

Institutional Repository Software

Features and Functionality

T.B. Rajashekar
National Centre for Science Information
Indian Institute of Science
Bangalore – 560 012
(E-Mail: raja@ncsi.iisc.ernet.in)

INDEST-NCSI Workshop on Institutional
Repositories,
IISc, Bangalore, 27-29 July 2004

IR Software

- Key component of an IR is the repository management software
- Several software now available under open source license
- Comply with OAI metadata harvesting protocol
- Released and publicly available

IR Software

- ARNO
 - Academic Research in the Netherlands Online, Tilburg University, The Netherlands
 - <http://www.uba.uva.nl/arno>
- CDSware
 - CERN Document Server Software (CDSware), CERN, Geneva, Switzerland
 - <http://cdsware.cern.ch/>

IR Software

- DSpace
 - MIT Libraries, Cambridge, MA USA
 - <http://www.dspace.org/>
- EPrints
 - University of Southampton, U.K.
 - <http://software.eprints.org/>
- Fedora digital object repository management system
 - University of Virginia, USA
 - <http://www.fedora.info/>

IR Software

- i-Tor
 - Tools and technologies for Open Repositories
 - Netherlands Institute for Scientific Information Services
 - <http://www.i-tor.org/en/toon>
- MyCoRe
 - Essen University Library, University of Duisburg-Essen, Germany
 - <http://www.mycore.de/engl/index.html>

Repository Software

- Installed base (as on January 2004)

Arno	CDSWare	DSpace	EPrints	Fedora	i-Tor	MyCore
7	7+	15+	106	20	10	10

EPrints and DSpace are more widely used today internationally

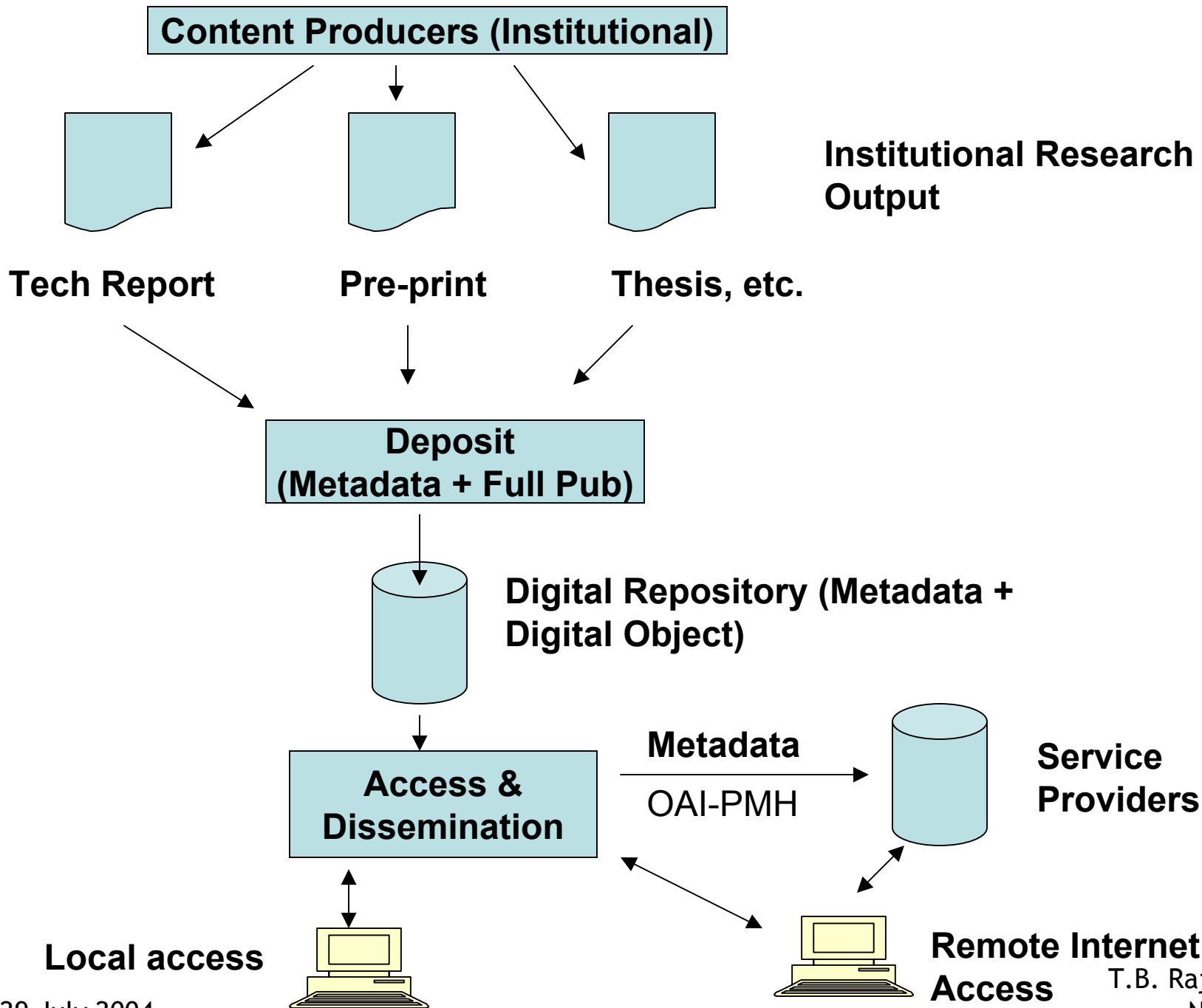
Source:

A guide to Institutional Repository Software. 2nd edition. Open Society Institute. January 2004. Contains summary information about each repository software and a very detailed feature and functionality table.

<http://www.soros.org/openaccess/software>

What IR software aim to do?

- Capture and describe digital material using a workflow
 - Provide interface for online submission of research material (intranet)
- Provide access to this material over the web (metadata and/or full pub)
- Preserve digital material over long period of time
- Expose metadata through OAI-PMH protocol
 - Default: Unqualified Dublin Core
 - Other metadata standards



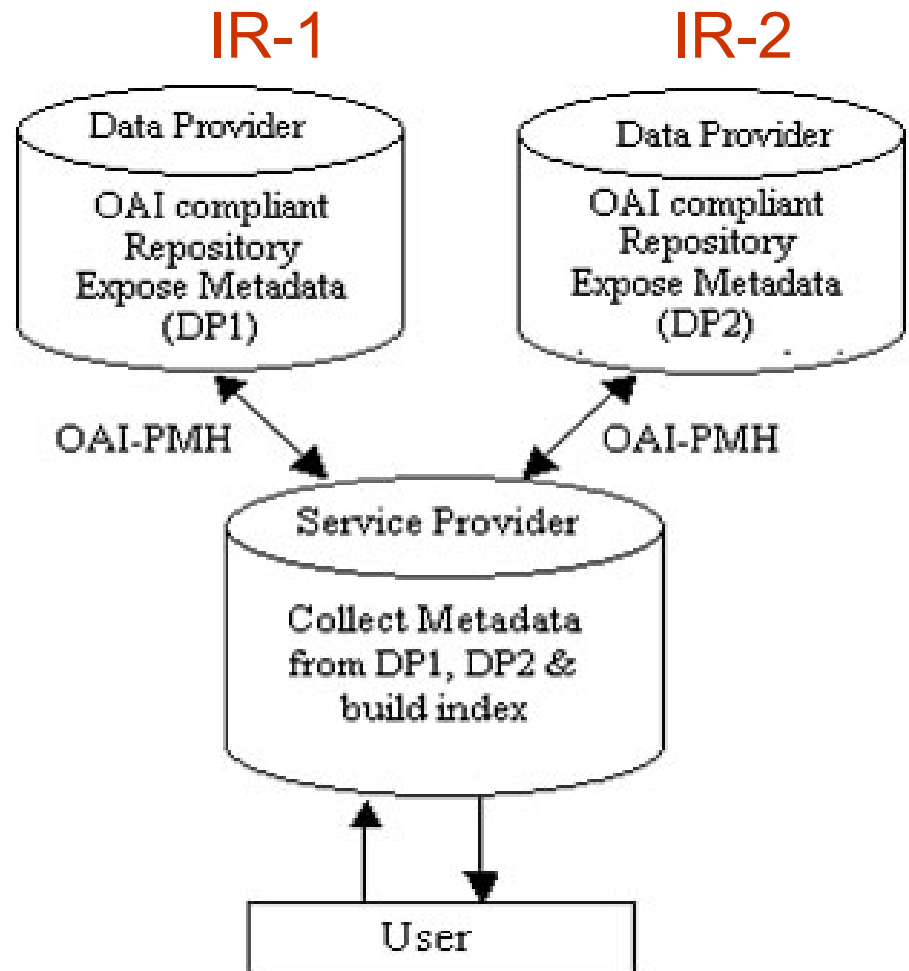
Interoperability through OAI-PMH Protocol*

- Data Provider (Ex. Institutional repository)

- Maintain repository
- Expose metadata according to a metadata standard (e.g. DC)
- Register with OAI

- Service provider

- Register with OAI
- Extract metadata from registered repositories ('harvest')
- Provide services (e.g. central index)



* <http://www.openarchives.org/>

What content can be handled?

- Example Criteria*
 - Scholarly: research/ teaching oriented
 - Produced by an institution's research community
 - Non-ephemeral: work in complete form, ready for dissemination
 - Perpetual license: author grants the right to the institution to preserve and distribute the work via the repository

* SPARC Institutional Repository Checklist & Resource Guide. Release 1.0, Nov 2002. http://www.arl.org/sparc/IR/IR_Guide.html

Content Examples

- Published material
 - Ex.: Journal papers (post-prints), book chapters, conference papers
- Unpublished/ gray material
 - Ex.: Pre-prints, working papers, minutes, theses and dissertations, technical reports, progress/ status reports, committee reports, course material, presentations, multimedia material, etc.
- Supporting material
 - Ex.: Data sets, models, simulations

Repository Software Systems

Key Features and Functionality

Specific features and functionality varies across
different systems

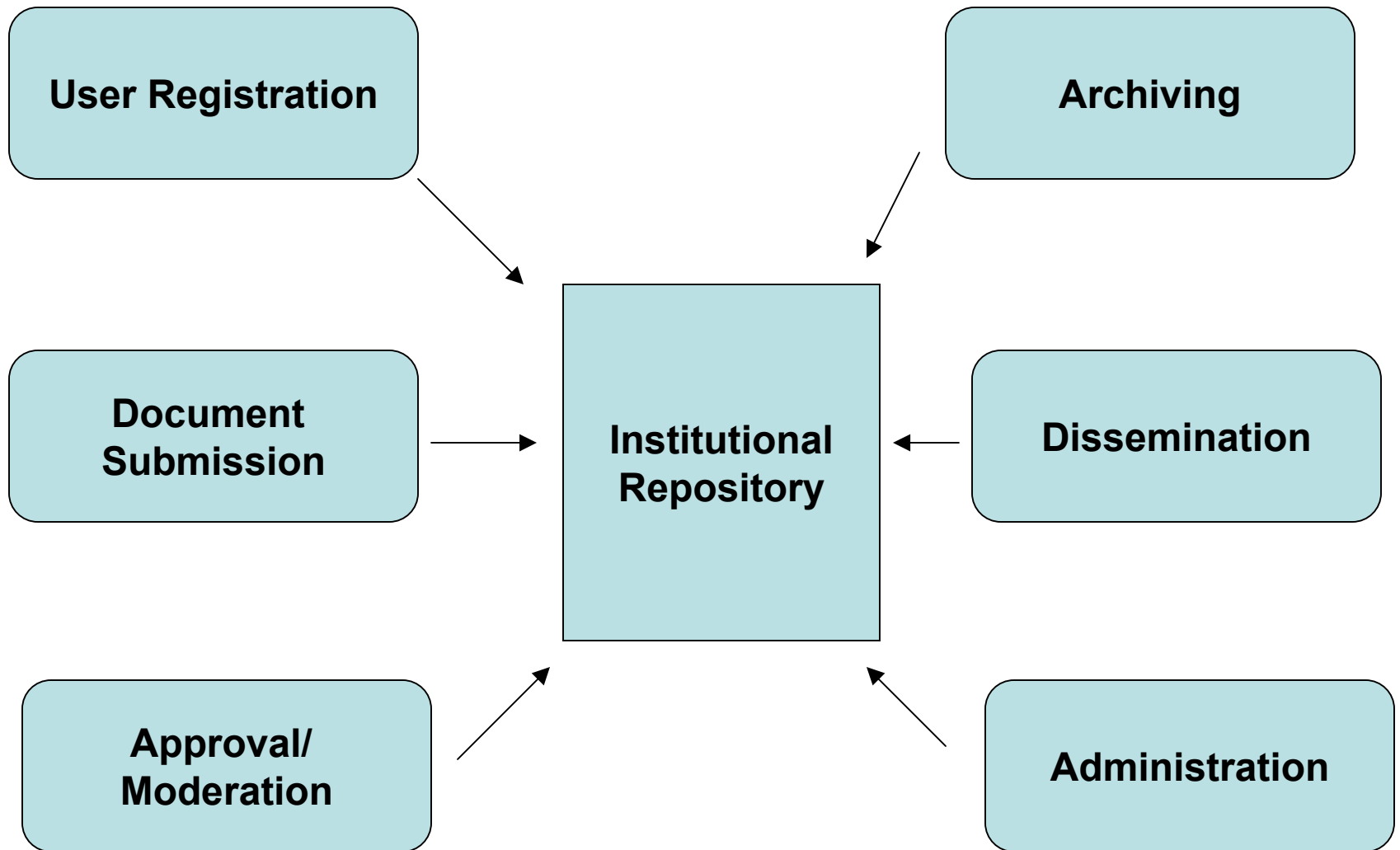
Key Features and Functionality

- **Administration**
 - Administer collections, users
 - Define/ customize document types, formats, metadata
 - Define workflow
 - Licenses (author permissions)
 - Customize submission forms, home page
- **Registration of institutional users (authors)**
 - For document submission and other privileged use
 - User authentication
 - Profile set up (user record)
- **Document submission**
 - Authentication
 - Assign Metadata, validation
 - Upload Document
 - Grant license (permission)

Key Features and Functionality

- **Approval/ moderation (workflow)**
 - Submission approval (metadata, format, affiliation, etc.)
 - Content approval (peer review)
- **Archiving**
 - Date stamping
 - Unique/persistent identifier assignment
 - Preservation support (checksum)
 - Indexing and storage
- **Dissemination**
 - Search, browse
 - Alerting service
 - OAI registration and compliance (metadata exposure)
 - Rights management

Repository Functionality



Additional Functionality

- Multilingual content/ interface support
 - Unicode
- Batch import/export
- Other interoperability protocols
 - Z39.50, OpenURL
- Other metadata standards and cross-walks (e.g. METS, MARC)
- Persistent URL (Identifier)

EPrints and DSpace

- Widely used IR software
- Platform:
 - EPrints: Unix/ Linux/ Perl/ Apache/ MySQL/ XML/ HTML/
 - DSpace: Unix/ Linux/ Java/ Tomcat or Apache/ XML/ HTML/ Ant/ PostGreSQL


Imply software knowledge required for installing, configuring, and maintaining archives developed using these packages.

Why we have chosen EPrints?



- [EPrints 2.3.4 and DSpace 1.2Beta3]
- Lower learning curve
- Better suited for research publications
- Ease of customization – document types, metadata, subject classification, home page, help instructions, etc.
- Better support for customization (e.g. ‘How to...’)
- Search and display – better designed for research publications

Why we have chosen EPrints?

- Both the software are evolving
- Strengths of DSpace
 - Communities/ collections (needs good insight to use these properly – study carefully current examples)
 - Backed up by MIT and HP!
 - Strong workflow support (but may be difficult to define and manage?)
 - Handle-based identifier
 - Better articulation of preservation strategy
 - Default support for qualified DC

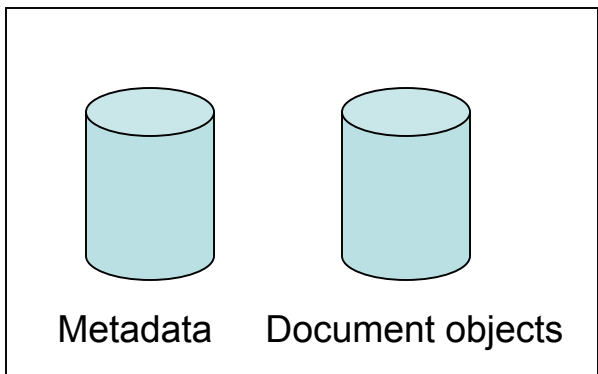
Why we have chosen EPrints?

- Either of these (or other) software can be used
- You can switch to other software after some time, if required
- Import/ export options can facilitate this process

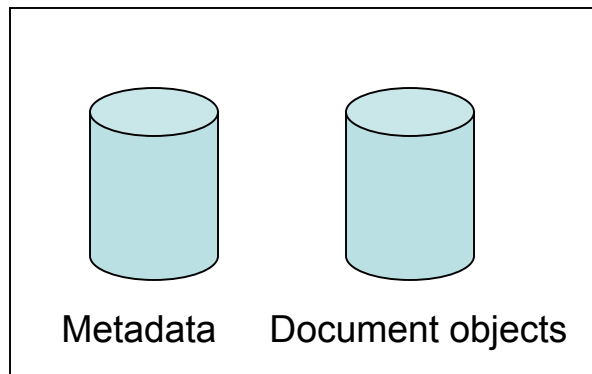
What we need to do?

- Key decisions for INDEST (loud thinking!)
 - INDEST Harvesting service, facilitating effective cross-search of all INDEST IRs
 - Registry of doc types?
 - Define mandatory data elements?
 - Rendering format for key fields (e.g. Author names)
 - Define OAI compliance for INDEST harvesting? (see for example DARE project specification)

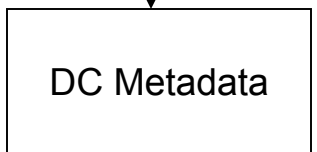
Repository 1



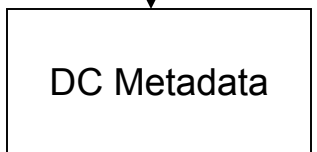
Repository 2



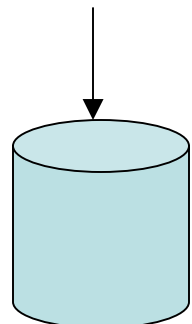
OAI-DC (unqualified)
(mapping)



OAI-DC (unqualified)
(mapping)

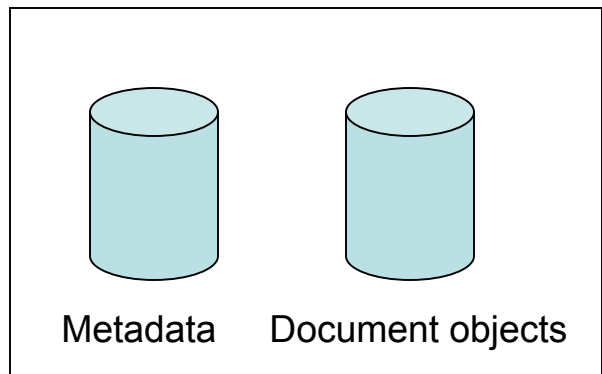


Metadata Harvester (e.g. ARC, OAIster)

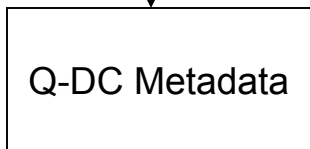


Cross-Index

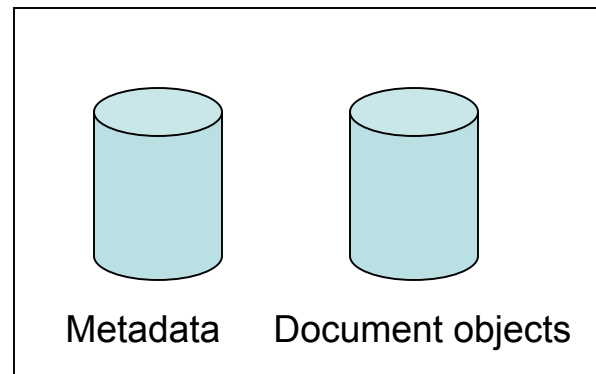
Repository 1



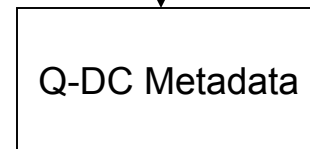
OAI-DC (qualified)
(mapping)



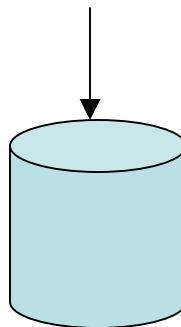
Repository 2



OAI-DC (qualified)
(mapping)



INDEST Metadata Harvester



INDEST Cross-Index

Related Resources

- Open Society Institute. A guide to institutional repository software. 2nd Edition. January 2004. <http://www.soros.org/openaccess/software>
- Open Archives Initiative (OAI). <http://www.openarchives.org/>
- Creative Commons. <http://www.creativecommons.org/>
- CNRI. Handle system. <http://www.handle.net/>
- DARE use of Dublin Core. Version 1.0, October 2003. (<http://www.surf.nl/DARE>)