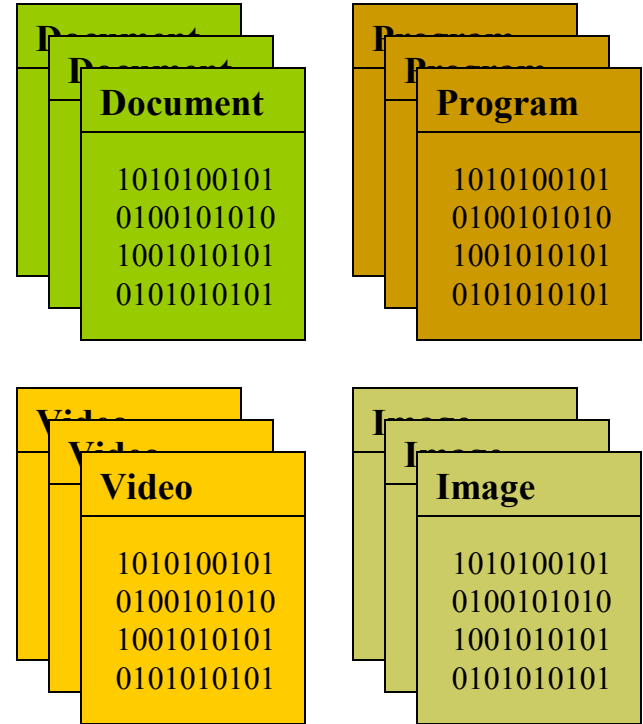
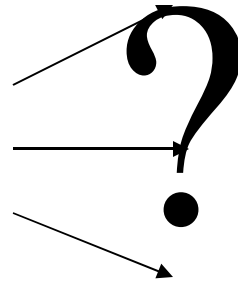
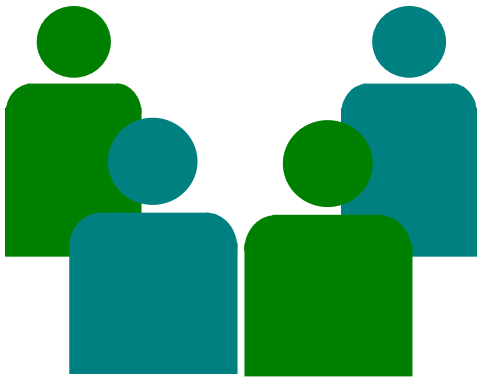


Digital Library Component Models



hussein suleman
uct cs honours 2005

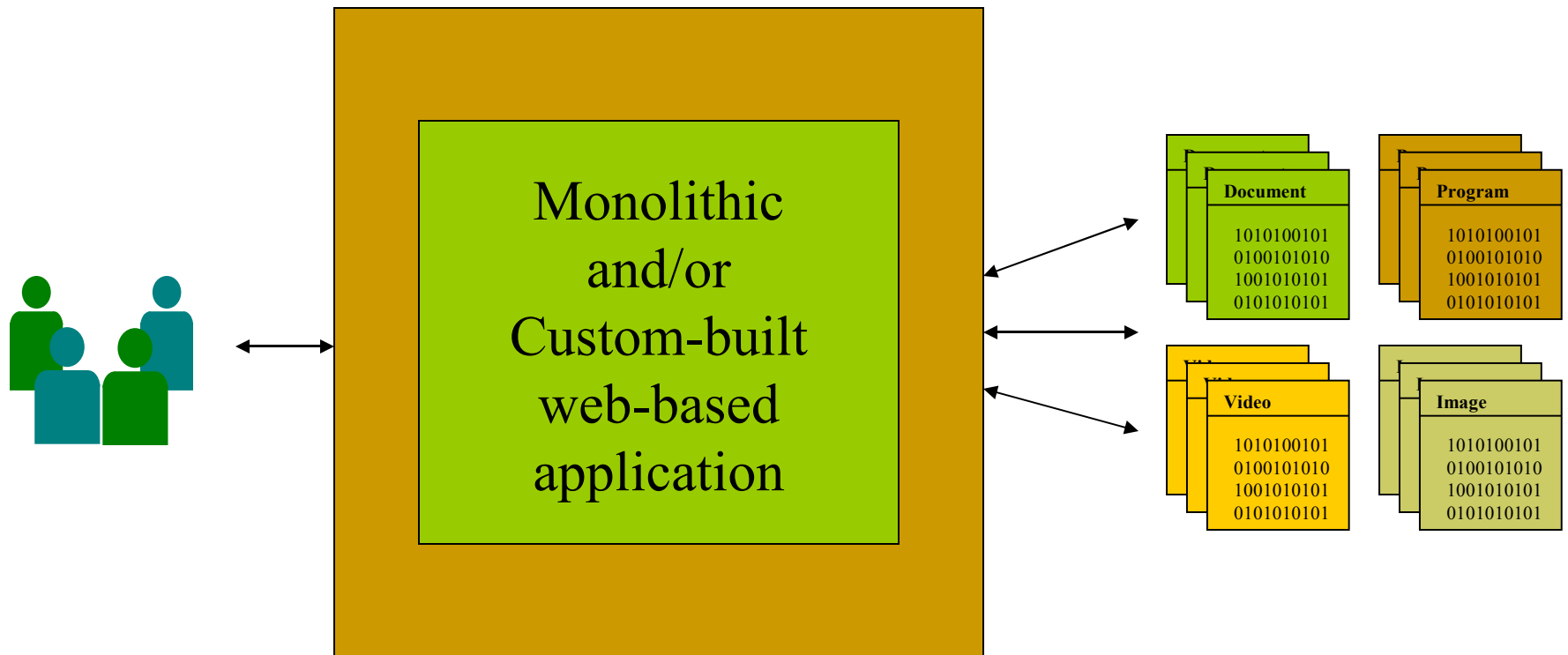
Introduction



users

digital objects

Introduction ...



digital library

Problems

- Digital Libraries are difficult to build – lots of standards and evolving architectures
 - e.g., DSpace, EPrints

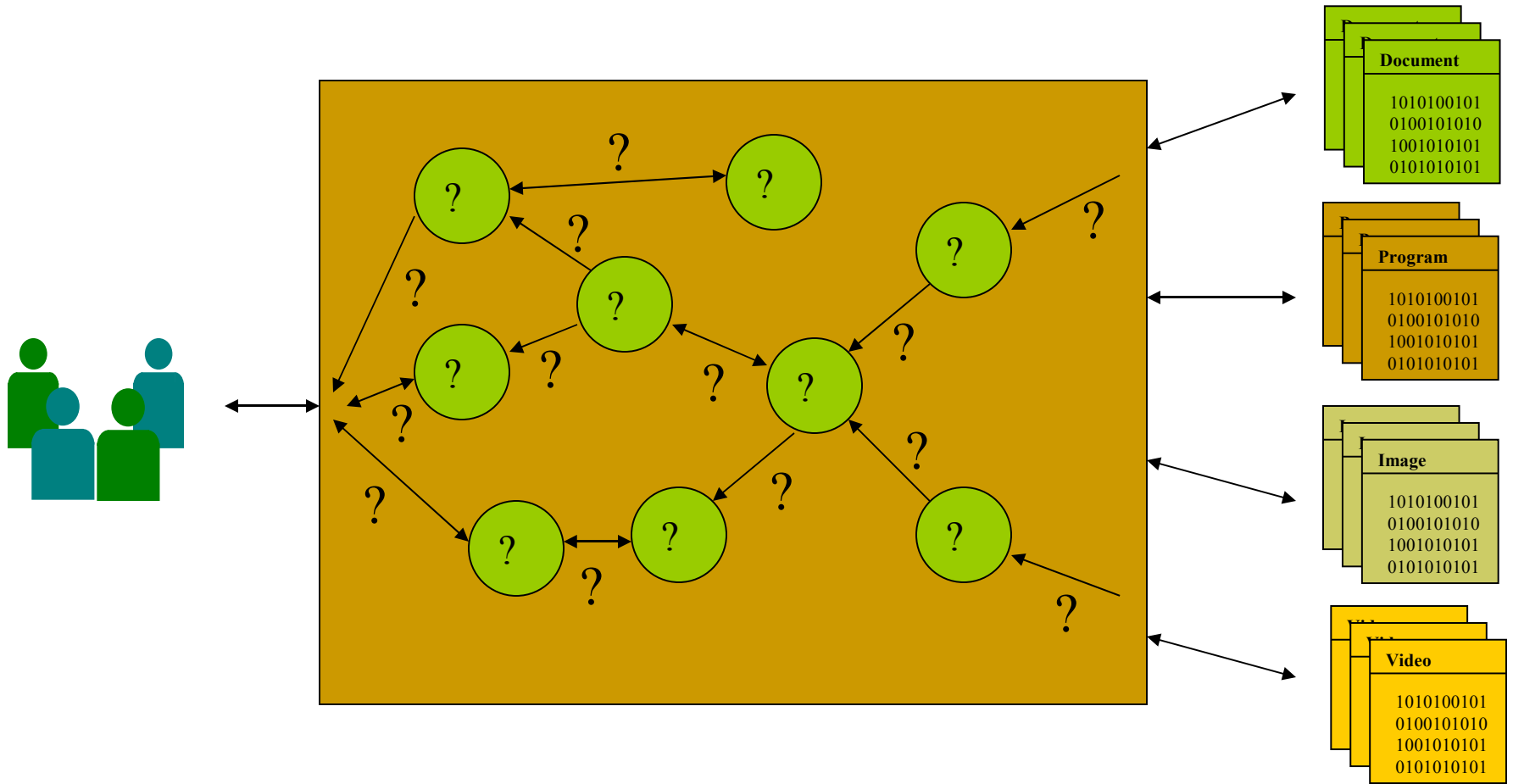
- Interoperability is (was) hard
 - e.g., NCSTRL, Z39.50

- Software development is time-consuming
 - e.g., CSTC, WCR, EPrints

More Problems

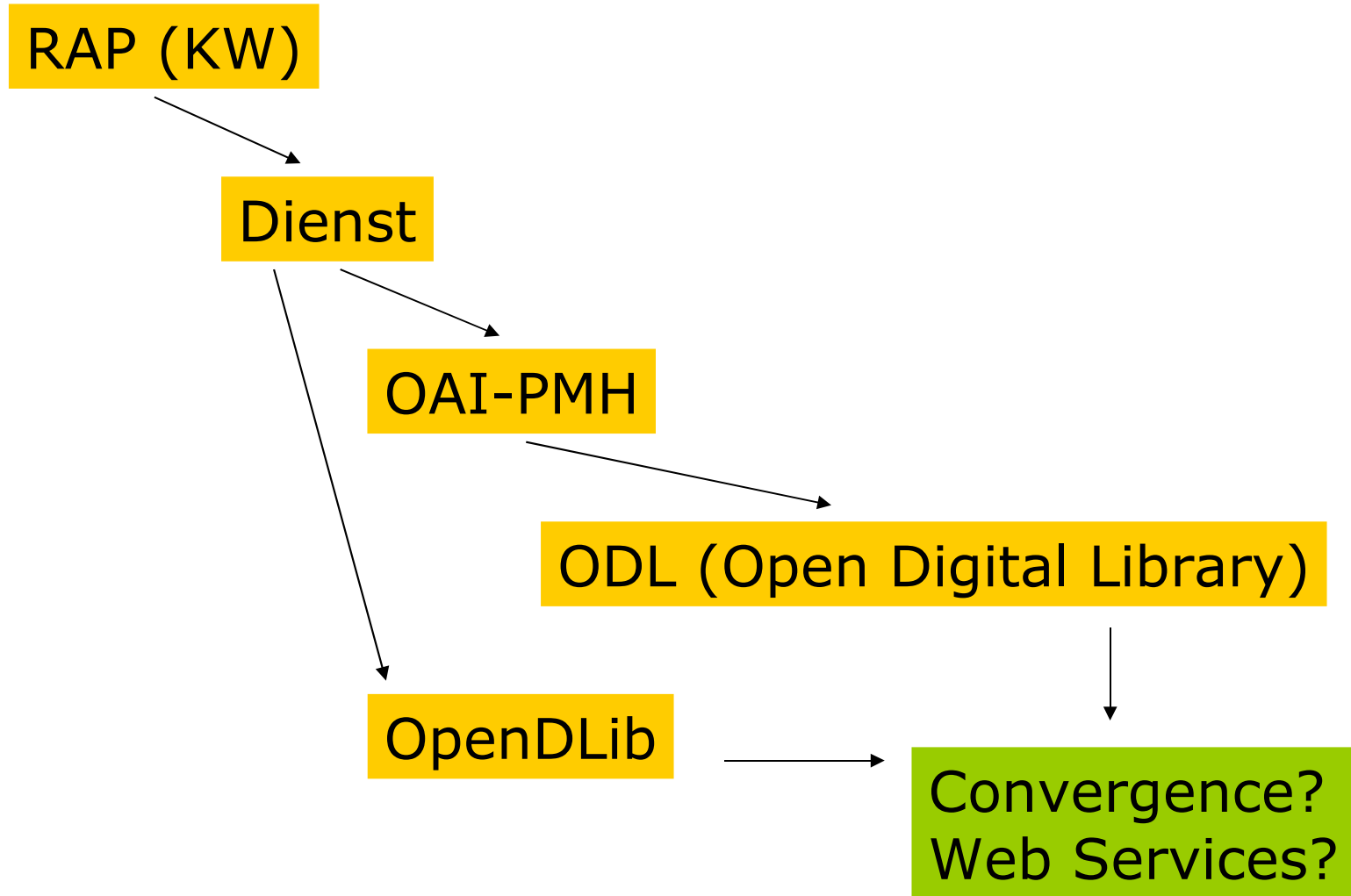
- Poor software engineering
 - Tight coupling
 - Too much complexity
 - Inadequate testing methods
- Lessons from Internet development ignored
 - Simplicity
 - Independence
 - Layering
 - etc.

Solution ?



componentised digital library

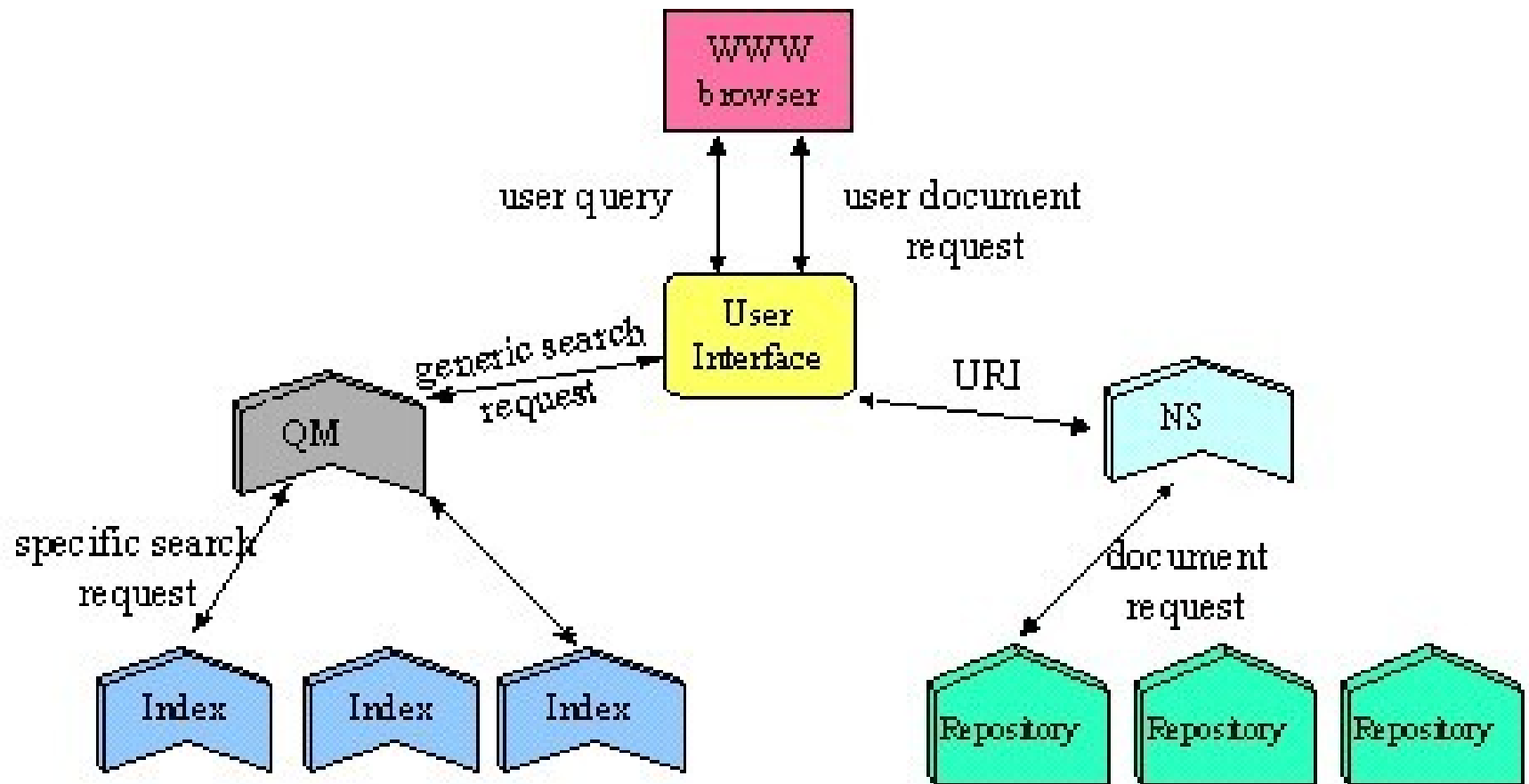
Some Component Architectures



Dienst

- Dienst (German for “service”) is a suite of protocols and components to build distributed digital libraries.
- Dienst is the software suite that supported document management at each of the older NCSTRL (Networked Computer Science Technical Reference Library) sites, and transparently linked them into an international federation of sites.
- Dienst uses federation for interoperability, with a “backup server” for robustness.

Dienst Service Architecture Example



Dienst Example

□ Example Request:

- List the handles in the high energy (hep) partition within the physics partition.

```
/Dienst/Repository/4.0/List-Contents?partitionspec=physics;hep
```

□ Example Response:

```
<?xml version="1.0" encoding="UTF-8"?>
  <List-Contents version="4.0">
    <record>
      handlecorp/970101
    </record>
    <record>
      handlecorp/970102
    </record>
  </List-Contents>
```

Dienst → OAI-PMH

- Dienst formed the foundation for the current OAI-PMH – hence the terminology is sometimes similar.
- NCSTRL has moved to a model based on harvesting and OAI-PMH is being used to connect sites together. In 2001, data from the existing NCSTRL sites was harvested and archived (for preservation) using an early version of an ODL component!
 - see <http://www.ncstrl.org>

Dienst → OpenDLib

- ❑ OpenDLib is a component model for DL services based on Dienst.
- ❑ OpenDLib attempts to define services (mediators) and repositories based on Dienst and updated best practices in DLs.
- ❑ OpenDLib uses a well-defined document model for structured content: the Document Model for Digital Libraries (DoMDL).

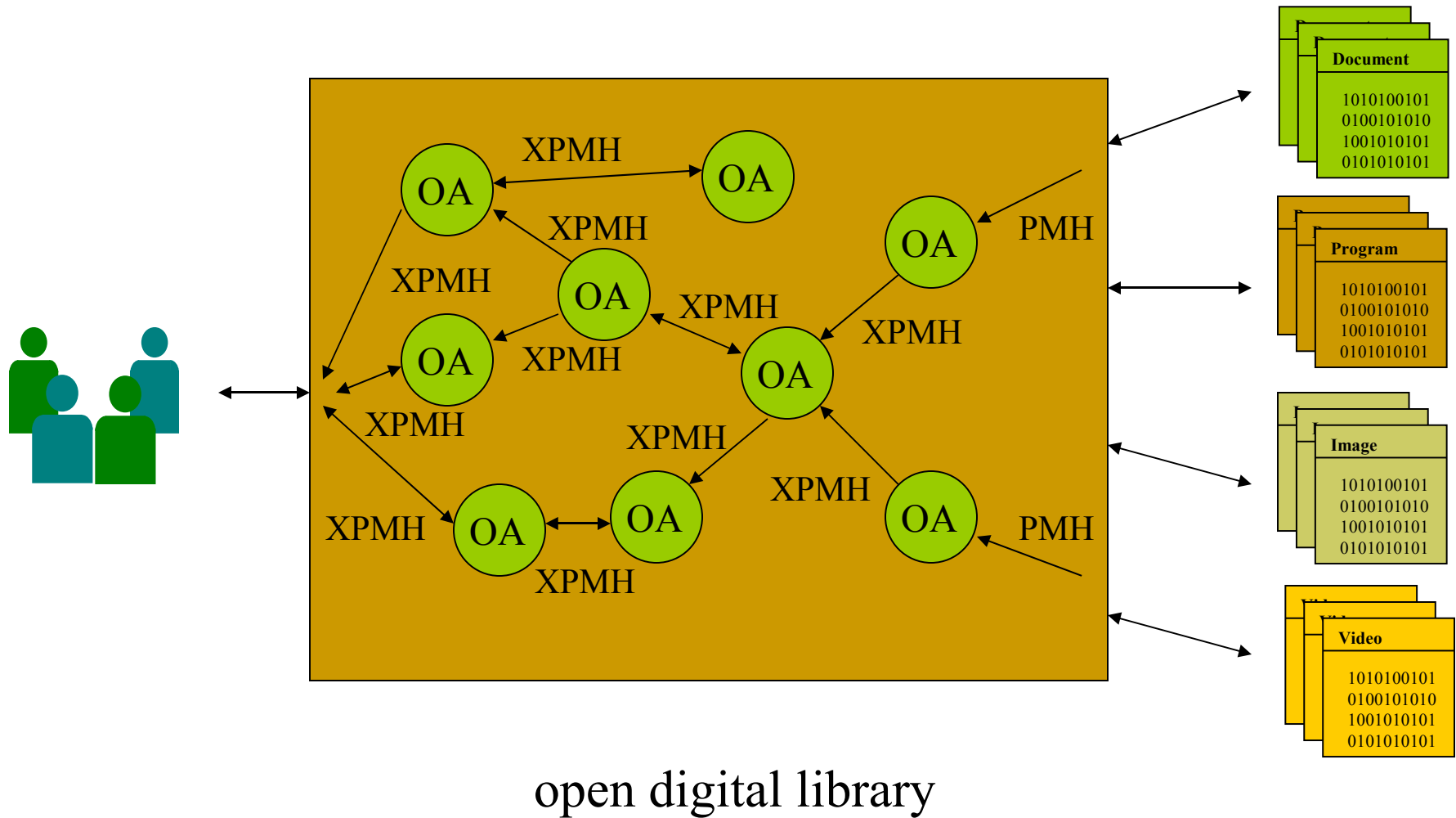
Open Digital Library (ODL)

- An experiment in digital library componentisation.
- Digital Libraries were modeled as networks of extended Open Archives, where each extended Open Archive is a source of data and/or a provider of services.
- Each component was independent, with well-defined external interfaces that are Web-based, e.g., OAI-PMH.

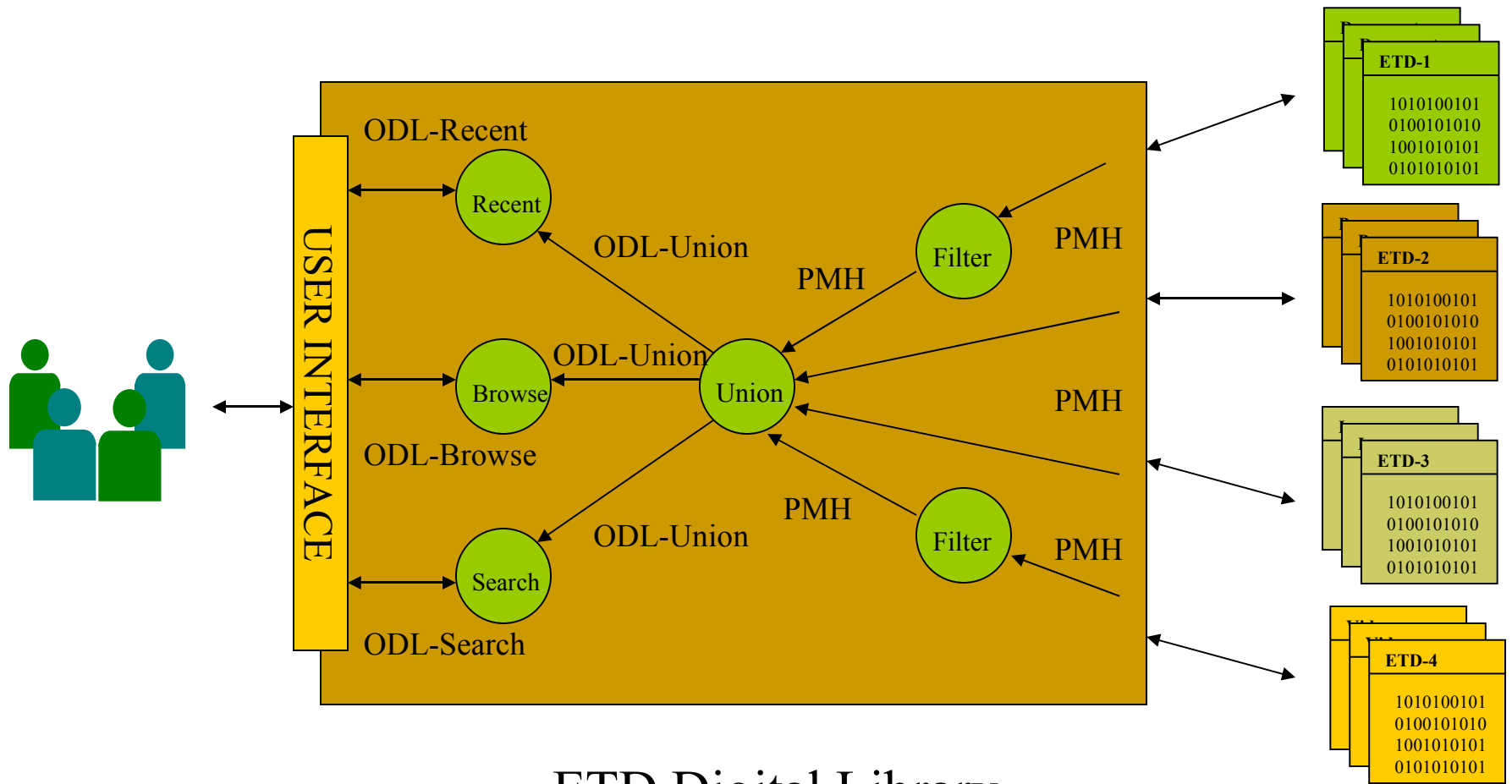
Open DL Design

- ❑ Each component is encapsulated in an extended Open Archive.
- ❑ Communication with other components and user interfaces use specialised versions of the extended OAI-PMH (XOAI-PMH).
- ❑ Digital Libraries are constructed as networks of extended Open Archives.

Problem Revisited



Example Open Digital Library



ETD Digital Library

Students and
researchers

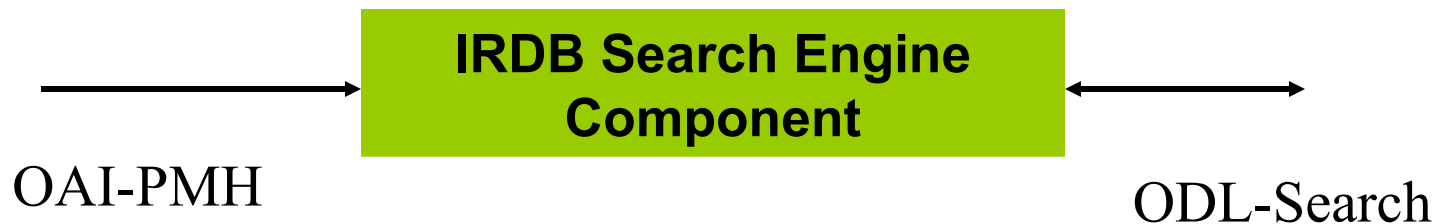
ETD collections

Protocols and Components

Protocol	Component	Description
ODL-Union	DBUnion	Merge archives together
ODL-Search	IRDB	Search engine
ODL-Browse	DBBrowse	Category-based browser
ODL-Recent	WhatsNew	Tracker for recent entries
ODL-Submit	Box	Archive supporting submit and retrieve operations
ODL-Annotate	Thread	Threaded annotation engine
ODL-Recommend	Suggest	Recommendation system
ODL-Rate	DBRate	Ratings manager
ODL-Review	DBReview	Peer review workflow manager

Example: IRDB Search Engine

- ❑ Encapsulate search capability in an OA
- ❑ OAI-PMH to gather data for indexing
- ❑ ODL-Search to submit queries and get results



Example: ODL-Search Protocol

□ Parameters

- query - list of searchable keywords
- query language - "odlsearch1"
- start/stop - subset of ranked list

□ Encoding

- `verb=ListIdentifiers&set=odlsearch1/query/start/stop...`
- `verb=ListRecords&set=odlsearch1/query/start/stop...`

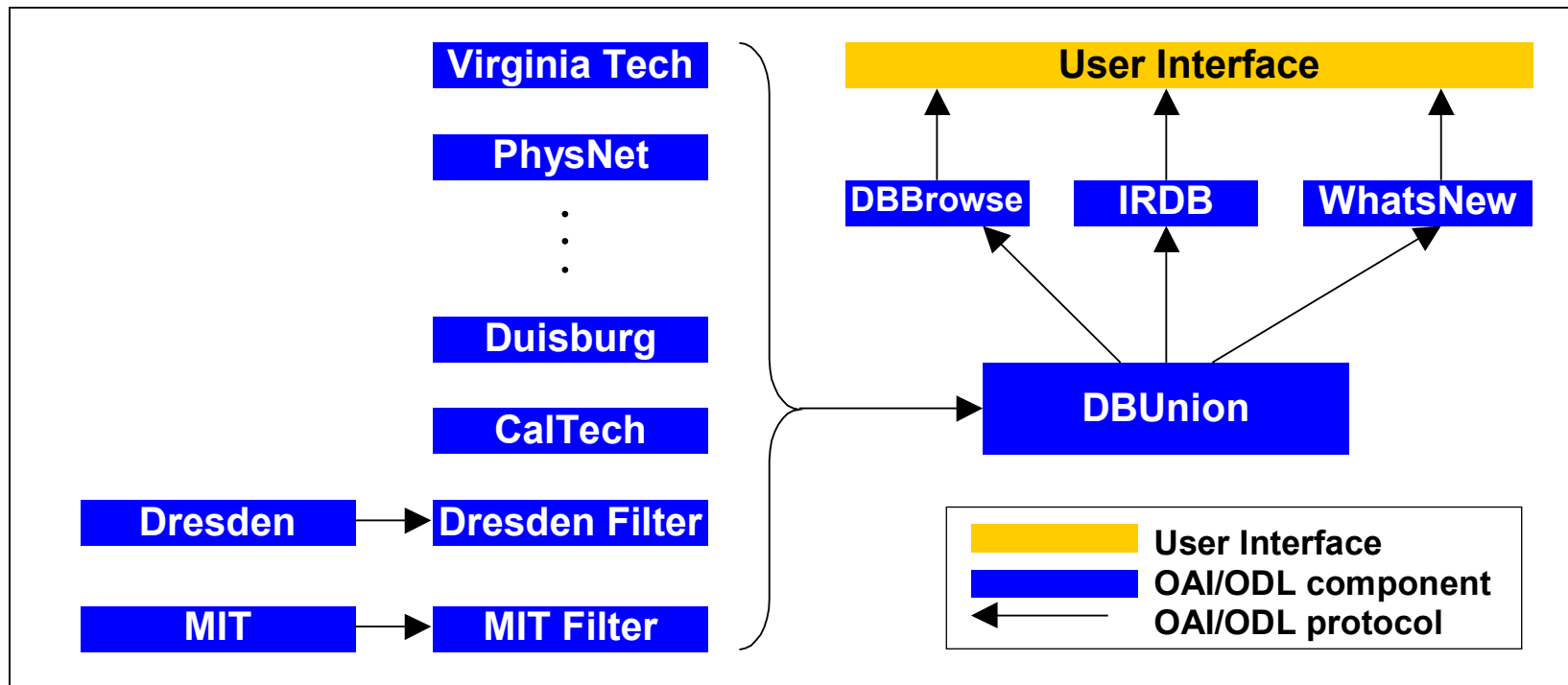
□ Results

- Standard OAI response - list of identifiers or records

□ Example

- `verb=ListRecords&set=odlsearch1/computer science/1/10...`

Case Study: EID Union Catalog



ETD Union Catalog - Front



Electronic Thesis/Dissertation OAI Union Catalog

[Home](#)
[Search](#)
[Browse](#)
[About](#)
[How to Join](#)

Related Sites

- [NDLTD](#)
- [Theses.org](#)
- [Open Archives Initiative](#)

Current Sites

1. [Caltech](#)
2. [PhysNet](#)
3. [Virginia Tech](#)
4. [Humboldt-Universität zu Berlin](#)
5. [Gerhard-Mercator-](#)

Some Recent Additions to our Collection

- ASSESSING AND EVALUATING RECREATION RESOURCE IMPACTS: SPATIAL ANALYTICAL APPROACHES, *Leung, Yu-Fai, Virginia Polytechnic Institute and State University, 1998-04-30* [[More Info](#)]
- The Ontology of Persistence, *Love, Shanon, Virginia Polytechnic Institute and State University, 2001-06-25* [[More Info](#)]
- Virginia Save Our Streams (SOS): Volunteers' Motivations for Participation and Suggestions for Program Improvement, *Haas, Steven Christopher, Virginia Polytechnic Institute and State University, 2000-08-03* [[More Info](#)]

Quick Search

Query :

Quick Browse

Institution : Year :
Sort By :

ETD Union Catalog - Search



Electronic Thesis/Dissertation OAI Union Catalog

[Home](#)
[Search](#)
[Browse](#)
[About](#)
[How to Join](#)

Related Sites

- [NDLTD](#)
- [Theses.org](#)
- [Open Archives Initiative](#)

Current Sites

1. [Caltech](#)
2. [PhysNet](#)
3. [Virginia Tech](#)
4. [Humboldt-Universität zu Berlin](#)
5. [Gerhard-Mercator-Universität](#)

Search Results

Page : [[1](#) [2](#) [3](#)] [Next](#)

1. Characterizing Web Response Time

Liu, Binzhang M.S.

- **Abstract :** It is critical to understand WWW latency in order to design better HTTP protocols. In this study we characterize Web response time and examine the effects of proxy caching, network bandwidth, traffic load, persistent connections for a page, and periodicity. Based on studies with four workloads, we show that at least a quarter of the total elapsed t...

• **Date :** 1998-05-07

• [[More Info](#)] [[Go To Document](#)] [[Find Similar Documents](#)]

2. Chemical Interferences on the Atomization Yield of High Reduction Potential Elements - Signal Suppression in the Plasma Source Spectrometry

LIU, Jian, jianliu78@yahoo.com, 1960-11-04, Changchun, CHINA

• **Date :** 2001-03-08

• [[More Info](#)] [[Go To Document](#)] [[Find Similar Documents](#)]

ETD Union Catalog - Browse

[Home](#)
[Search](#)
[Browse](#)
[About](#)
[How to Join](#)

Related Sites

- [NDLTD](#)
- [Theses.org](#)
- [Open Archives Initiative](#)

Current Sites

1. [Caltech](#)
2. [PhysNet](#)
3. [Virginia Tech](#)
4. [Humboldt-Universität zu Berlin](#)
5. [Gerhard-Mercator-Universität Duisburg](#)
6. [Technische Universität Dresden](#)

Browse ETDs

Institution : Year :

Sort By :
Aarhus University, Chemistry Department
California Institute of Technology
Caltech Library System
Forskningsscenter Risoe
Gerhard-Mercator-Universität Duisburg
Humboldt-Universität zu Berlin
Massachusetts Institute of Technology
Other
Risoe National Laboratory
Technische Universität Dresden
Universitaet Stuttgart
Virginia Tech

[10](#) > [11](#) >> [361](#)] [Next](#)

of the linked bibliography capability.

2. Microstructure and Crystallization Behavior in Bulk Glass Forming Alloys.

Bossuyt, Sven

- **Abstract :** The solidification microstructure in wedge-shaped castings of Cu-Ni-Ti-Zr glass forming alloys is investigated, while the composition was systematically varied. Near the critical thickness for glass formation, a spatially inhomogeneous dispersion of nanocrystals is observed, where spherical regions contain a much higher density of nanocrystals than...
- **Date :** 2001-07-13
- [[More Info](#)] [[Go To Document](#)] [[Find Similar Documents](#)]

3. All-Optical Logic Circuits based on the Polarization Properties of Non-Degenerate Four-Wave Mixing

Search Protocols

- Z39.50 is the traditional remote search protocol for library systems.
 - ANSI/NISO/ISO standard
 - Comparatively complicated syntax/operation
 - Based on older standards (1998)
- ZING (Z39.50 International Next Generation) is the latest updated version.
 - SRW – Search/Retrieve for the Web
 - SRU – Search/Retrieve URL mechanism?
 - `http://myserver.com/myurl/searchRetrieve?query=dc.title=cat&maximumRecords=10&recordSchema=http%3a//www.loc.gov/mods/&sortKeys=title,dc&startRecord=1` (excerpt from ZING website)

SRW Request

SOAPAction: "searchRetrieve"

```
<SOAP:Envelope>
  <SOAP:Body>
    <SRW:searchRetrieveRequest
xmlns:SRW="http://www.loc.gov/zing/srw/v1.0/">
      <SRW:query>(dc.author exact "jones" prox///5 title >=
"smith")</SRW:query>
<SRW:sortKeys>/record/title,"http://www.loc.gov/zing/srw/dcsche
ma/v1.0/",1,0,highValue
/record/datafield[@tag="100"]/subfield[@code="a"],"http://www.l
oc.gov/marcxml/,,,,"Smith"</SRW:sortKeys>
      <SRW:startRecord>1</SRW:startRecord>
      <SRW:maximumRecords>10</SRW:maximumRecords>

      <SRW:recordSchema>http://www.loc.gov/mods/</SRW:recordsSchema>
    </SRW:searchRetrieveRequest>
  </SOAP:Body>
</SOAP:Envelope>
```

(excerpt from ZING website)

SRW Response

```
<SOAP:Envelope>
<SOAP:Body>
<SRW:searchRetrieveResponse xmlns:SRW="http://www.loc.gov/zing/srw/v1.0/"
  xmlns:DIAG="http://www.loc.gov/zing/srw/v1.0/diagnostic/">
  <SRW:numberOfRecords>2</SRW:numberOfRecords>
  <SRW:resultSetId>8c527d60-c3b4-4cec-a1de-1ff80a5932df</SRW:resultSetId>
  <SRW:resultSetIdleTime>600</SRW:resultSetIdleTime>
  <SRW:records>
    <SRW:record>
      <SRW:recordSchema>http://www.loc.gov/mods/</SRW:recordSchema>
      <SRW:recordData> &lt;?xml version="1.0"
encoding="UTF-8"?&gt; &lt;mods
xmlns:xlink="http://www.w3.org/TR/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.loc.gov/mods/"
xsi:schemaLocation="http://www.loc.gov/mods/
http://www.loc.gov/standards/mods/mods.xsd"&gt; &lt;titleInfo&gt;
&lt;title&gt;Sound and fury : the making of the punditocracy
/&lt;/title&gt; &lt;/titleInfo&gt; &lt;name
type="personal"&gt; &lt;namePart&gt;Alterman,
Eric.&lt;/namePart&gt; &lt;role&gt;creator&lt;/role&gt;
&lt;/name&gt; ...
      </SRW:recordData>
      <SRW:recordPosition>1</SRW:recordPosition>
    </SRW:record>
  ...
  (excerpt from ZING website)
```

Some Production Protocols

- OAI-PMH – data transfer (OAI)
- SRU/SRW – remote searching (LoC)
- Annotea – annotation management (W3C)

- Note:
 - ODL, OpenDLib and Dienst are research-driven!

The Ultimate Goal

- Package different configurations of components into instant DL systems.
- DL building = component configuration.
- All DLs speak the same language(s).
- Basic services are trivial to provide so more effort is spent on advanced capabilities of DLs.
- Information is more accessible to users.

The FDL Project – as of 2005

- Design digital library middleware using IDE-like visual interface (BLOX, Linda Eyambe).
- Design user interface to connect to middleware using visual interface (Kevin Feng).
- Package custom configurations for rapid redeployment (Siya Mhlongo).
- Extend component architecture to scale on clusters of computers (Muammar Omar).

References

- Suleman, H. and E. A. Fox (2001) "A Framework for Building Open Digital Libraries", in D-Lib Magazine, Vol 7., No. 12, December 2001. Available <http://www.dlib.org/dlib/december01/suleman/12suleman.r>
- Lagoze, Carl and James Davis (1995) "Dienst: an architecture for distributed document libraries", Communications of the ACM, ACM, Vol. 38, No. 4, p. 47.
- Castelli, Donatella and Pasquale Pagano (2002) "OpenDLib: A Digital Library Service System", in Proceedings of Research and Advanced Technology for Digital Libraries: 6th European Conference (ECDL 2002), Rome, Italy, September 2002, Lecture Notes in Computer Science 2458, p. 292-307. Maristella Agosti, Costantino Thanos (eds.). Springer, 2002.