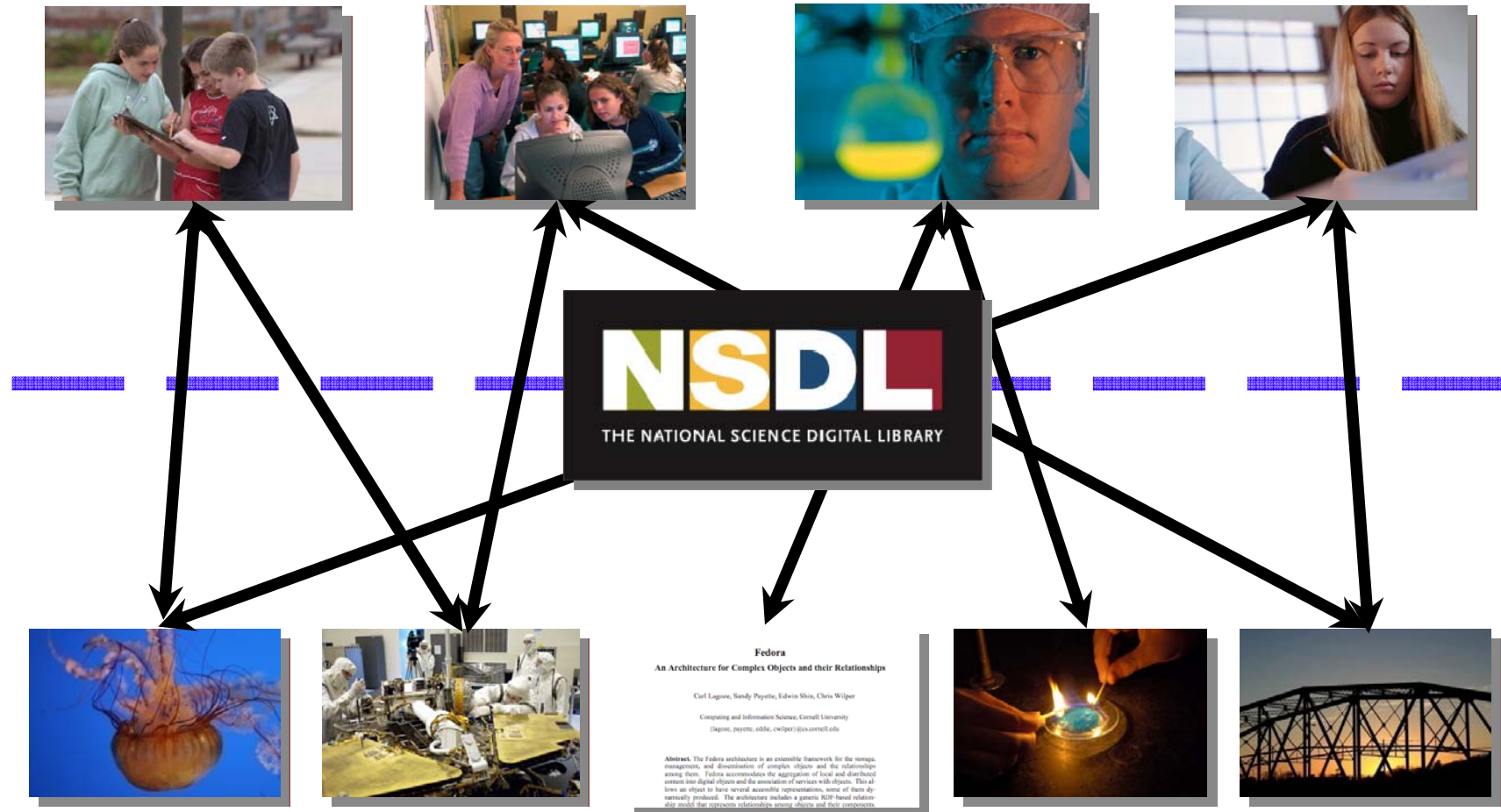


# Moving beyond the union catalog: NSDL 2.0

## NSDL 2.0 vs. NSDL 1.0

- Supports better editorial and community collection development
- Supports resource content and context
- Builds on semantic technologies to situate resources in their context: linked to related concepts; with user ratings; with codes and data
- Enables community tools for selecting, organizing, evaluating, annotating, contributing, and collaborating
- Provides two-way data flow: NSDL ↔ users

# Object-Centered Sociality

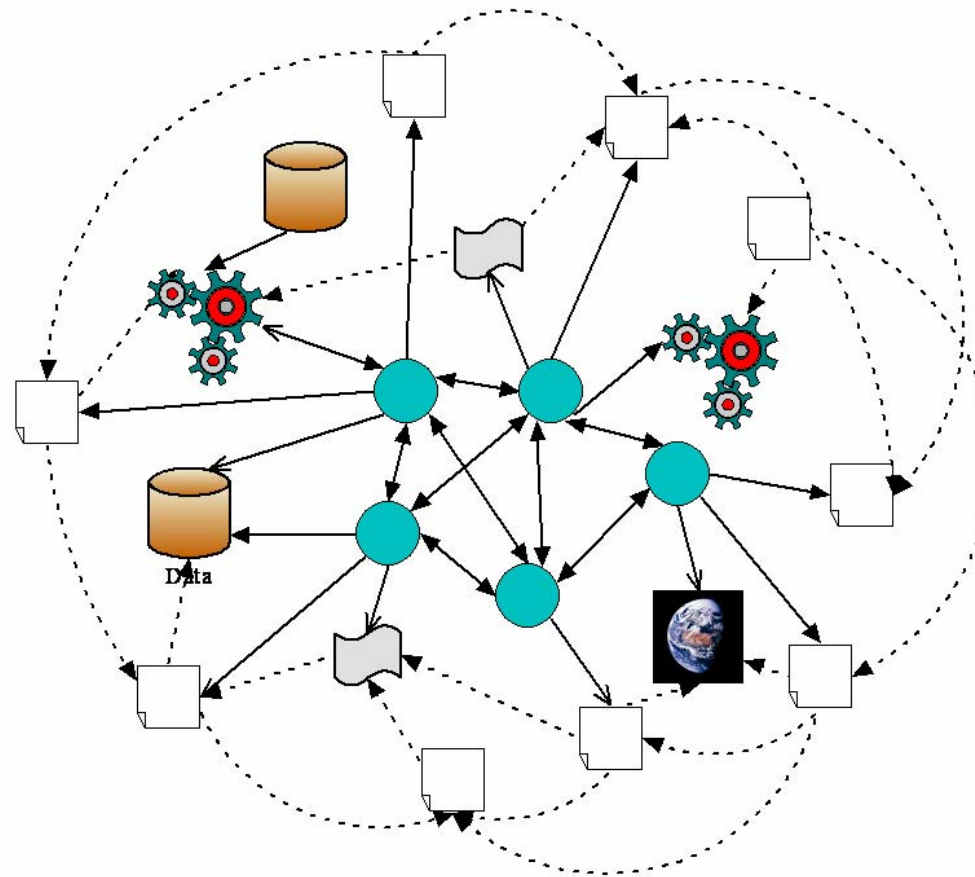
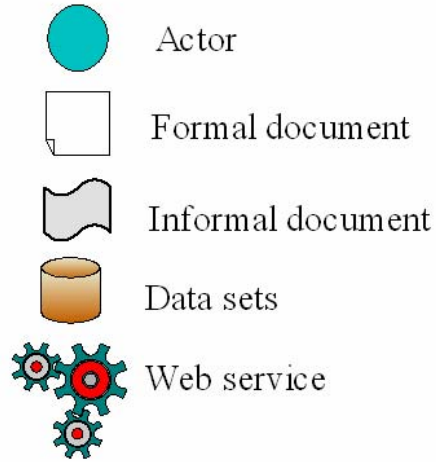




## In Architectural terms, create an NSDL Data Repository that

- Stores both content and metadata
- Allows arbitrary relationships among resource and metadata objects: organization, annotation, citation
- Easily integrates with existing semantic technologies
- Is accessible through web service architecture of remixable data sources and transformations

# The Fedora Vision: A Repository for Rich Information Networks



# Fedora: the NDR middleware

- A Flexible, Extensible **D**igital **O**bject **R**epository **A**rchitecture (<http://www.fedora.info>)
- Open source project with \$2.2 million in Mellon funding 2002-2007, evolving into the Fedora Commons
- Collaboration of Cornell and Univ. of Virginia
- Key funded users include:
  - eSciDoc project (collaboration of the Max Planck Society and FIZ Karlsruhe)
  - Public Library of Science (Topaz Foundation)
  - VTLS Corp., Harris Corp., Library of Congress
  - Australian Research Repositories Online to the World
  - Royal Library Denmark, National Library, and DTU

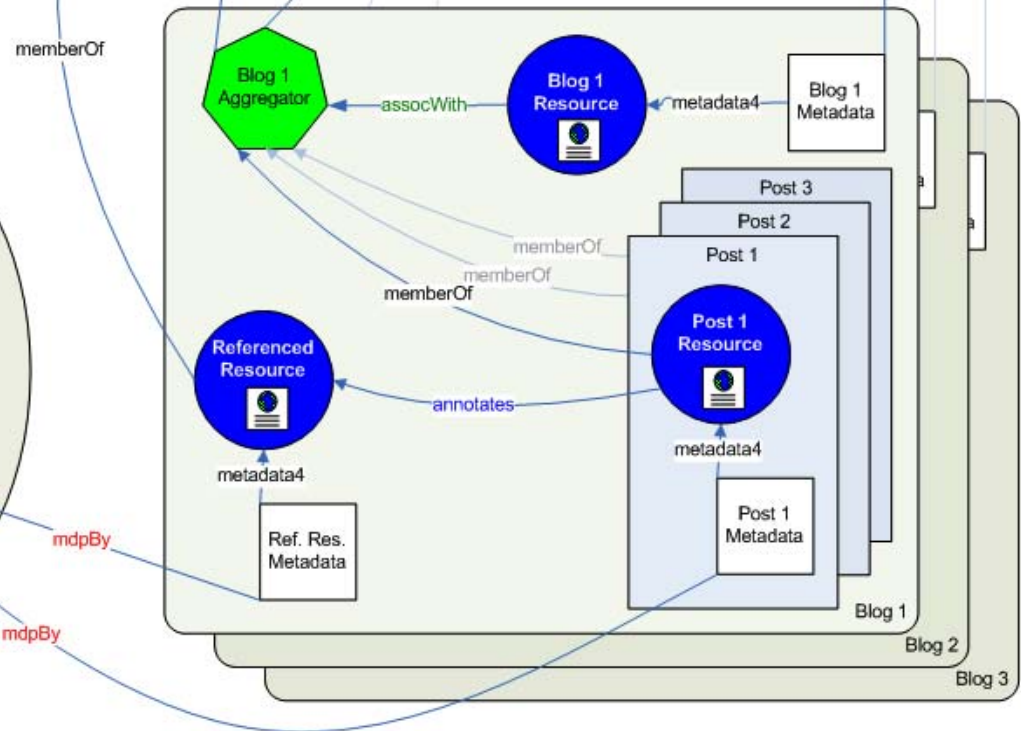
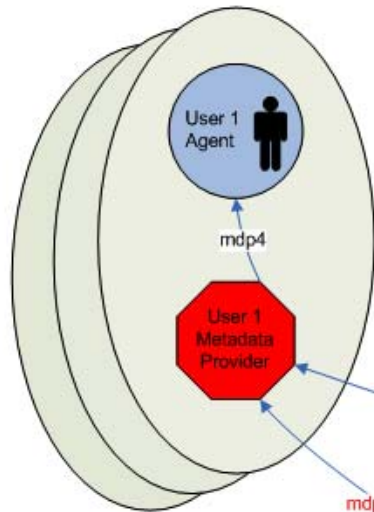
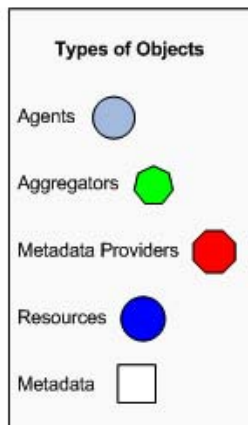
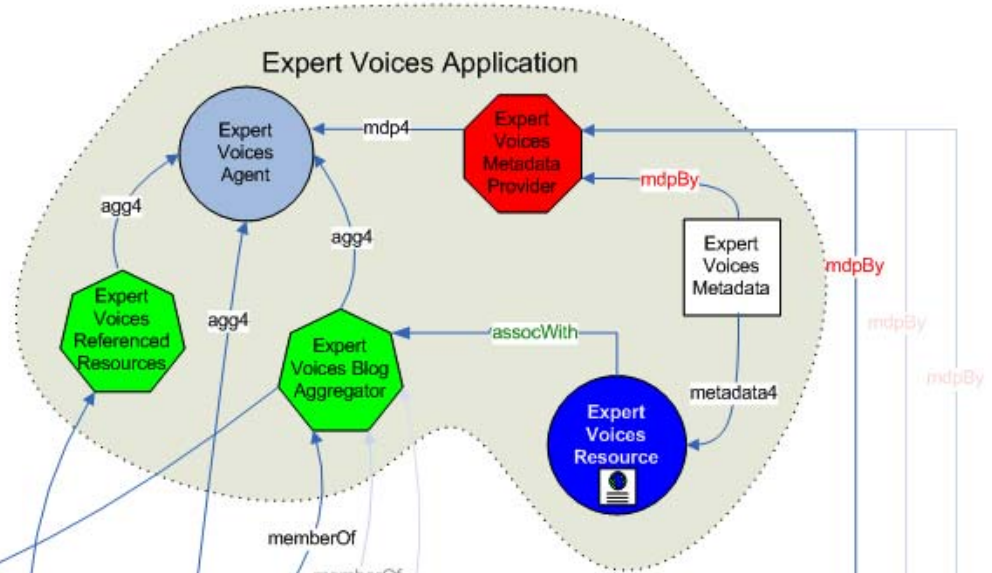
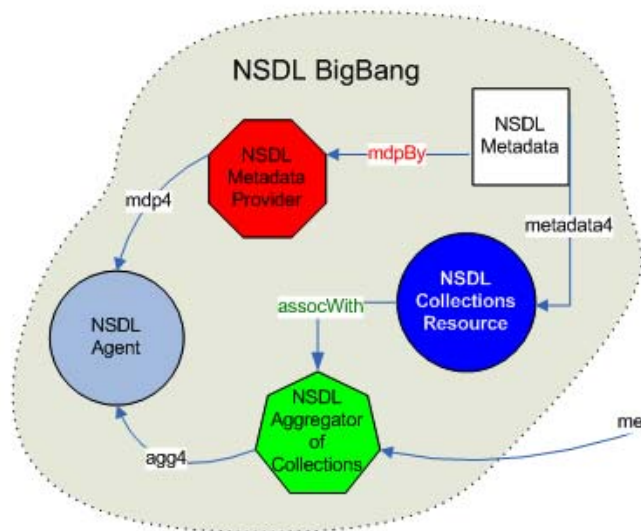


## What is Fedora?

- An architecture, toolkit, and implementation: middleware, not a vertical application
- Stores arbitrary internal and external digital objects, disseminations (transformations and combinations), relationships among objects
- Entirely SOAP/REST based, disseminations are URLs
- XML data store; RDBMS cache; RDF triplestore supports relationship queries

- Multiple Object Types:
  - Resources (with local or remote content)
  - Metadata
  - Aggregations (collections)
  - Metadata Providers (branding)
  - Agents
- Relationships with arbitrary graph queries:
  - Structural (part of)
  - Annotation (relates to)

# NSDL DATA REPOSITORY EXPERT VOICES



## NSDL Data Repository (NDR)

- References to roughly 2 million selected STEM resources on the web
- Sourced metadata statements about those resources
- A REST API to allow authenticated access by Pathways, providers, tool builders
- Since January 2007, in production at [nsdl.org](http://nsdl.org)

# NDR API Characteristics

- Uses REST calls for all interactions; uses handles (DOIs) for all external references
- Ensures external applications can't violate the NDR model constraints
- Disseminations allow combining metadata from multiple sources, or related content
- Authentication: Requests signed with private key associated with an agent
- Authorization: Agent can become a metadata provider or aggregator; can create resources
- API/NDR instance available for development and testing ([ndrtest.nsdlib.org](http://ndrtest.nsdlib.org))



- NDR API provides authenticated access to the repository - Expert Voices (EV), NCS
- Search service REST API supports programmatic search - Pathways, Strand Maps
- OAI-PMH ingest and server allow batch aggregation and dissemination - all collections, search
- Shibboleth Community Sign-On for user authentication - EV, nsdl.org, Engr Pathway, ...
- SDSC Archive - REST access to archived resources - SERC, nsdl.org
- RSS feeds - NSDL editorial content, EV



# NSDL 2.0 as a Digital Library Infrastructure



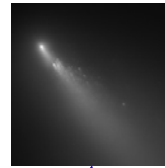
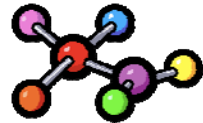
- NSDL 2.0 specializes Fedora for digital library use – provides model and API
- NSDL 2.0 provides a set of tools: collection management, ingest control, search, authentication
- NSDL itself is an instantiation of this DLMS for STEM education
- Tools and architecture are agnostic
- By exposing all the information in the library on the web, P2P, remixes, etc. are all possible

# An Information Network Overlay

- Think of the NDR as a lens for viewing science content on the net
- Content can be:
  - Local: stored directly in the NDR
  - Remote: accessed through a URL
  - Computed: derived from a database or web service
  - Archived: an older version stored at SDSC
- It all has a repository-based URL

# Network Overlay View

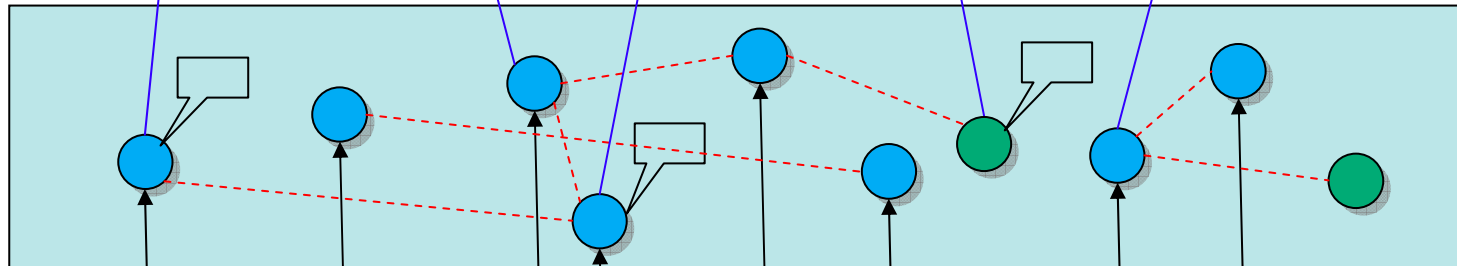
User View



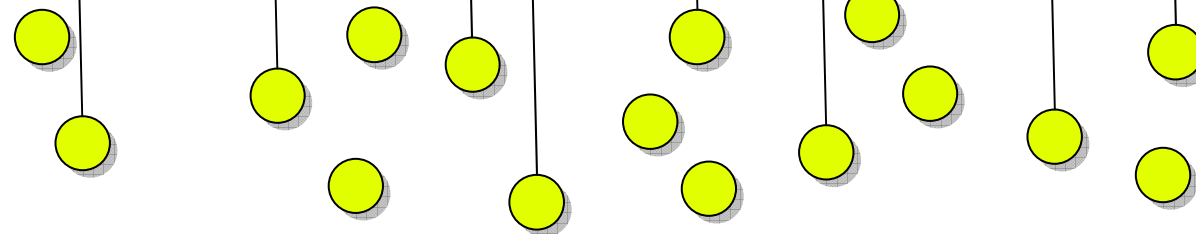
API/UI



Repository  
View with  
Relations &  
Annotations



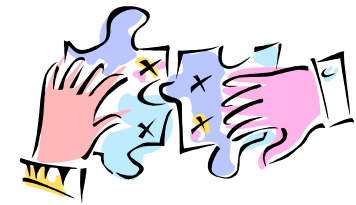
Resources  
on the Web



# What is a Semantic Digital Library?

## Semantic digital libraries

- **integrate** information based on different metadata, e.g.: resources, user profiles, bookmarks, taxonomies
- provide **interoperability** with other systems (not only digital libraries) on either metadata or communication level or both
- delivering more robust, **user friendly and adaptable search and browsing** interfaces empowered by semantics



# How are Semantic Digital Libraries different?

- Semantic digital libraries extend digital libraries by
  - describing and exposing its resources in a machine ‘understandable’ way
  - resources can be
    - contents, digital artifacts
    - organization of objects (e.g. collections)
    - users, user communities
    - controlled vocabularies, thesauri, taxonomies
  - expose the semantics of their metadata in terms of an ontology
    - defined using a formal language
  - deliver mediation services for communication with other systems

# Some Existing Semantic Digital Library Systems

- **FEDORA**
  - delivers flexible service-oriented architecture to managing and delivering content in the form of digital objects
- **SIMILE**
  - extends and leverages DSpace, seeking to enhance interoperability among digital assets, schemata, metadata, and services
- **JeromeDL**
  - a social semantic digital library makes use of Semantic Web and Social Networking technologies to enhance both interoperability and usability
- **BRICKS**
  - aims at establishing the organizational and technological foundations for a digital library network in order to share knowledge and resources in the cultural heritage domain.





# Digital Libraries as Collaborative Social Spaces



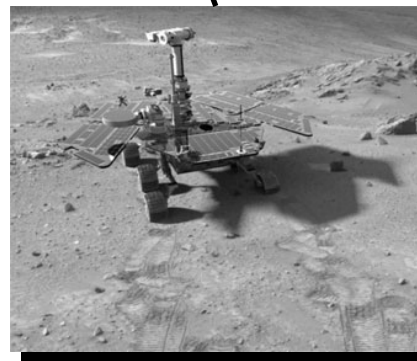
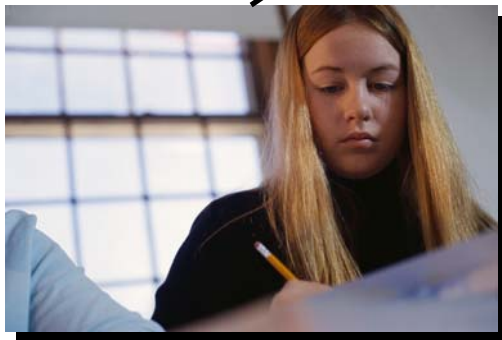
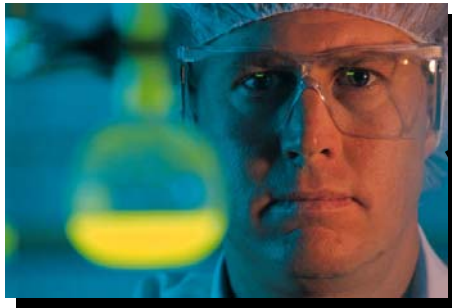
## NSDL 2.0 Platform is a set of capabilities

- Supports creating context around resources
- Enables the NSDL community to directly contribute resources and context
- Represents a web of relationships among science resources and information about those resources
- Provides web services interfaces to accomplish these tasks

## Putting the capabilities in the hands of the users

- The NDR and services provide the platform, but we still need the applications
- Solution 1: Leverage the existing successful models: blogs, wikis, bookmarking/tagging
- Solution 2: Leverage the existing software: WordPress, MediaWiki, Connotea, Sakai
- Solution 3: Engage with partners and the broader community to build applications to the platform

# ExpertVoices



- The NSDL Blogosphere, live at <http://expertvoices.nsdlib.org>
- Topic-based discussions (e.g. forensics) linked to related library resources
- A way for NSDL community members to become NSDL contributors: of resources, questions, reviews, annotations, metadata
- Wordpress-based multi-user multi-blog application (open source, plug-in architecture)
- Owner controls publication of entries as NSDL resources and visibility of comments
- Entries can contain linked references to NSDL resources, references to URLs that should become resources, and new resource metadata



Hurricane Floyd/Photo by NASA





Photo by Jon Crispin



Expert Voices: Speaking of Something Interesting.

Search NSDL

NSDL.org > Expert Voices >

Larger Text



All blogs grouped by audience category

[Change View]

### K12 Teachers

[Teaching Measurement at the Middle School Level](#) Updated: Jun 2nd, 2006

[Bringing the Field to the Classroom: Birds](#) Updated: Oct 4th, 2006

[Meeting web kids on their own turf](#) Updated: Sep 21st, 2006

[Boneyard Science: Investigating Forensics](#) Updated: Apr 29th, 2006

### University Faculty

[Presentation of Math on the Web](#) Updated: Jun 20th, 2006

[How can digital education help the Gulf Coast?](#) Updated: Oct 3rd, 2006

### Librarians

[NSDL Whiteboard Report Talk Back](#) Updated: Oct 5th, 2006

[How can digital education help the Gulf Coast?](#) Updated: Oct 3rd, 2006

### NSDL Community

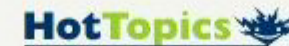
[NSDL News Topic Center: Current News Information About S.T.E.M.](#) Updated: Oct 3rd, 2006

[NSDL Whiteboard Report Talk Back](#) Updated: Oct 5th, 2006

### Informal Learners

[Bringing the Field to the Classroom: Birds](#) Updated: Oct 4th, 2006

[Meeting web kids on their own turf](#) Updated: Sep 21st, 2006



General

Education Health

Mathematics **Science**

Social Studies

Technology Blogroll

[Collecting Data Seminar 2 Blog](#)

[Related Links Bio](#)

RecentPosts

[ObservationTower](#)

Logged in:

[About Expert Voices](#)

[Expert Voices Help](#)

[Recommend a Blog Topic](#)

[Expert Voices User Feedback](#)

Expert Voices Pre-Beta Version

Powered by [WordPress MU](#).



## NSDL News Topic Center: Current Science Information

« Previous Post: [Baby Plesiosaur Discovered in Antarctica](#)

### Polar Bears Face Sudden Meltdown

Wednesday, February 7th, 2007 2:55 pm

Written by: **Sarah Birns**

According to a report from Wednesday, February 07, the polar bear is officially being viewed by the U.S. government as an endangered species: <http://www.polarbearsinternational.org/in-the-news/threatened-species/>. Due to global warming, the waters in the north are beginning their annual freeze later than usual, and the ice is, inversely, breaking up earlier, decreasing the time of solid ice. This change is detrimental to the mammals, as they need to be on solid ice in order to hunt. NSF's International Polar Year website, [http://nsf.gov/news/special\\_reports/arctic/index.jsp](http://nsf.gov/news/special_reports/arctic/index.jsp), explains, "Besides warmer air temperatures and melting permafrost, sea ice, and glaciers, other substantial changes have taken place in the Arctic. Out-of-place plant communities are appearing in the high latitudes, subsurface ocean currents are warmer and precipitation patterns have changed, all of which affect animal habitats and migration routes. With fewer than 25,000 polar bears remaining on the planet, the situation appears grim unless serious actions are undertaken to address the issue of global warming. For more information, please visit the following NSDL resources:

#### **Slide Presentation on Climate Science and Potential Consequences of Climate Change:**

<http://oaspub.epa.gov/eims/eimsapi.dispdetail?deid=19495>

A powerpoint presentation on climate changes.

Grade Level: Undergraduate, Graduate

#### **The Polar Bear Tracker:**

Sign In

Information

About

Categories

Education

General

Health

Mathematics

Science

Social Studies

Technology

Bookmarks

ExpertVoices FAQ

Previous Posts

February 2007

December 2006

November 2006

October 2006

June 2006

May 2006

Expert Voices Help

Expert Voices User Feedback

**Expert Voices Beta Version**  
Powered by [WordPress MU](#).







NSDL Home > Search

Larger Text

- Home
- Search Options**
- Browse
- Resources For
  - K12 Teachers
  - Librarians
  - NSDL Community
  - University Faculty
  - First Time Users
- News
  - New in NSDL
  - Newsfeeds
  - Press
- Publications
  - Whiteboard Report
  - Annual Report
  - Newsletter Sign Up
- Participate in NSDL
- About NSDL
- Annual Meeting
- Expert Voices
  - NSDL's Blogosphere

**NSDL Search Results**

[Search Tips](#)

endangered polar bears

Search Resources  Search NSDL.org

94 Results • 1 2 3 4 5 → [Next](#)

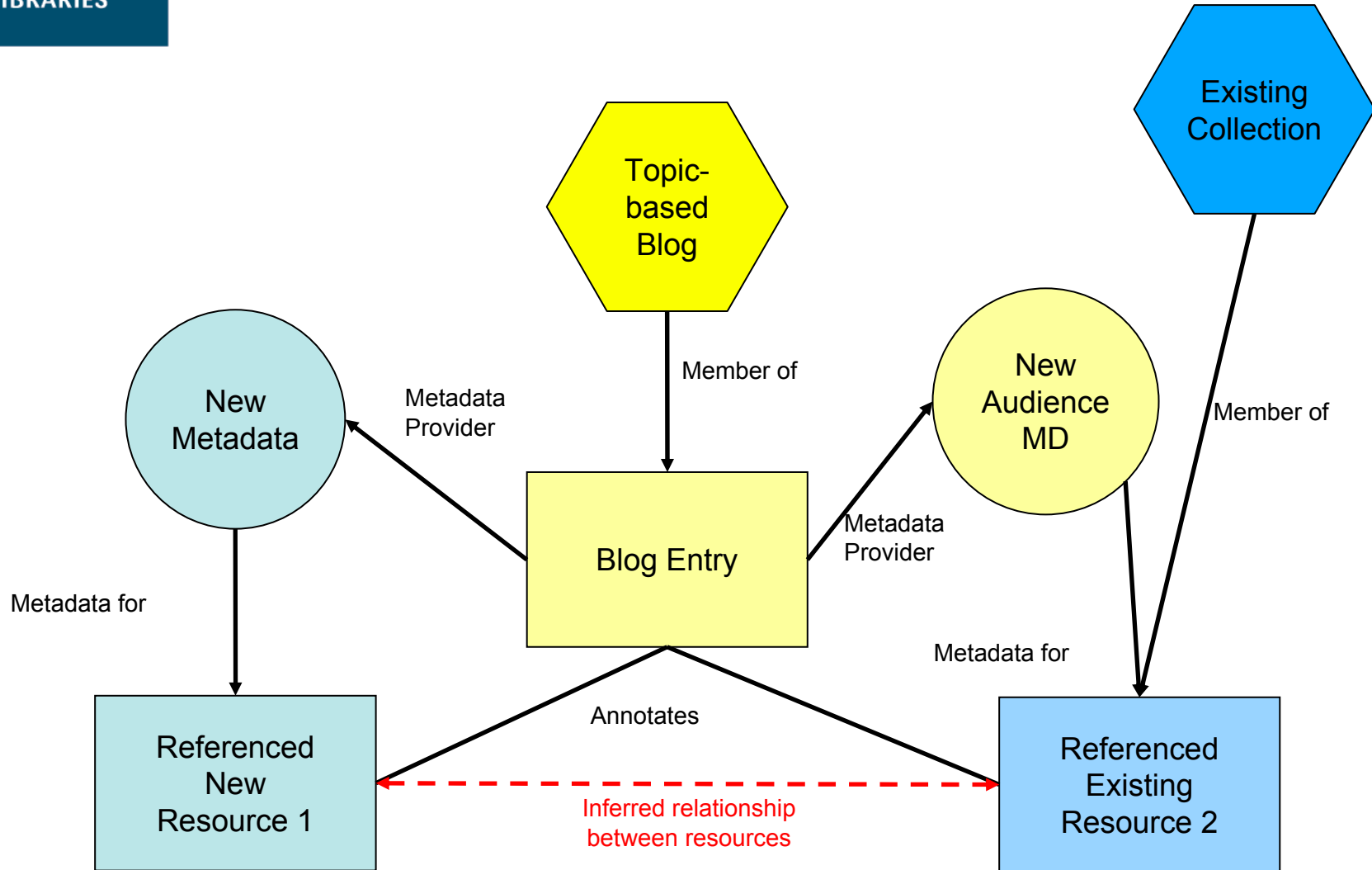
Title/Description	Resource Format
<p><a href="http://www.polarbearsinternational.org/in-the-news/threatened-species/">http://www.polarbearsinternational.org/in-the-news/threatened-species/</a> Information about the polar bear's newly acquired status as endangered. <a href="#">View all related information</a></p>	
<p><b>Polar Bears Face Sudden Meltdown</b> Global warming is having an increasingly adverse effect on polar bears, and the mammals have recently joined the ranks of the endangered species list. With... <a href="#">View all related information</a></p>	
<p><b>Toxic Chemicals Initiative</b> This World Wildlife Fund initiative is aimed at banning the most deadly, persistent pollutants (such as DDT, PCBs, and dioxins) that endanger every species.</p>	[text]

## What isn't EV?



- Expert Voices ≠ LiveJournal
  - Contributors are carefully selected, contributions are about science, the process of science, and education

## NDR Entry for Expert Voices



## OurNSDL: NDR- integrated Wiki

- Community of approved contributors (e.g. teachers, librarians, scientists) are granted edit access on OurNSDL wiki
- New resources and metadata are created as wiki pages and reflected into the NDR
- Non-wiki-based NDR resources and metadata are displayed as read-only wiki pages, subject to comment and linking
- User and project pages organize NDR resources
- Now implementing MediaWiki extensions

#### Navigation

- [Main Page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

#### Search

#### Toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)

[Article](#)[Discussion](#)[Edit](#)[History](#)[NSDL](#)

## AboutOurNSDL

### Contents [hide]

- 1 [MediaWiki Version](#)
- 2 [NSDL Data Repository Extension](#)
  - 2.1 [MediaWiki Repository Manager](#)
  - 2.2 [NSDL MediaWiki Resource Manager](#)
- 3 [NSDL Search Popup](#)
- 4 [Third Party Plugins being used](#)
- 5 [OurNSDL skin](#)

### MediaWiki Version

OurNSDL is built using Ant and configuration files. It is based on MediaWiki 1.9.3. The source code is altered only under extreme circumstances where a bug fix is not available from MediaWiki.

### NSDL Data Repository Extension

#### MediaWiki Repository Manager

The [Special:RepositoryManager](#) is a "Special Page" that only sysops can view. It gives information about the status of the application in the NSDL Data Repository. If the application has not been initialized, the sysop can do so with this page. The initialization creates an aggregator and metadata provider for the application. The required inputs are: setSpec, setName.

#### NSDL MediaWiki Resource Manager

The Resource Manager is the "NSDL" tab present with all wiki articles. When viewed, it checks if the article exists in the NSDL Data Repository and presents the information in the repository if it does exist. If it does not exist and the user has the authority to do so, they can edit the metadata for this resource.



Navigation

- [Main Page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

Search




Toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)

- [Article](#)
- [Discussion](#)
- [View source](#)
- [History](#)
- [NSDL](#)

## NSDL interaction

**Contents** [\[hide\]](#)

- 1 [About This Resource](#)
- 2 [Referenced Resources](#)
- 3 [NSDL Context](#)

*Main Page* is a resource in the NSDL. This tab allows you to manage this resource as it appears in the library. I'm wondering how much stuff we would want to put on this page.. we could potentially make this like a "master" control panel that can do almost everything. Another thing to consider - Does this seem like a user-oriented feature, or a librarian/curator/developer oriented feature?

### About This Resource

**NSDL Resource URL** [http://our.nsd.org/index.php/Main\\_Page](http://our.nsd.org/index.php/Main_Page) [↗](#)

<b>Title</b>	Main Page
<b>Description</b>	The great main page
<b>Audience</b>	K-12 Teachers Students

### Referenced Resources

This section lists any resources used by this article, wither in reference, citation, inclusion, etc. Is this feature useful to have here?

- (add to NSDL) <http://nsdl.org> [↗](#)
- (add to NSDL) <http://nsdl.org/about/index.php> [↗](#)
- (add to NSDL) <http://nsdl.org/about/contactus/> [↗](#)
- NSDL (edit) <http://matdl.org/matdlwiki> [↗](#)



#### Navigation

- [Main Page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

#### Search

#### Toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)

[Article](#) [Discussion](#) [Edit](#) [History](#) [Protect](#) [Delete](#) [Move](#) [Watch](#) [NSDL](#)

## NSDL interaction

### Editing

We are editing now!

Title	<input type="text" value="AboutOurNSDL"/>
Description	<input type="text" value="Great page!"/>
Resource Contributor	<input type="text" value="Dean"/>
Author	<input type="text" value="Elly"/>
Keywords	<input type="text" value="nsdl"/>
Audience	<input type="text" value="K-12 Teachers"/> <input type="text" value="Faculty"/> <input type="text" value="Librarians"/> <input type="text" value="NSDL Community"/> <input type="text" value="Students"/> <input type="text" value="Informal Learners"/>
Education Level	<input type="text" value="Grades PreK to 12"/> <input type="text" value="Elementary School"/> <input type="text" value="Middle School"/> <input type="text" value="High School"/> <input type="text" value="Postsecondary"/> <input type="text" value="Higher Education"/>



### navigation

- Main Page
- Community portal
- Current events
- Recent changes
- Random page
- Help
- Donations

### search




### toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

[article](#) [discussion](#) [edit](#) [history](#)

[Log in / create account](#)

## Main Page

### Soft Matter Wiki

[\[edit\]](#)

Soft materials are materials such as **polymers**, **biomolecules**, **liquid crystals**, **surfactants**, and **proteins** that are typically organic and can be melted and processed at moderate temperatures as compared with inorganic materials like **metals** and **ceramics**. Typically, soft materials have weak interactions among molecular or supramolecular components and are often either amorphous or can **self-assemble** from the liquid state. There are often many levels of complexity with heirarchical, supramolecular structures that can be cooperative and far from equilibrium. We are most often concerned with the structural arrangements, viscoelastic rheology, and/or mechanical behavior of these materials. Within these pages, you will find information pertinent to soft matter and nanomaterials, with a specific focus on computational methods and modeling.

[Soft Matter Wiki-Overview of Contents](#)



This page was last modified 17:20, 11 September 2006. This page has been accessed 857 times.  
 Content is available under Attribution-NonCommercial-ShareAlike 2.5. [Privacy policy](#)  
[About NSDL Materials Digital Library Soft Matter Wiki](#) [Disclaimers](#)





## MyNSDL: NDR-integrated tagging, bookmarking, and recommendation

- Based on Connotea open-source folksonomic tagging/bookmarking system
- Tags and bookmarking structure are reflected back into the NDR
- Authorized users can “automatically” recommend new NSDL resources simply by tagging them
- Gives user a personal view of NSDL resources
- Planned for late 2007



Search

My library



Find results



Organize. Share. Discover.

you are logged in as  
deanbkrafft

My library

Log out

- Home
- Latest News
- About Connotea
- Site Guide
- Community pages

deanbkrafft's tags:

By Usage    A - Z

[digital library](#)

[education](#)

[frogs](#)

[NSDL](#)

[science education](#)

deanbkrafft's bookmarks

EXPORT LIST

RSS

?

Create a [Profile](#) on the [Community Pages](#).

Number of bookmarks per page: 10 | 25 | 50 | 100

edit    delete

**[NPR - Radio Expeditions: Disappearing Frogs](#)**

[www.npr.org](#)

Posted by [deanbkrafft](#) to [frogs NSDL](#) on [Thu Apr 27 2006](#) at 17:54 UTC | [info](#)

edit    delete

**[Exploratorium: Frogs](#)**

[www.exploratorium.edu](#)

Posted by [deanbkrafft](#) to [frogs NSDL](#) on [Thu Apr 27 2006](#) at 17:53 UTC | [info](#)

edit    delete

**[Deformed Frogs in Minnesota - Minnesota Pollution Control Agency](#)**

[www.pca.state.mn.us](#)

Posted by [deanbkrafft](#) to [frogs NSDL](#) on [Thu Apr 27 2006](#) at 17:53 UTC | [info](#)

edit    delete

**[What Is a Digital Library Anyway? Beyond Search and Access in the NSDL](#)**

Carl Lagoze *et al.*

*D-Lib Magazine* **11** (11), (Nov 2005)

[doi:10.1045/november2005-lagoze](#)

DLib paper describing how NSDL creates context and enrichment for digital library resources

Posted by [deanbkrafft](#) (who is an author) and [8 others](#) to [digital library NSDL](#) on [Thu Apr 27 2006](#) at 17:31 UTC | [info](#)

edit    delete

**[NSDL - The National Science Digital Library](#)**

Toolbox

[Add a bookmark](#)

[Create a new group](#)

[Create a tag note](#)

[Rename a tag](#)

[Import from local file](#)

[Export my library](#)

[Report a problem](#)

Related tags:

[digital libraries](#)

[National Science Digital Library](#)

[libraries](#)

[library](#)

[digital](#)

[science](#)

[metadata](#)

[soasym2005](#)

[engineering](#)

[mathematics](#)

[technology](#)

[digital library](#)

[research](#)

[BioEd](#)

[Archives - National](#)

[case studies](#)

[metasearch](#)

- Developed by DLESE from DCS
- Allows creation and editing of collection and item metadata records
- Extensive guidance and help for various categories of metadata
- Syncs records using the NDR API
- First prototype developed
- Released version expected 2Q07

Search for:  term  id  url

- Collection
- Last Editor
- Metadata Format
- Validity
- Status

Results per page

10

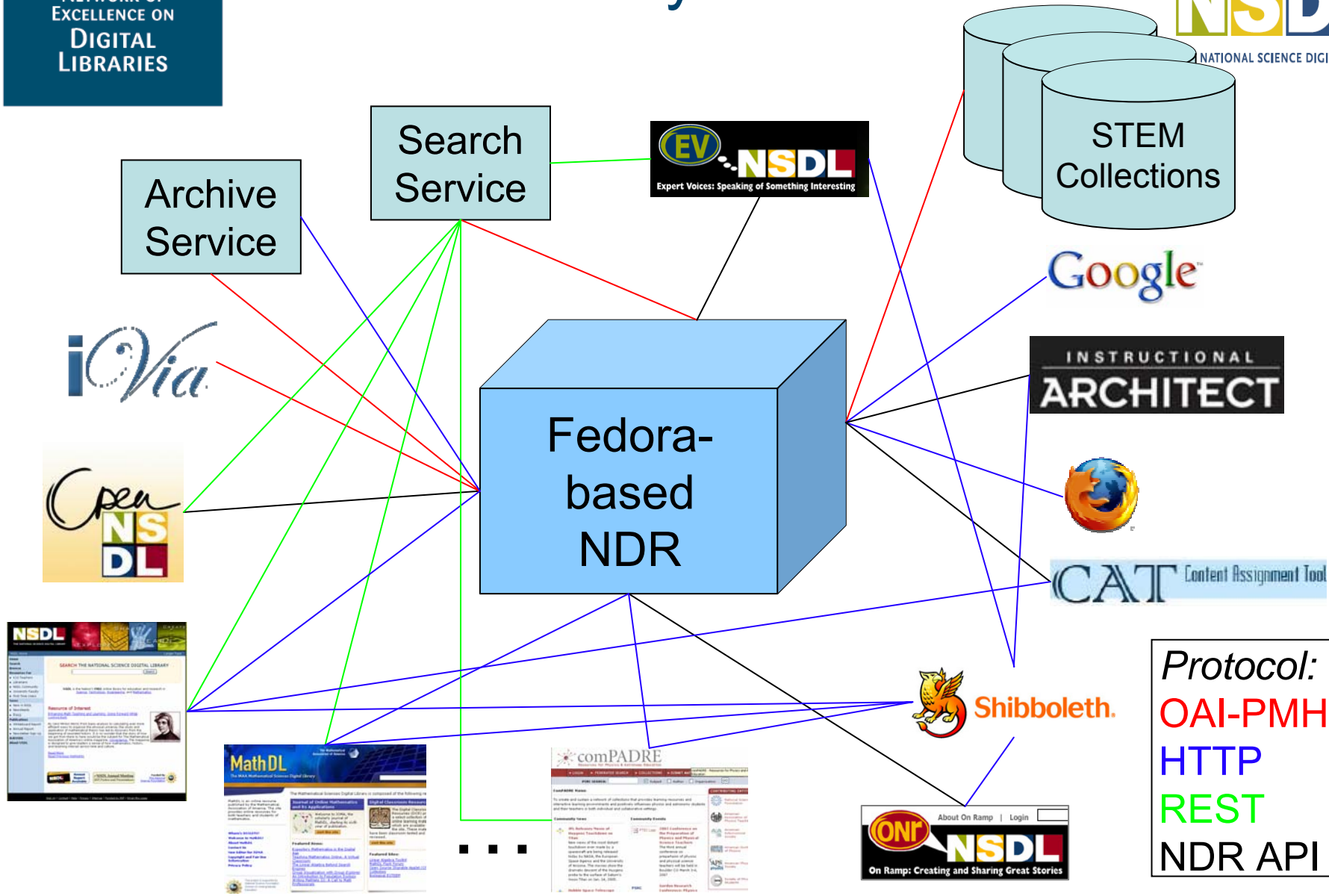
Your selections: **Collection: NCS Demo Collection**

**Last Editor: all + Format: all + Validity: all + Status: all**

Your search had 8 matches.

Record ID <input type="checkbox"/>	1 - 8 out of 8	Last Editor	Status	Last Touch
<a href="#">NSDL-000-000-000-015</a>	[ <a href="#">Validate Record</a>   <a href="#">View XML</a> ]	Unknown	<a href="#">New</a>	2006-12-12 10:28 AM
<b>Expert Voices Test</b> <a href="http://expertvoices.local.net">http://expertvoices.local.net</a> Collection: NCS Demo Collection Record format: nsdl_ncs File location: /ldevel/preview/ostwald/records/nsdl_ncs/1165611334360/NSDL-000-000-000-015.xml				
<input type="button" value="view"/> <input type="button" value="edit"/> <input type="button" value="copy"/> <input type="button" value="move"/> <input type="button" value="delete"/>				
<input checked="" type="checkbox"/> <b>Status Note</b> [ <a href="#">edit</a> ]				
<a href="#">NSDL-000-000-000-014</a>	[ <a href="#">Validate Record</a>   <a href="#">View XML</a> ]	Unknown	<a href="#">New</a>	2006-12-12 10:11 AM
<b>Play Record by Katy</b> <a href="http://www.comet.ucar.edu/index.html">http://www.comet.ucar.edu/index.html</a> Collection: NCS Demo Collection Record format: nsdl_ncs File location: /ldevel/preview/ostwald/records/nsdl_ncs/1165611334360/NSDL-000-000-000-014.xml				
<input type="button" value="view"/> <input type="button" value="edit"/> <input type="button" value="copy"/> <input type="button" value="move"/> <input type="button" value="delete"/>				
<input checked="" type="checkbox"/> <b>Status Note</b> [ <a href="#">edit</a> ]				
<a href="#">NSDL-000-000-000-012</a>	[ <a href="#">Validate Record</a>   <a href="#">View XML</a> ]	Unknown	<a href="#">New</a>	2006-12-12 10:03 AM
<b>Stuff On My Cat</b> <a href="http://www.stuffonmycat.com/">http://www.stuffonmycat.com/</a> Collection: NCS Demo Collection Record format: nsdl_ncs File location: /ldevel/preview/ostwald/records/nsdl_ncs/1165611334360/NSDL-000-000-000-012.xml				
<input type="button" value="view"/> <input type="button" value="edit"/> <input type="button" value="copy"/> <input type="button" value="move"/> <input type="button" value="delete"/>				
<input checked="" type="checkbox"/> <b>Status Note</b> [ <a href="#">edit</a> ]				
<a href="#">NSDL-000-000-000-007</a>	[ <a href="#">Record is Valid</a>   <a href="#">View XML</a> ]	jonathan	<a href="#">Done</a>	2006-12-08 9:11 AM
<b>Demo Record 2</b> <a href="http://www.dlese.org/dds/services/index.jsp">http://www.dlese.org/dds/services/index.jsp</a>				
<input type="button" value="view"/> <input type="button" value="edit"/> <input type="button" value="copy"/> <input type="button" value="move"/> <input type="button" value="delete"/>				
<input checked="" type="checkbox"/> <b>Status Note</b> [ <a href="#">edit</a> ]				

# NSDL 2.0 Ecosystem



**Protocol:**  
 OAI-PMH  
 HTTP  
 REST  
 NDR API

What are the challenges in creating a collaborative digital library?

# Trust







navigation

- Main Page
- Community Portal
- Featured articles
- Current events
- Recent changes
- Random article
- Help
- Contact Wikipedia
- Donations

search




toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

in other languages

- 粵語
- العربية
- Български
- বাংলা
- Català

- 

[Sign in / create account](#)

*Your continued donations keep Wikipedia running!*

## Portal:Mathematics

From Wikipedia, the free encyclopedia

[Arts](#) | [Biography](#) | [Geography](#) | [History](#) | **[Mathematics](#)** | [Philosophy](#) | [Science](#) | [Society](#) | [Technology](#)

### The Mathematics Portal

[edit](#)



**Mathematics** is often defined as the study of quantity, structure (pattern), change, and space. Some people call mathematics the study of "figures and numbers", but this is an oversimplification. In the formalist view, it is the investigation of axiomatically defined abstract structures using logic and mathematical notation. In the realist view, it is the investigation of objects or concepts that exist independently of our reasoning about them. Other views are described in the philosophy of mathematics article. Due to its applicability in practically every scientific discipline, mathematics has been called "the language of science", "the language of the universe", and "the queen of sciences".

Currently, there are approximately 14790 mathematical articles in Wikipedia.

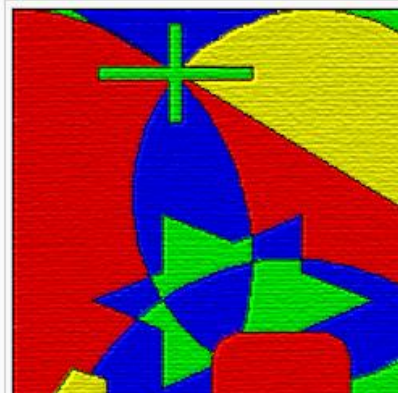
[Where to start](#) | [Did you know...](#) | [WikiProjects](#) | [Things you can do](#) | [Categories](#) | [Topics in mathematics](#)

### Selected article

[edit](#)

The **four color theorem** states that given any plane separated into regions, such as a political map of the counties of a state, the regions may be colored using no more than four colors in such a way that no two adjacent regions receive the same color. Two regions are called *adjacent* if they share a border segment, not just a point.

It is often the case that using only three colors is inadequate. This applies already to the map with one region surrounded by



### Where to start

[edit](#)

- [Mathematics](#)
- [History of mathematics](#)
- [Areas of mathematics](#)
- [Pure mathematics](#)
- [Applied mathematics](#)

### Did you know...

[edit](#)

- ...that the Klein Bottle gives a two-fold covering space of itself?
- ...that it is impossible to devise a single formula involving only polynomials and radicals for solving an



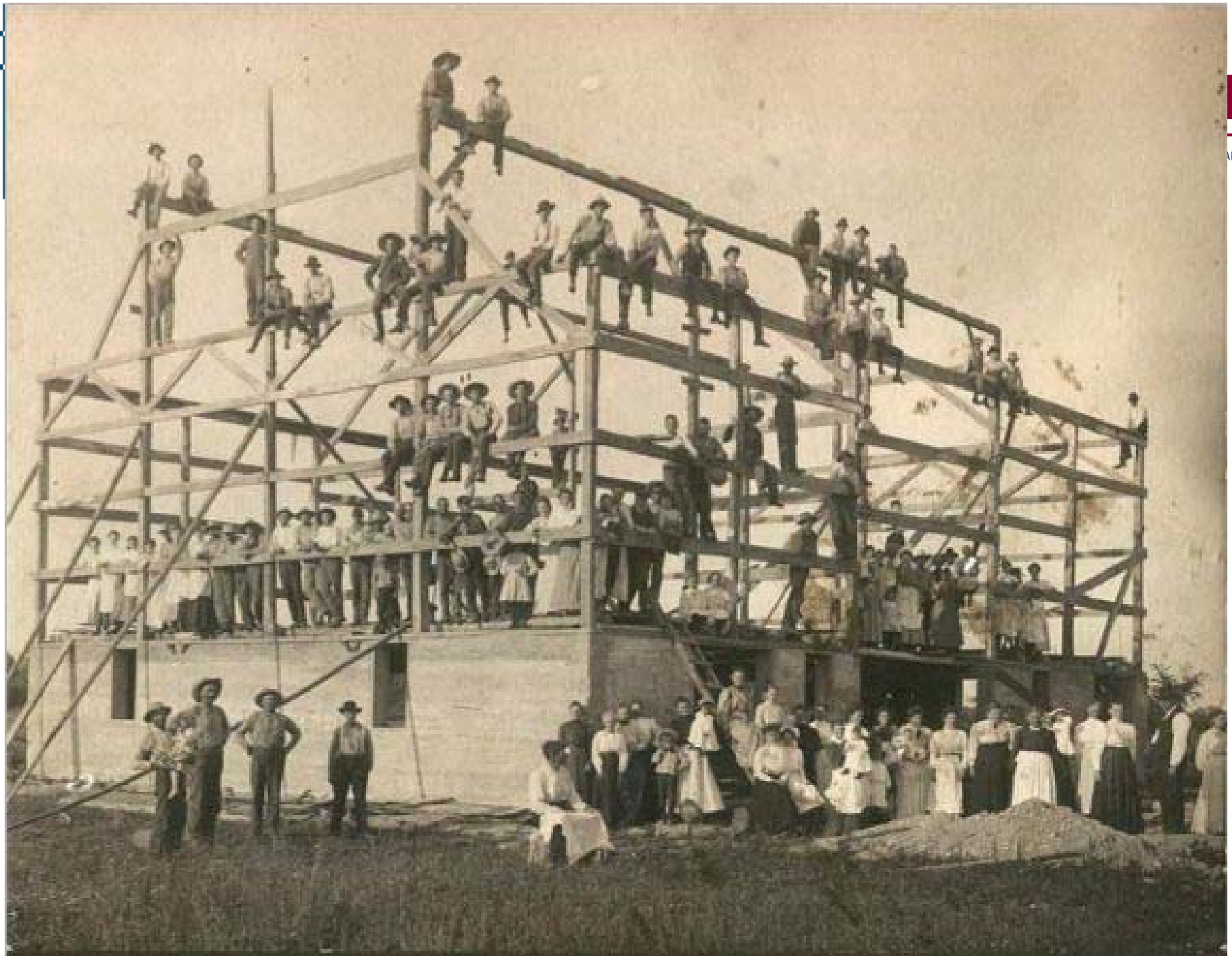






# Trust and reputation in a digital library

- Unlike Wikipedia, we typically want a digital library viewed as a source of “trusted” resources
- What is our trust mechanism?
  - Transitive trust approval
  - Community rating/filtering/reputation
- Trusted vs. complete “views”
- What is the right balance of trust vs. community contribution?



## Community Formation

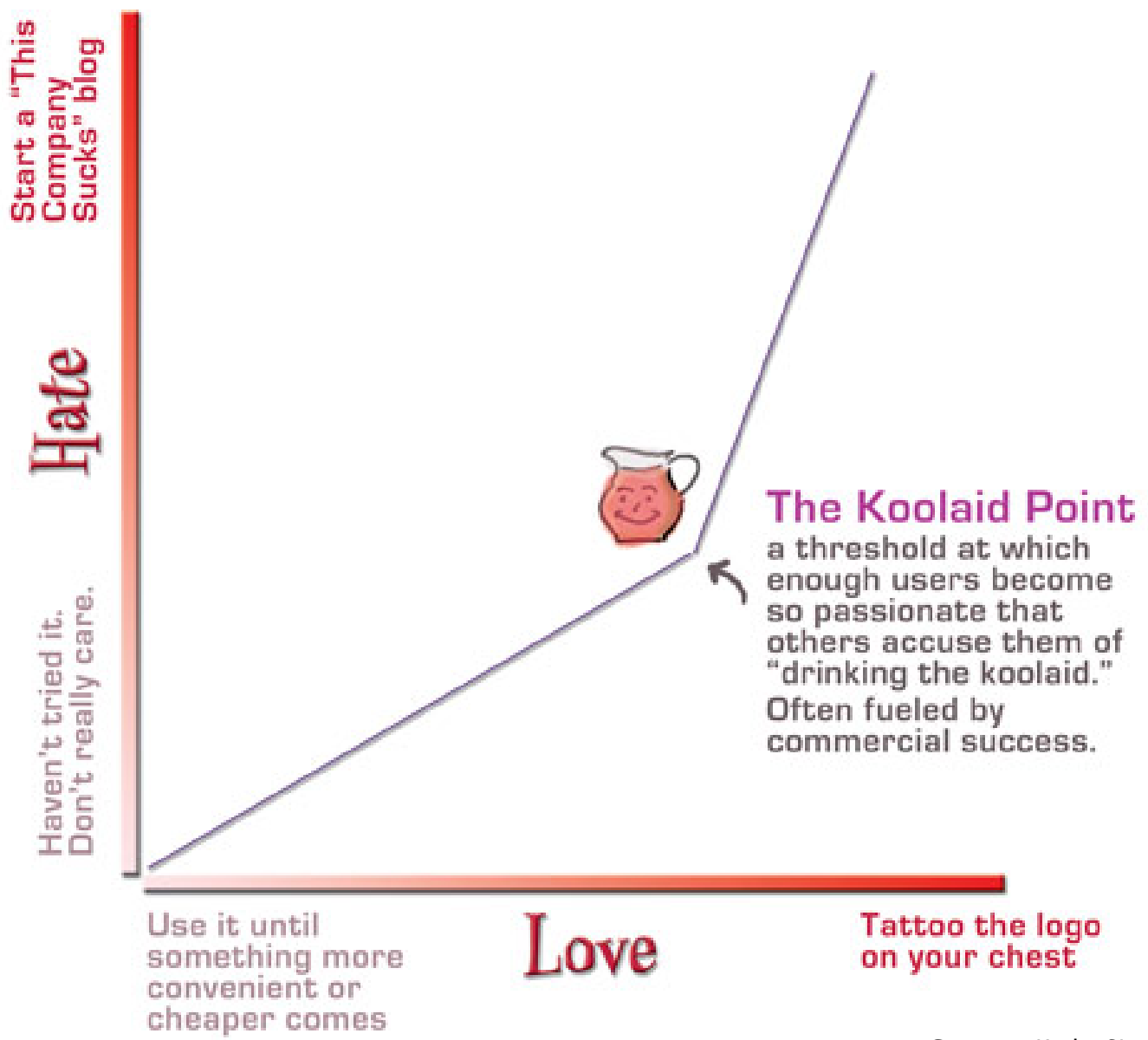
- Build the tools and they will come?
- What can we learn from Wikipedia, MySpace, Flickr, and YouTube?
- How do we leverage existing professional societies and groupings?
- For each digital library – is there one community, or are there many small communities?





Photo by Jon Crispin

# The Physics of Passion

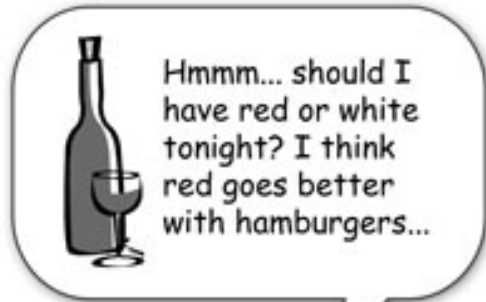


# Creating Passionate Users

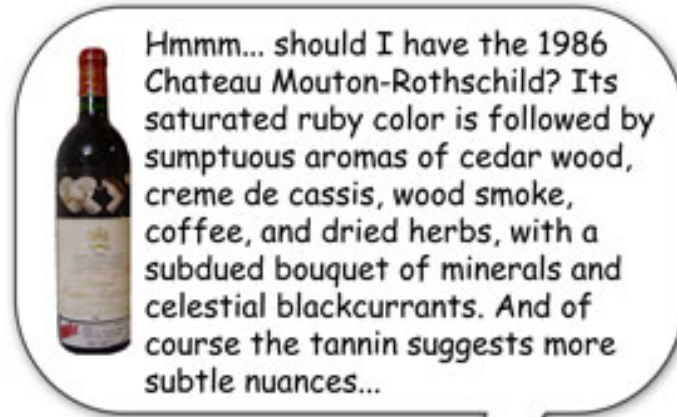
- How do we help digital library users “kick ass”?
- What can we learn from game design?
  - Motivating goal
  - Challenging interaction
  - Meaningful payoff
  - Multiple levels
- Can we use fun, emotion, seduction, surprise, and visuals – and still be academics?



## Learning increases resolution



**Before**



*After*



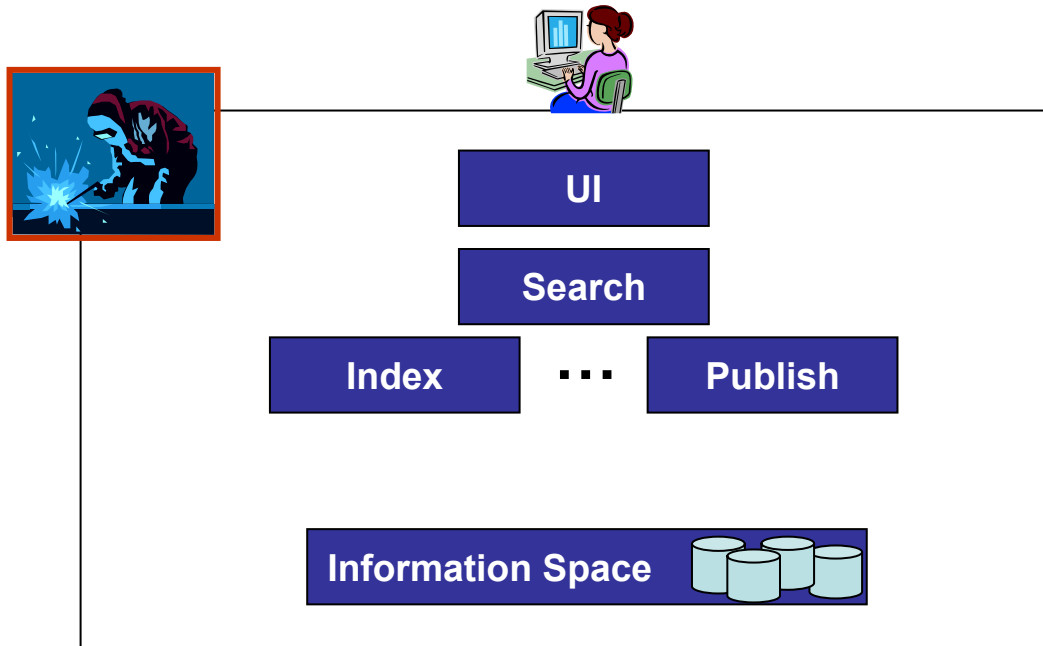
# PART III

## DLMSs and infrastructures

## The new demand for DLs

- Many of the organizations that demand a DL are small, distributed, and dynamic; they use the DL to support temporary activities such as courses, exhibitions, projects, etc.

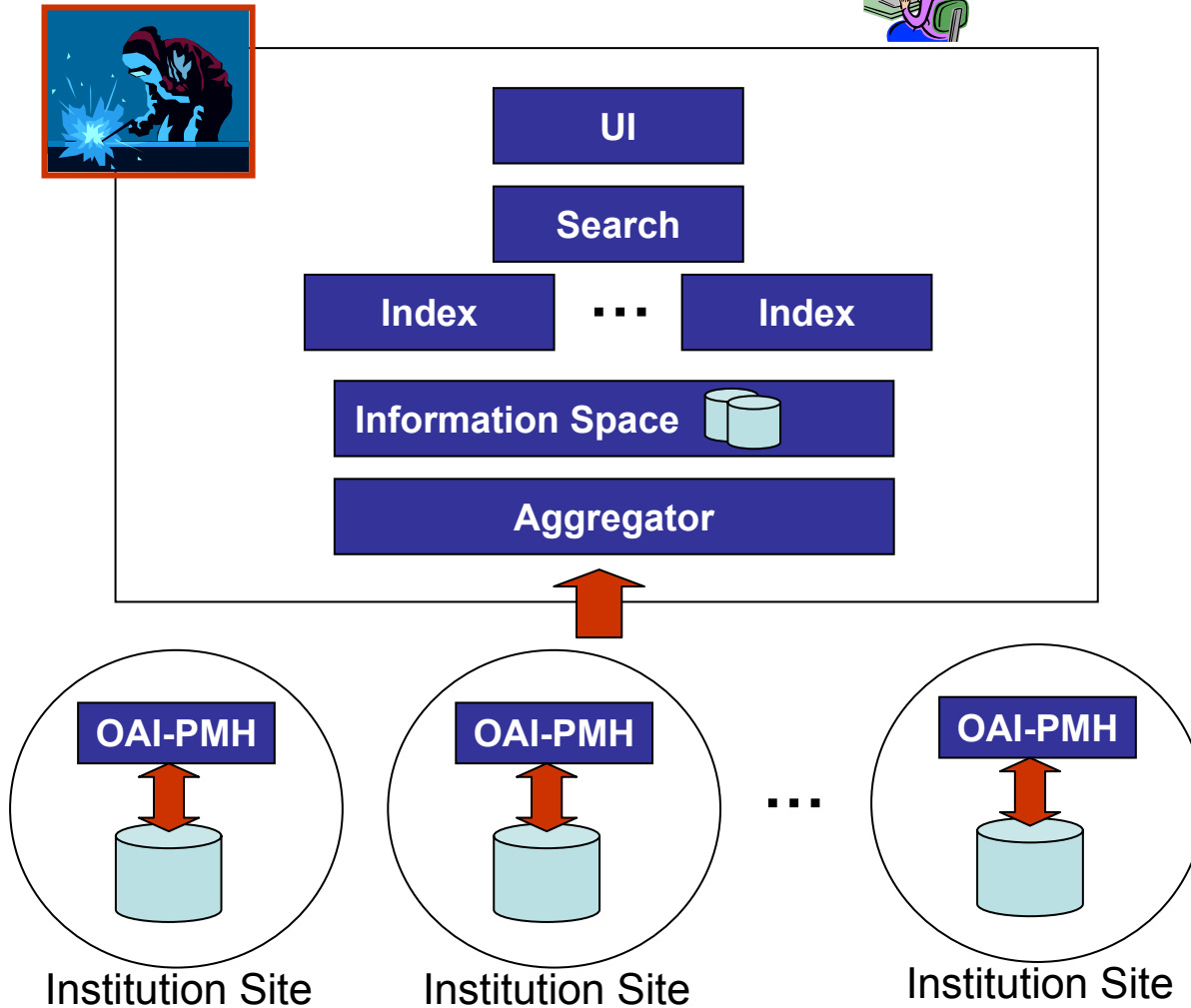
# DL Systems in the past



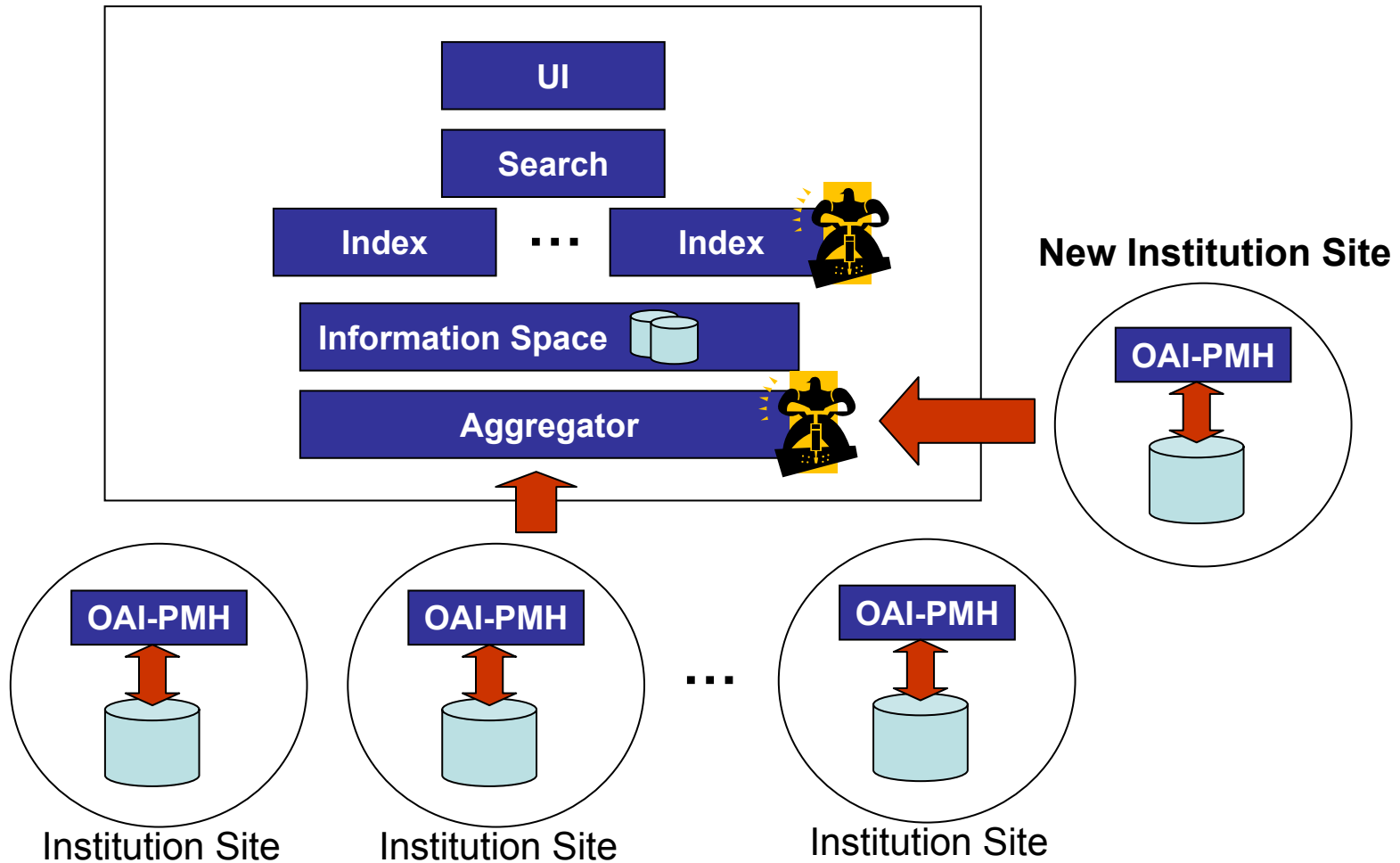
# Sustainability of the DL development models

- The construction and management of a DLS requires high investments and specialized personnel, content production is very expensive, multimedia and data handling requires high computational resources
- Years are spent in designing and setting up a DLS
- The systems lack interoperability and the services provided are difficult to reuse
- This development model is not suitable to satisfy the demand of many small organisations

# A solution: DL Systems – sharing of content

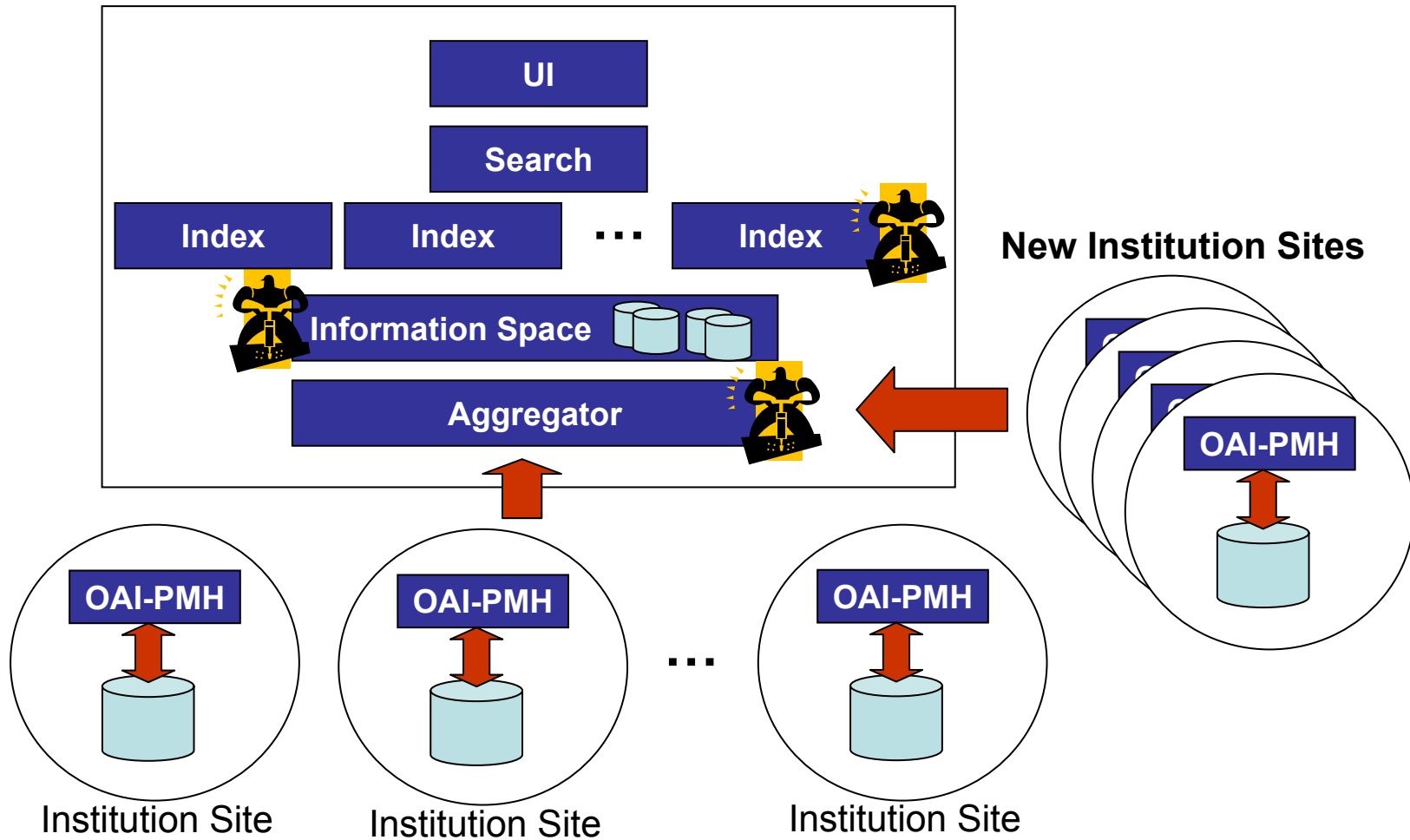


# Still high maintenance costs

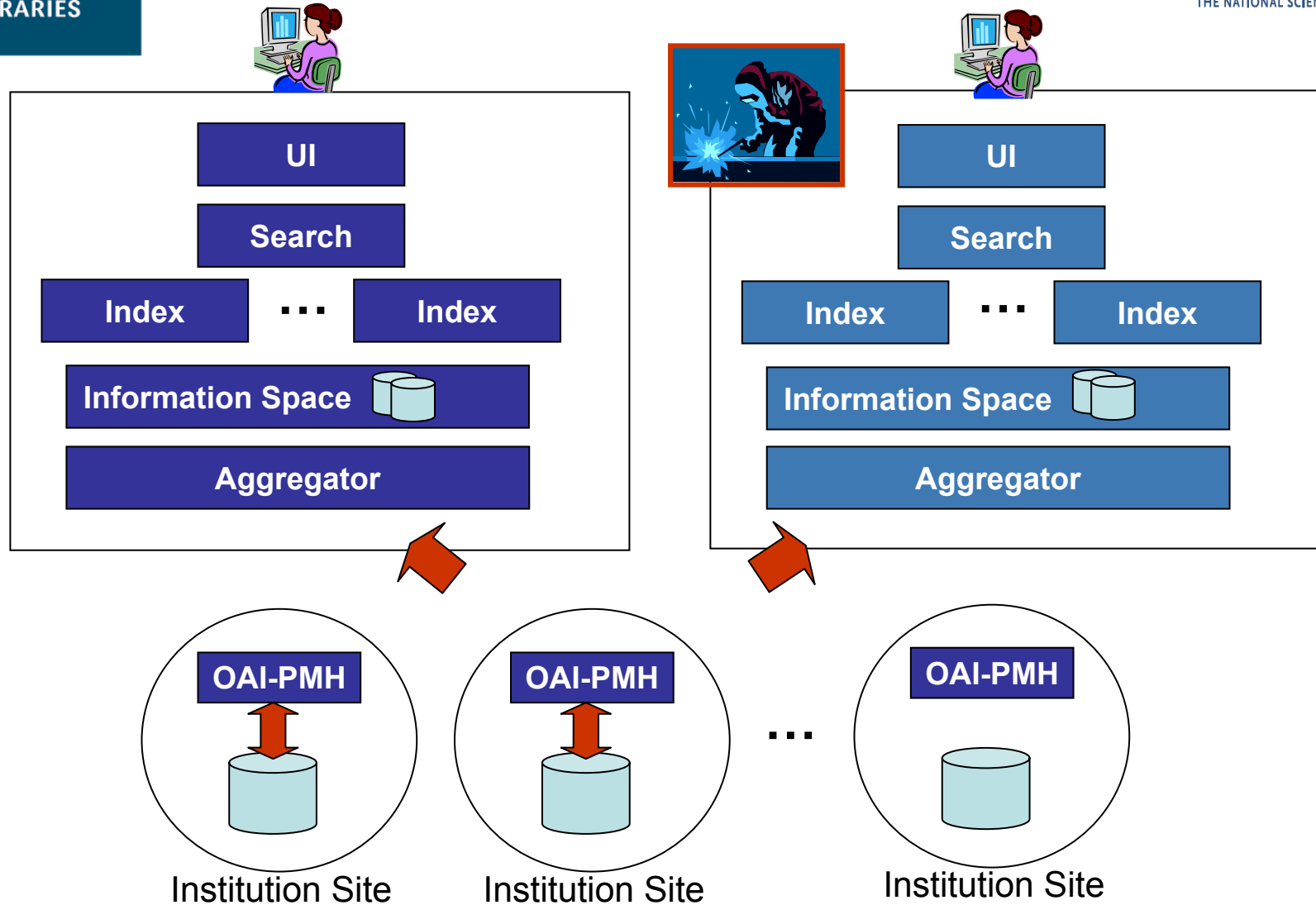




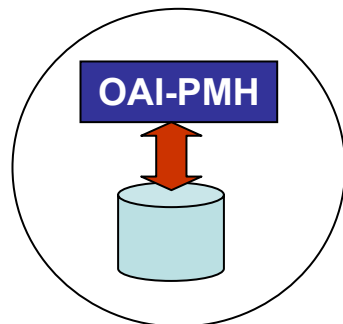
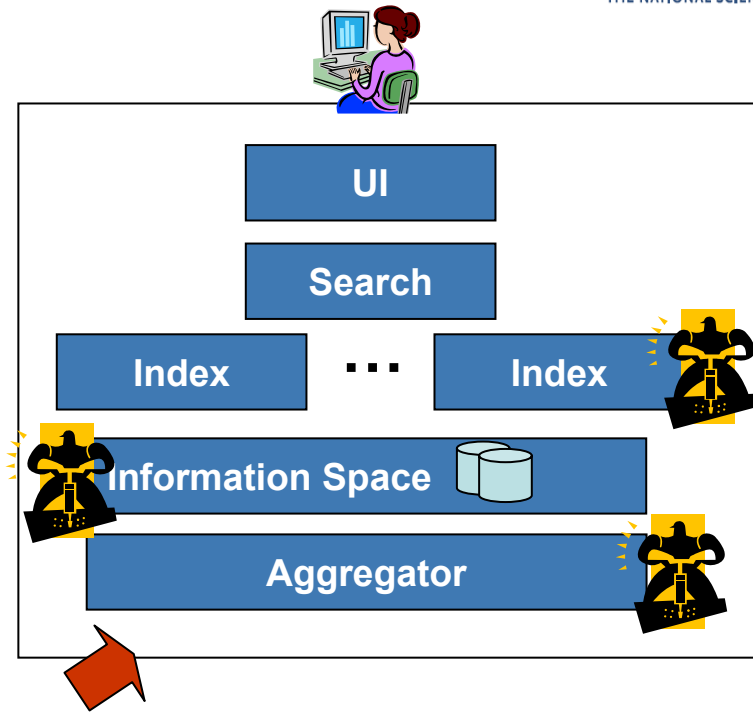
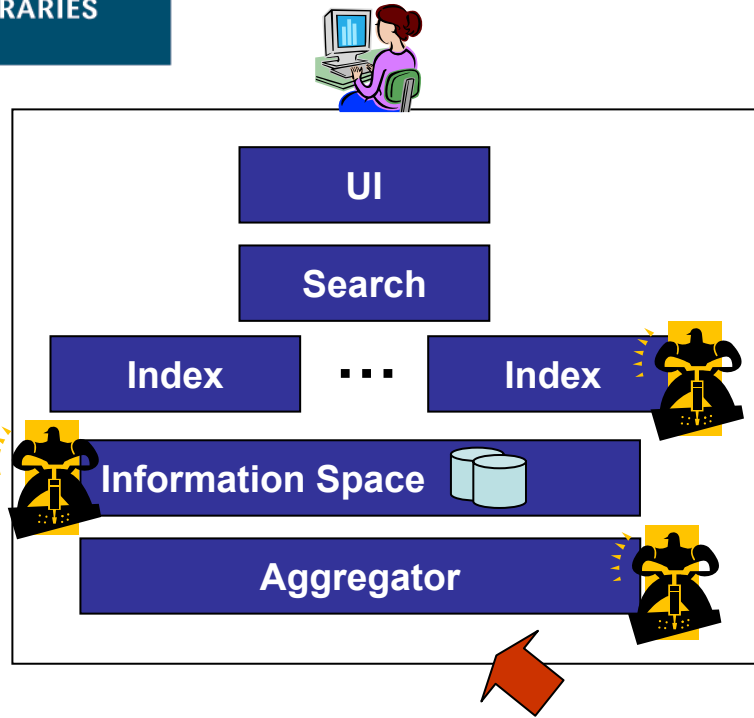
# Still hardly scalable



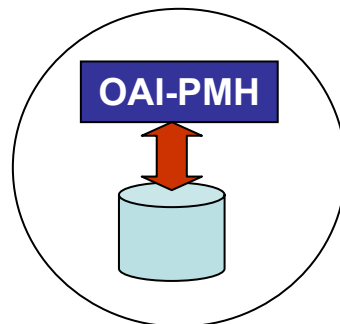
# Still not reusable



# Duplicate efforts

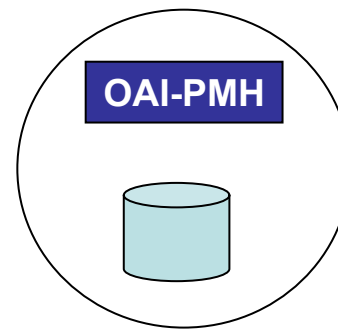


Institution Site

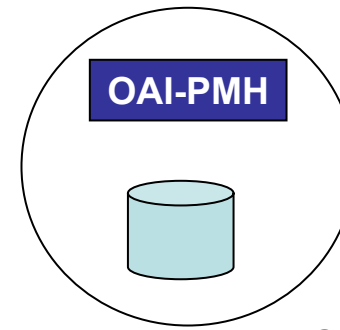


Institution Site

...



Institution Site



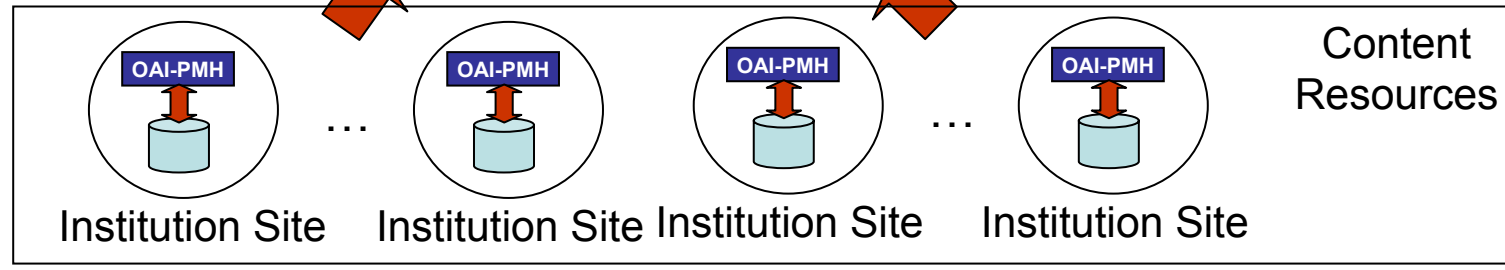
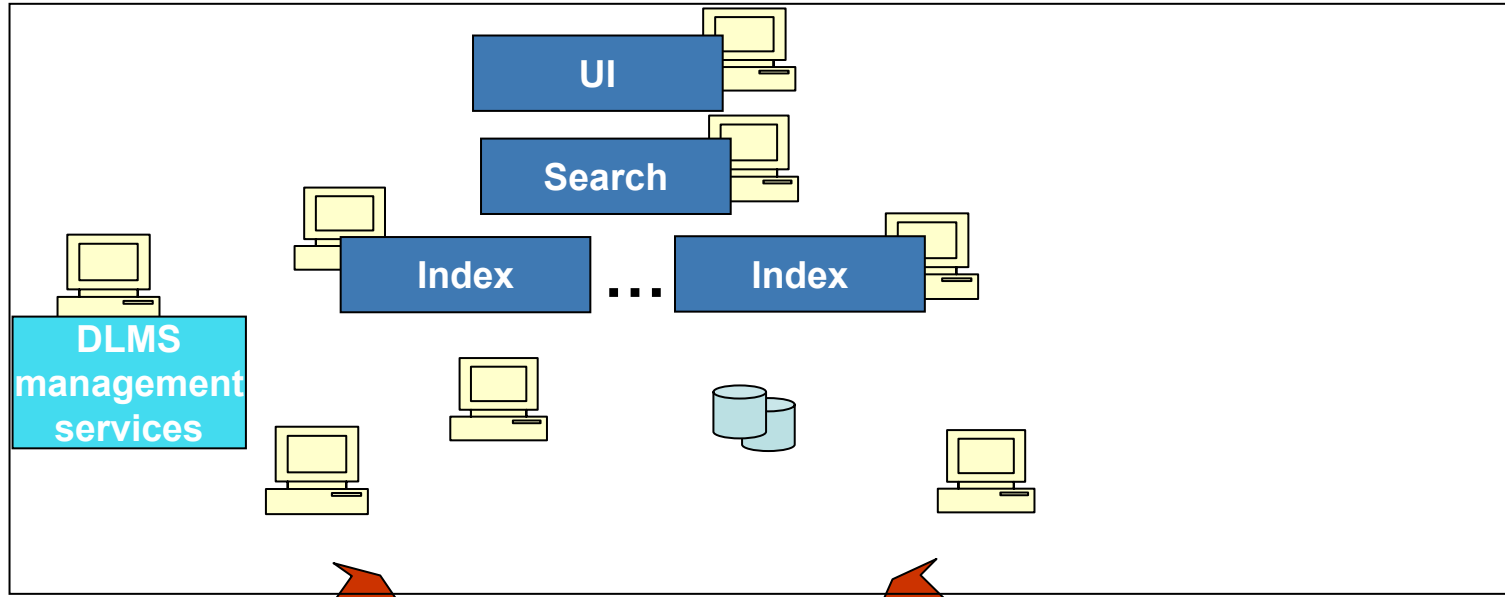
**New Institution Site**

# DLMS



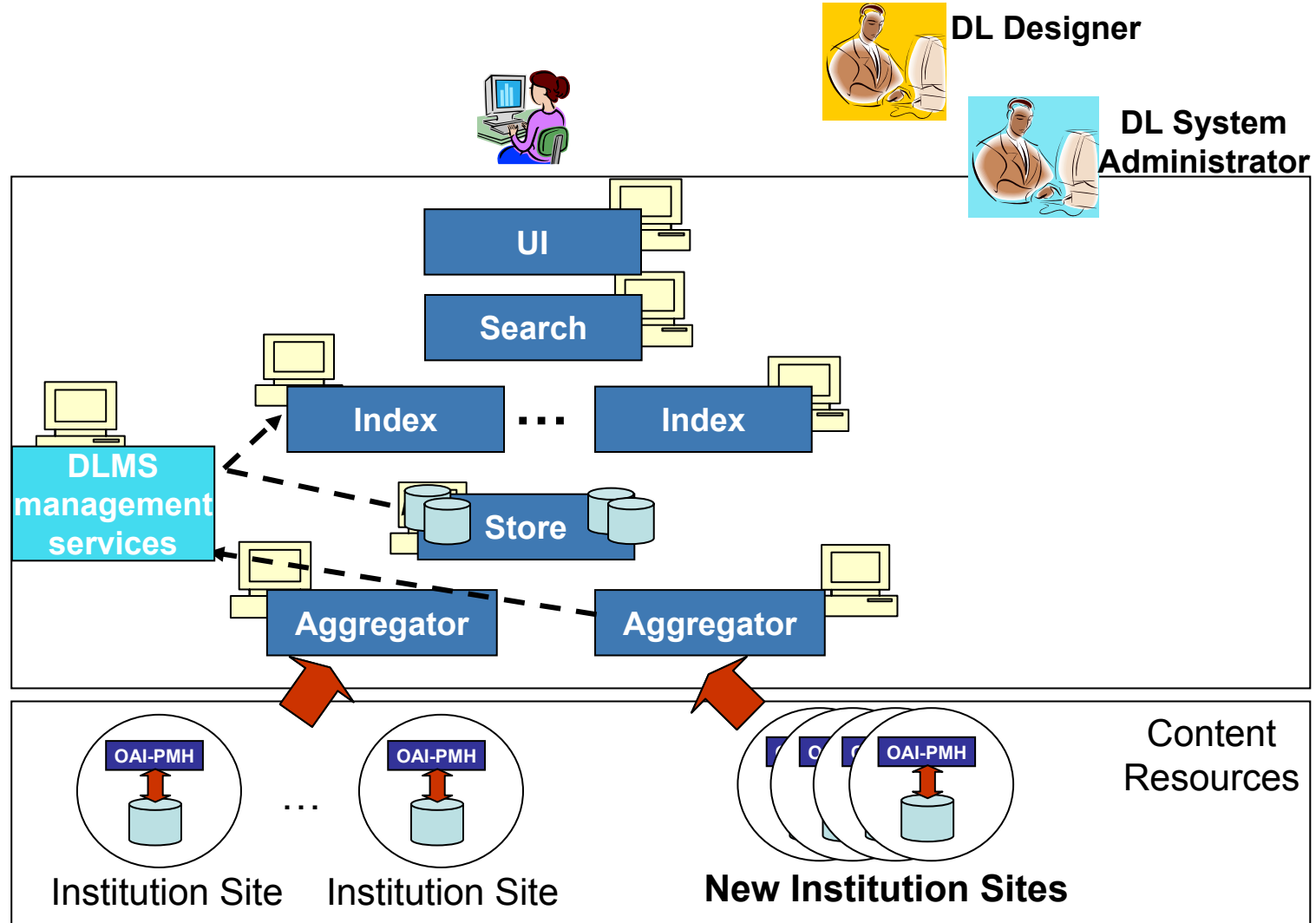
**Service Repository**

- UI
- Search
- Index
- Store
- Aggregator
- User Profiling
- ...
- Others

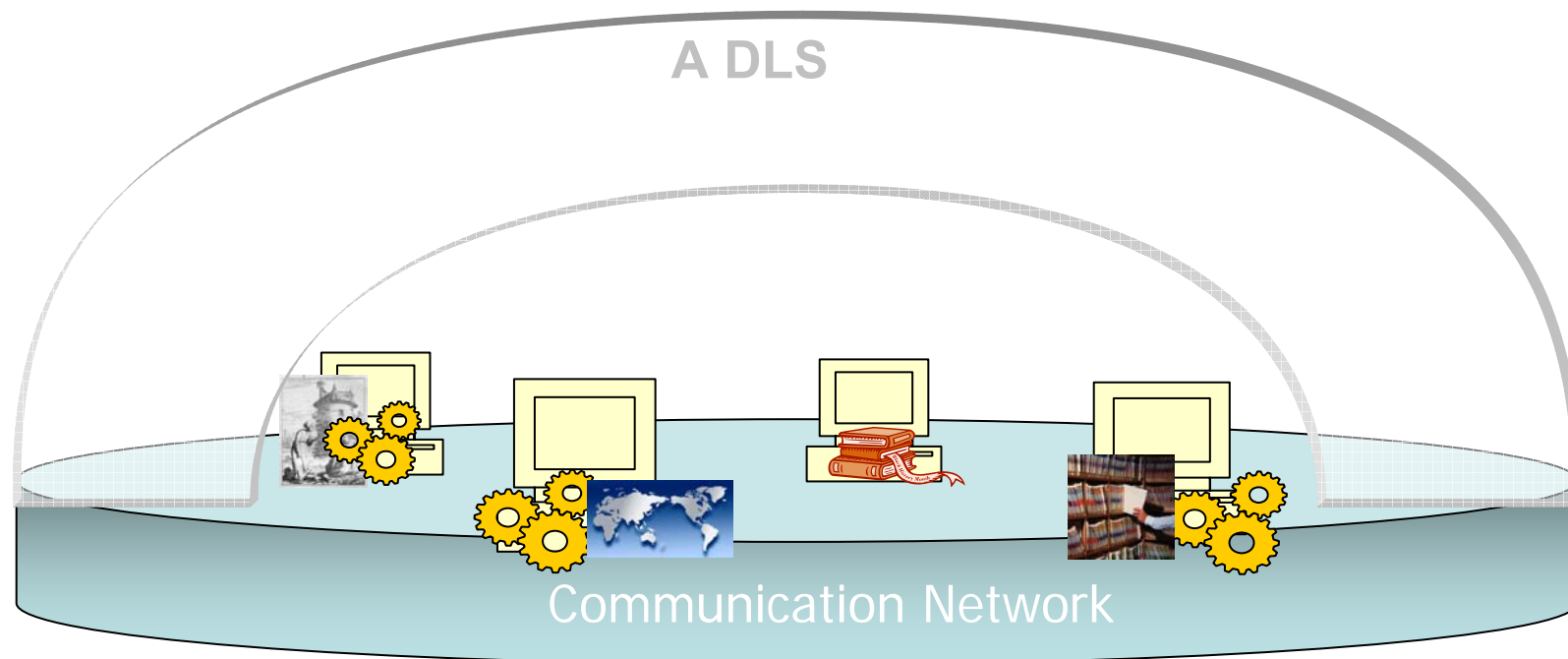


# Maintenance through the DLMS

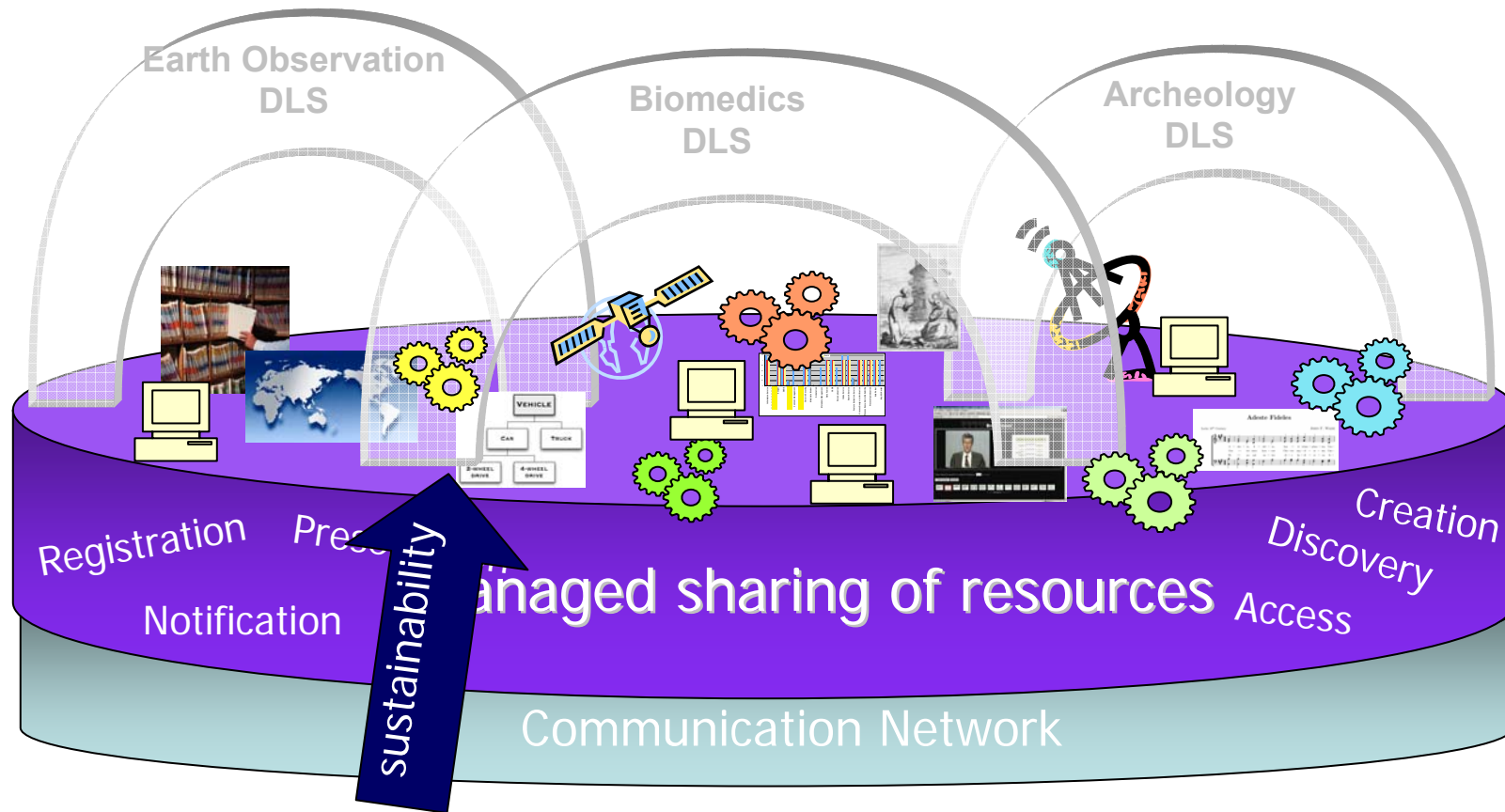
- Service Repository
- UI
  - Search
  - Index
  - Store
  - Aggregator
  - User Profiling
  - ⋮
  - Others



“The underlying foundation or basic framework (as of a system or organization)” *[Merriam-Webster]*



# Resource sharing infrastructure





# DRIVER



## Digital Repository Infrastructure Vision for European Research

<http://www.driver-repository.eu/>

- **PARTNERS**

- Univ.of Athens(GR)
- Univ. of Bielefeld (DE)
- CNR-ISTI (IT)
- STICHTING SURF (NL)
- Univ. of Nottingham (UK)
- CNRS-CCSD (FR)
- Univ. of Bath (UK)
- Univ. of Warszawski (PO)
- Univ. of Gent (BE)
- Univ. of Gottingen (GE)





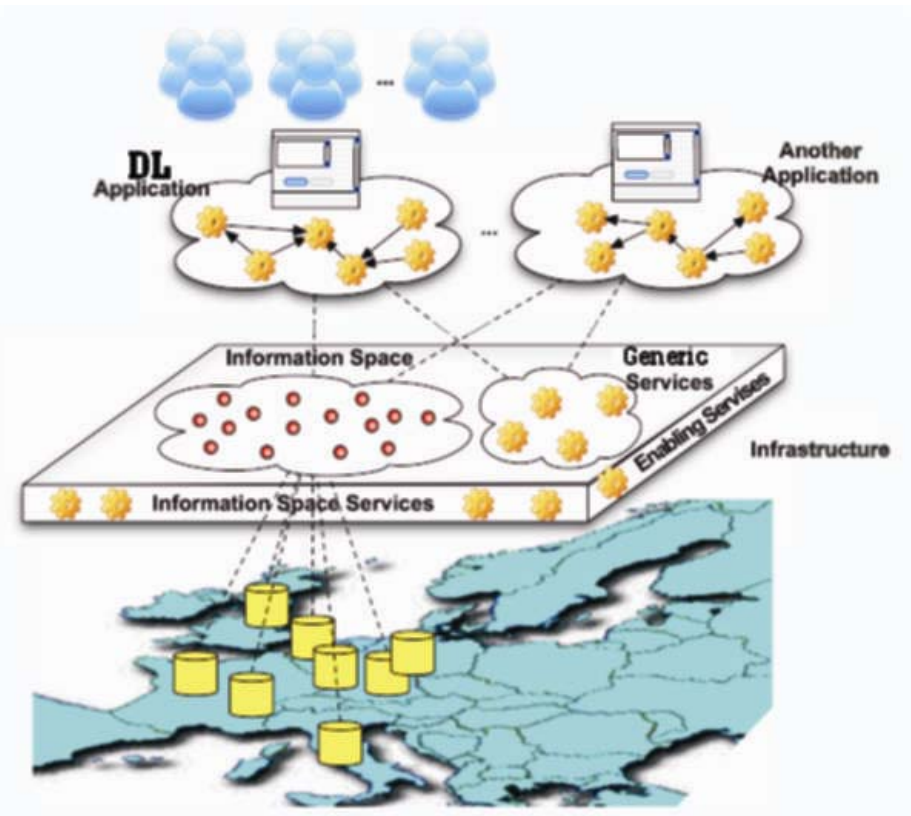
## DRIVER Objectives



- To develop a *test-bed* for integrating existing national, regional and thematic repositories in order to create a *production-quality European Repository Infrastructure*
- To identify and promote the use of a relevant set of standards
- To prepare the future expansion and upgrade of a Digital Repository Infrastructure across Europe and to ensure widest possible user involvement

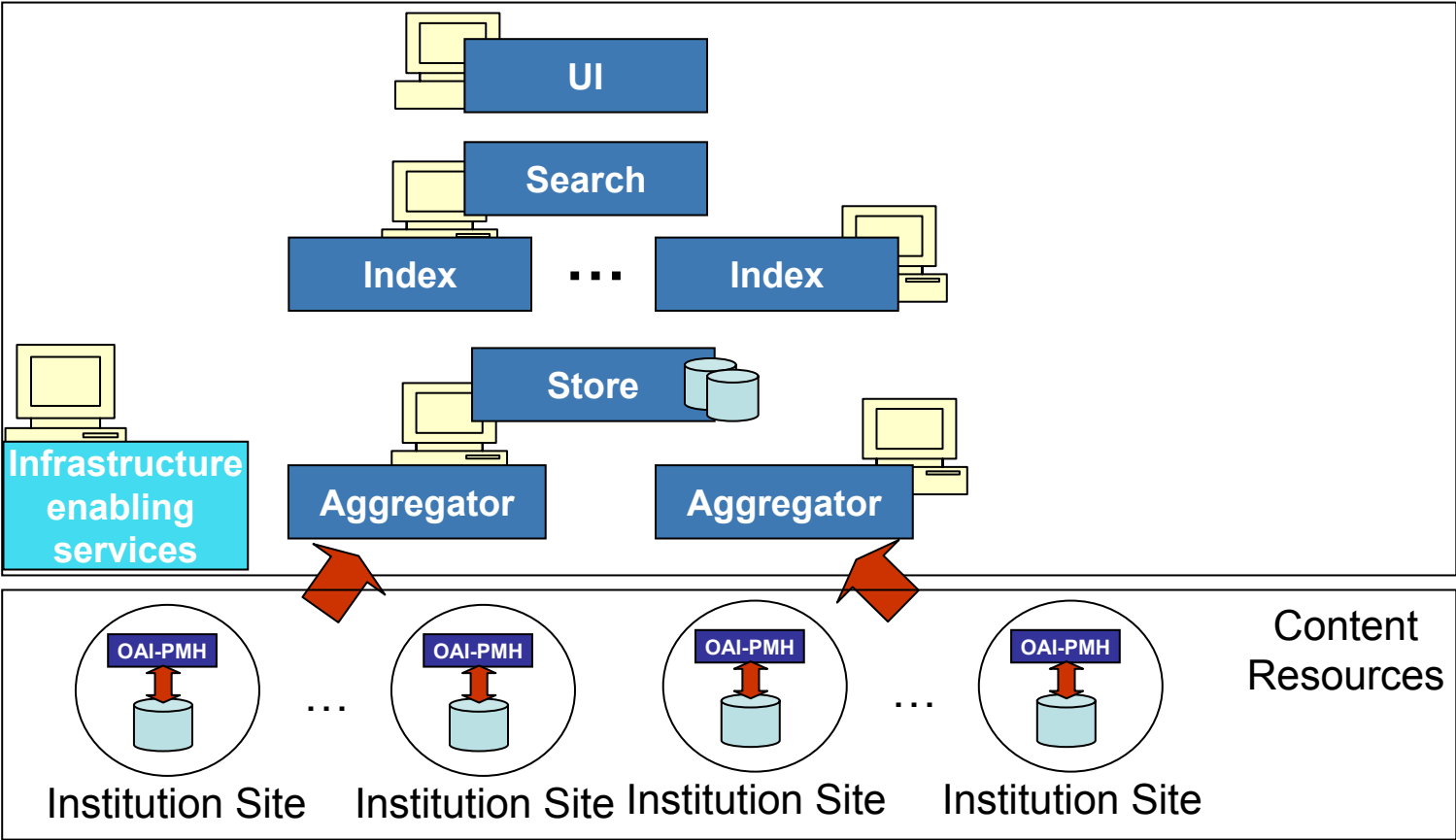
# DRIVER repositories

58 institutional repositories  
publicly accessible through  
the first DRIVER public release  
(June 2007)



# DRIVER Infrastructure

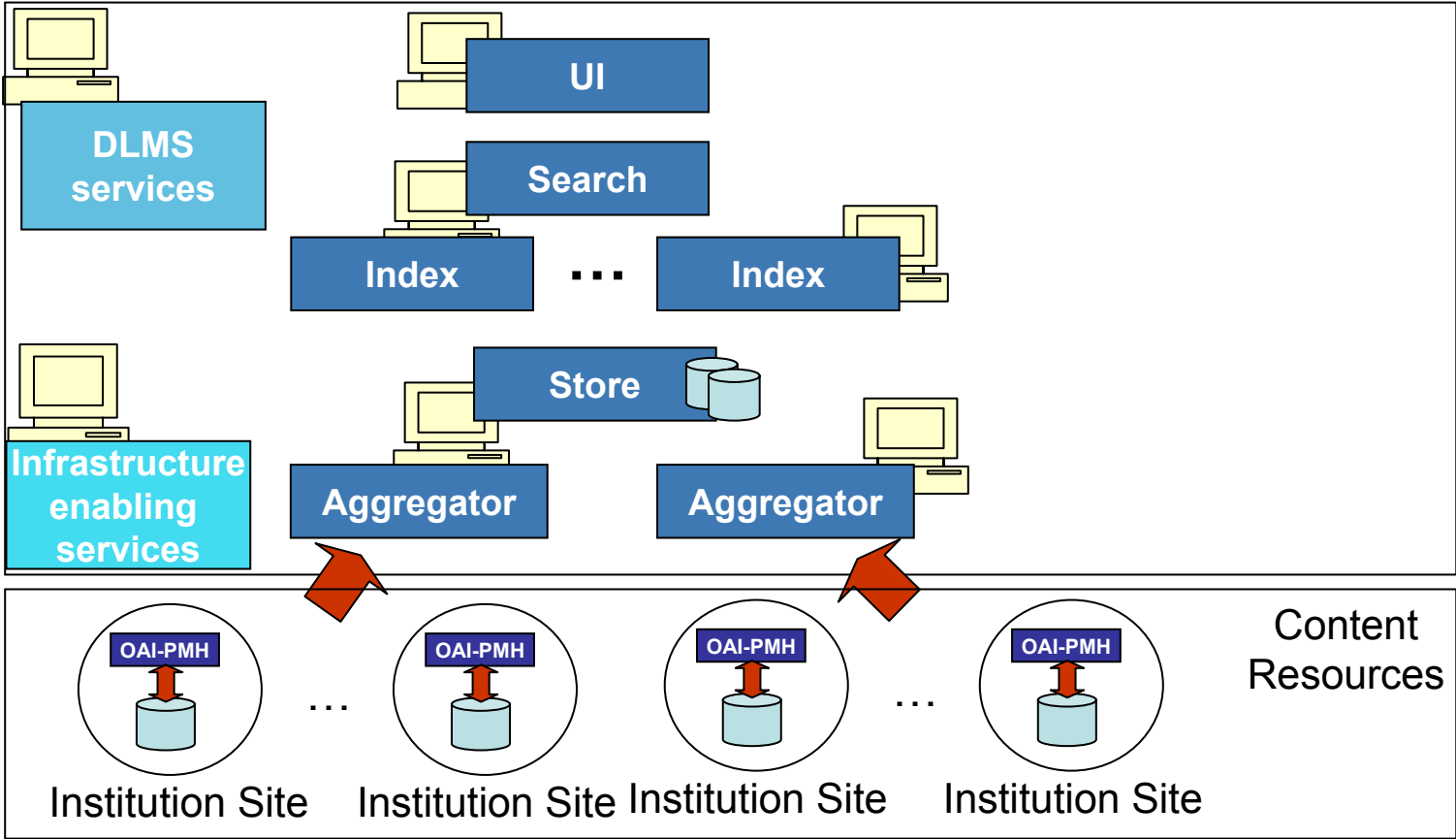
- Infrastructure enabling services
- Authentication
  - Directory
  - Res. validator
  - Notification
  - Monitoring
  - ⋮
  - Others



# DRIVER Infrastructure

**DLMS services**

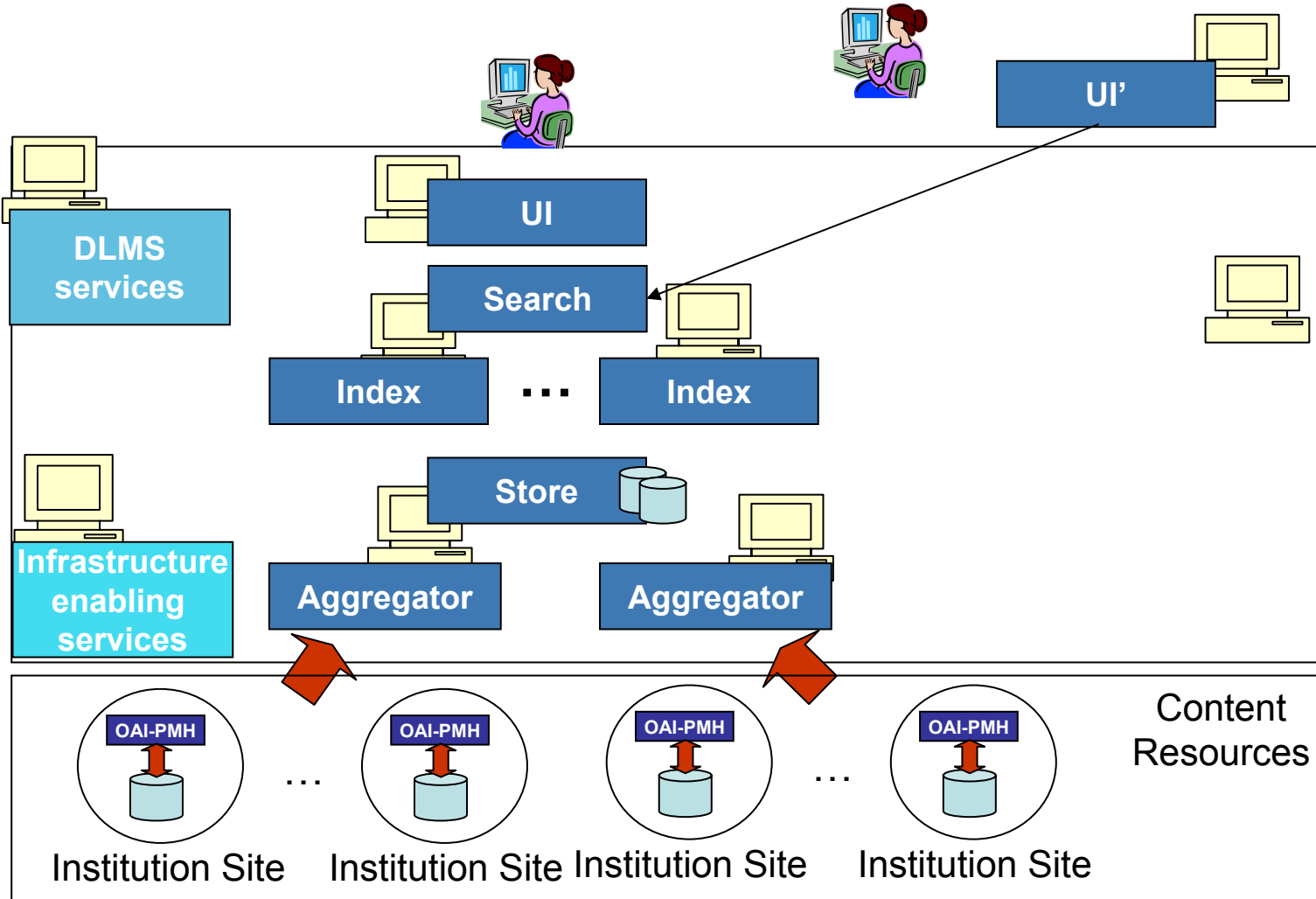
- Orchestration
- Configuration
- ⋮
- Others



# A step further in the reuse

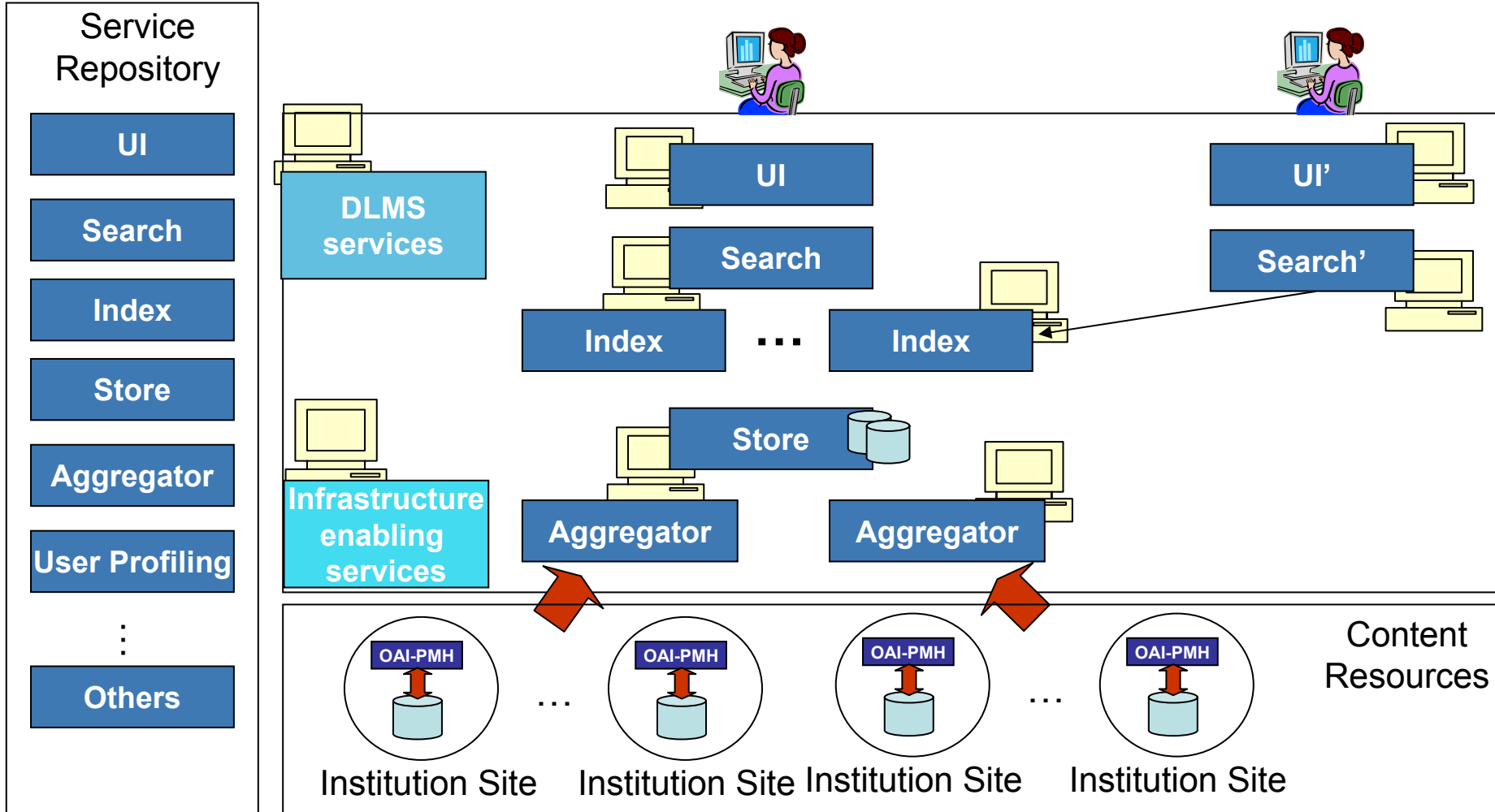
Functionality sharing #1

- Infrastructure enabling services
- Authentication
  - Directory
  - Res. validator
  - Notification
  - Monitoring
  - ⋮
  - Others



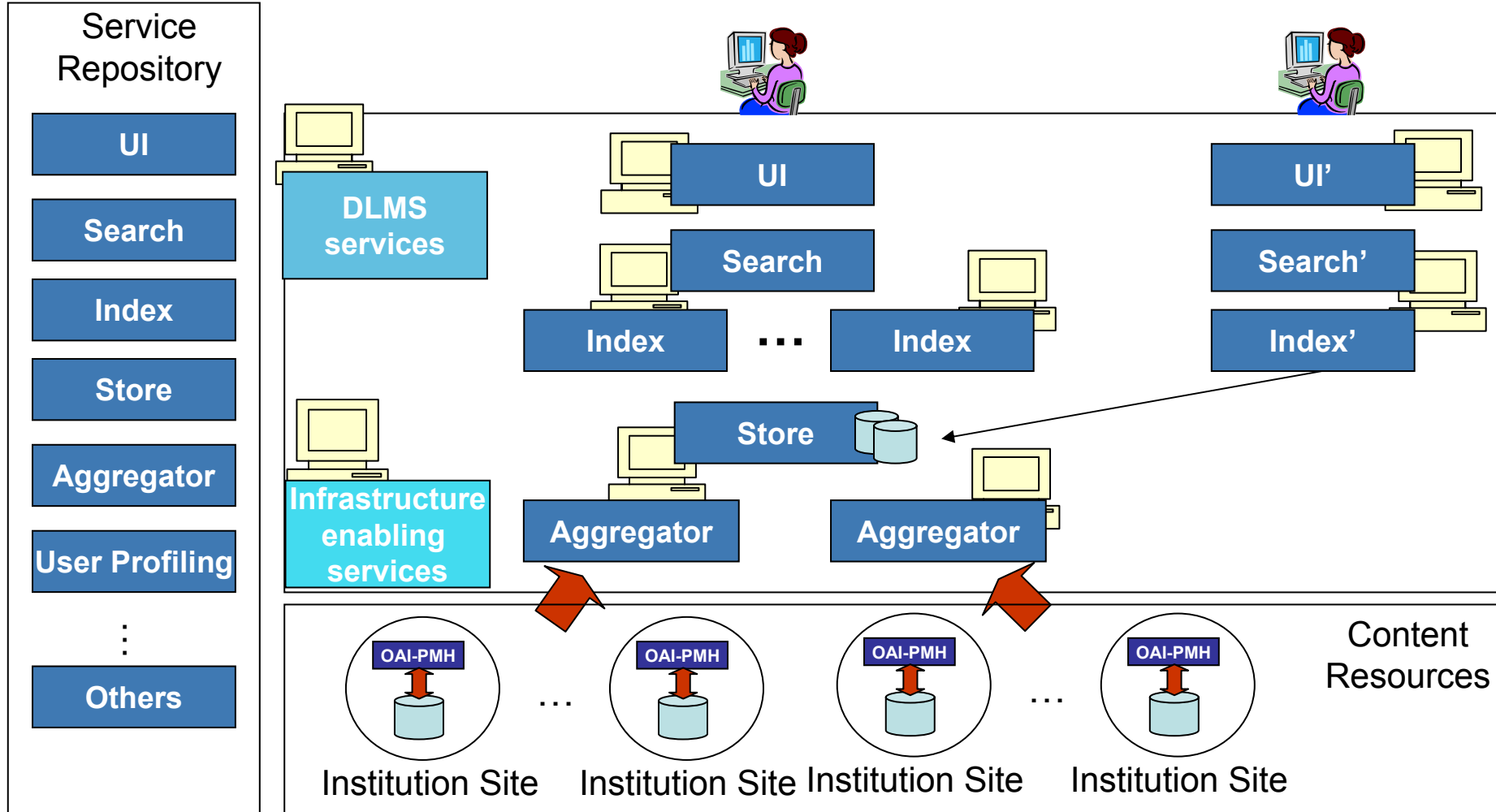
# A step further in the reuse

Functionality sharing #2



# A step further in the reuse

Functionality sharing #3



# DRIVER Demo





DILIGENT



## A Digital Library Infrastructure on Grid-Enabled Technology

<http://www.diligentproject.org>

### • PARTNERS

- ERCIM (FR)
- CNR-ISTI (IT)
- Univ. of Athens (GR)
- Univ. of Basel (SW)
- Engineering (IT)
- CERN (IT)
- Univ. of Strathclyde (UK)
- European Space Agency (IT)
- FAST (SW)
- RAI (IT)
- 4D-Soft (HU)



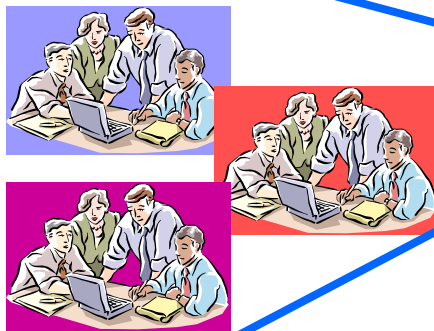
## Objective

To create an Infrastructure on Grid-ENabled Technology that allows members of dynamic **virtual organizations** to create on-demand transient virtual digital libraries based on **shared resources**, i.e. **processing and storage capabilities, multi-type content and applications**



# DILIGENT DL infrastructure

Consumers

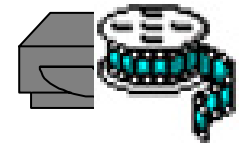


Infrastructure  
enabling  
services

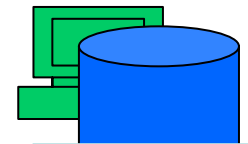
DLMS  
services



3D processing



simulation

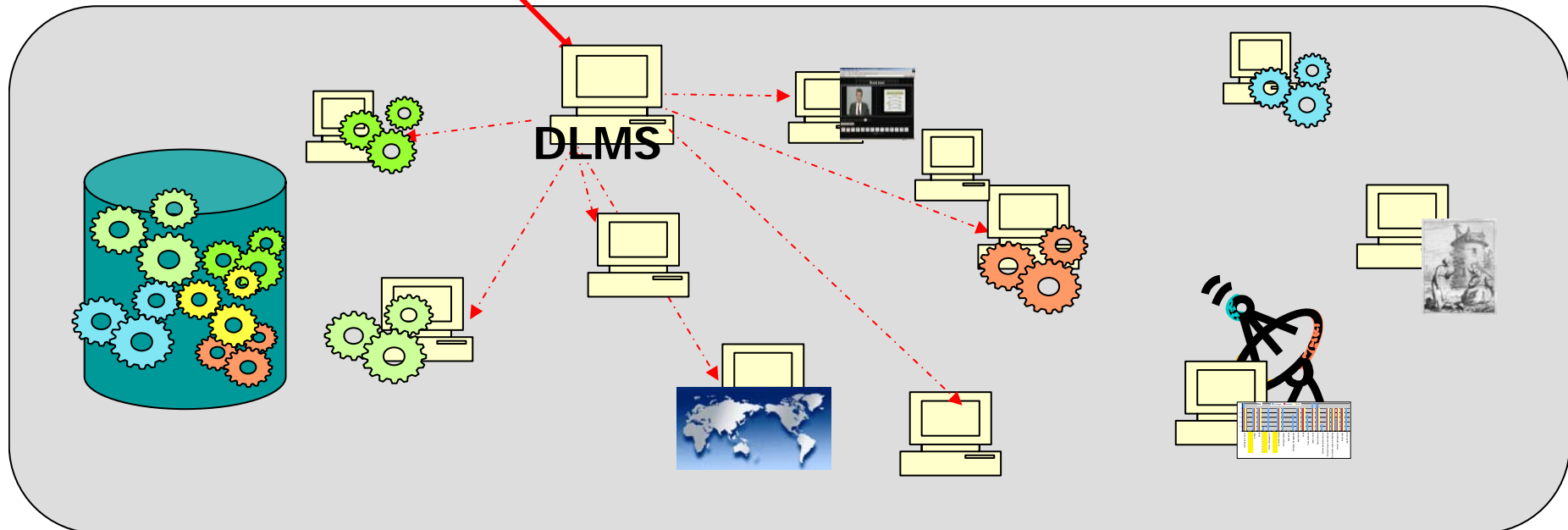


Feature  
extraction



Speech  
recognition

# DILIGENT Infrastructure



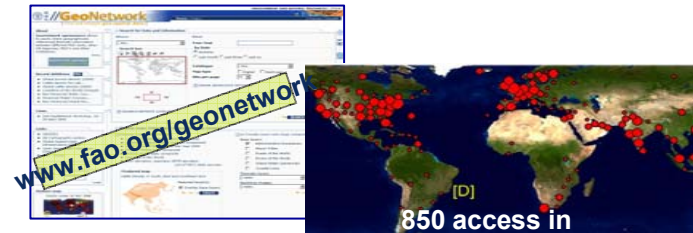
## DILIGENT and EGEE

- DILIGENT exploits the computation resources of the EGEE infrastructure
- **EGEE (Enabling Grid for E-scienceE)** Grid infrastructure consists of:
  - over 30,000 CPU available to users 24 hours a day, 7 days a week,
  - 5 Petabytes (5 million Gigabytes) of storage
- 90 institutions in 32 countries

## DILIGENT additional infrastructure features

- Resources are shared and used “on-demand”
- DL functionality whose implementation requires huge processing capabilities become “sustainable” e.g. many diverse virtual organizations may exploit them  
e.g. generation of products from a huge amount of observation data, generation of statistical reports from large sets of experimental data, similarity searches based on complex features extraction, video watermarking, ...

- User community represented by the European Space Agency



- Characteristics:
  - well-established tradition in exploitation of new technologies
  - wide variety of content types (maps, satellite images, etc.)
  - very large, dynamic data sets

- Earth Science Domain DLs

- preparation of periodical reports
- creation of decision supporting mechanisms in case of environmental accidents



# The ARTE Scenario

- User community represented by:
  - Scuola Normale Superiore
  - RAI Radiotelevisione Italiana
- Characteristics
  - IT technology exploitation still in infancy
  - multidisciplinary collaborative research
  - image based retrieval/semantic analysis of images
- Teaching and e-Learning in Humanities
  - organization of courses
  - support multidisciplinary research revolving around images





# DILIGENT Demo

## Concluding Remarks

- The notion of DL is changing ....  
*“The potential exists for digital libraries to become the Universal knowledge repositories and communication conduits of the future, a common vehicle by which everyone, will access, discuss, evaluate, and enhance information in all form”*
- DLs are exploited as tools in many different application areas
- The technology must appropriately evolve for supporting the new notion of DL

# Acknowledgements

- NSDL NSF Program Officers
  - Lee Zia
  - David McArthur
- NSDL Core Integration Team
  - UCAR: Kaye Howe, PI and Executive Director
  - Cornell: Dean Krafft, PI
  - Columbia: Kate Wittenberg, PI
- Fedora Development Team
  - Cornell: Sandy Payette & Carl Lagoze
  - Univ. of Virginia: Thornton Staples

• DRIVER Team



• DILIGENT Team



## Contact Information

- Dean B. Krafft
- Cornell Information Science
- 301 College Ave.
- Ithaca, NY 14850
- USA
- [dean@cs.cornell.edu](mailto:dean@cs.cornell.edu)
- Donatella Castelli
- ISTI-CNR
- Via Moruzzi, 1
- 57128 Pisa
- Italy
- [donatella.castelli@isti.cnr.it](mailto:donatella.castelli@isti.cnr.it)



This work is licensed under the Creative Commons Attribution-ShareAlike 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA. When separated from this work, some images may be covered by separate copyright or license terms.