Identifiers and Repositories

hussein suleman

uct cs honours 2008

Features of (Good) Identifiers

- Must uniquely name digital objects and/or metadata.
  - e.g., ISBN numbers
- Must be unique within global/local domain.
  - e.g., Email addresses
- May have scheme associated to indicate source or naming convention.
  - e.g., webpage URLs – http://something …

Example Repository

Identifier	MetadataX	DC
Proverb/1	Id: P1 Data: D1	Source: Africa Type: Proverb Identifier: D1
Proverb/2	Id: P2 Data: D2	Author: Albert Einstein Type: Quotation Identifier: D2

Identifier	Data
D1	When an old man dies, a library burns down
D2	Imagination is more important than knowledge

Conceptual Model
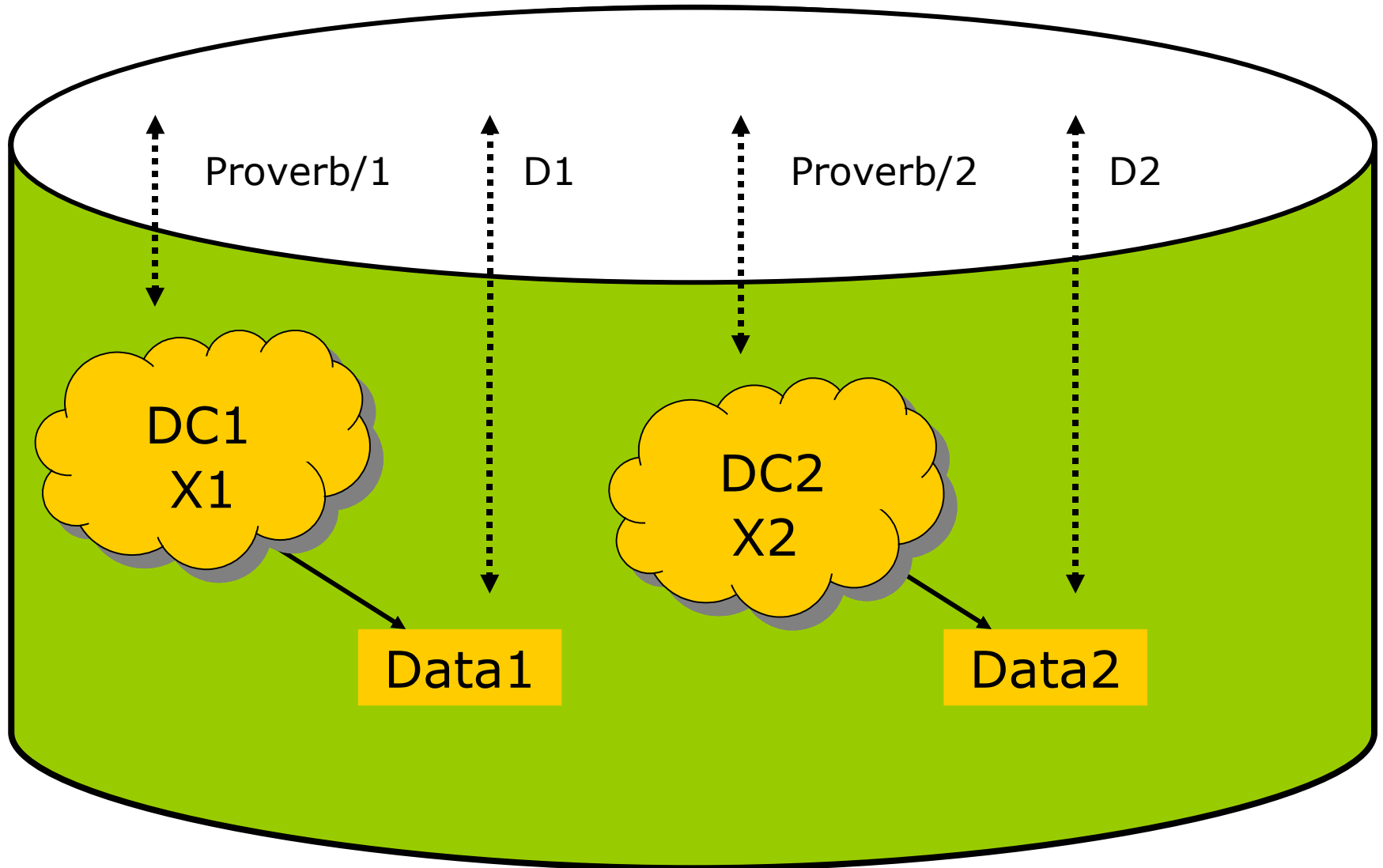
Storage Models

- Generalisation of database.
- Collection of metadata records.
  - in XML or other flat files
  - in database BLOBs
  - in columns of database tables
  - embedded in digital objects
- Abstract interface to data collection.
  - no concept of how (meta)data is stored.
  - think distributed file systems.
  - think NFS for remote file systems.

Repository Access Protocol (RAP)

- A repository can be defined as a network-accessible server.

- RAP specifies a simple interface to access and manage digital objects in a repository.

- RAP is an abstract model, with concrete implementations in Dienst, OpenDLib, OAI, ODL, Fedora, etc.

- This is usually referred to as the "Kahn/Wilensky architecture".
  - does Kahn ring any bells?

RAP Operations

- ACCESS_DO
  - Return a manifestation (dissemination) of a digital object based on its identifier and a specification of what service is being requested.
- DEPOSIT_DO
  - Submit a digital object to the repository, assigning or specifying an identifier for it.
- ACCESS_REF
  - List services and their access mechanisms for the repository.

RAP: Naming of Digital Objects

- Each digital object must have a location-independent name (handle), made up of a repository identifier and a local name.
  - Example:
    - berkeley.cs/csd-93-712
    - where berkeley.cs is the repository and csd-93-712 refers to a technical report.
- Handles are resolved by a handle server to redirect a service provider to a repository containing an object identified only by its location-independent handle.

Handle Servers

- A handle server stores the association between handles and physical locations of objects.

- Handle servers follow a DNS model:
  - they are distributed and replicated
  - there are global and local servers
  - handles may be cached locally after being resolved to minimise resolution traffic
  - management of servers/handles requires an authority system for management, accountability, delegation, etc.

Handle Example

PORTAL
University of Cape Town

Search: ○ The Guide   ● The ACM Digital Library

[SEARCH]

THE ACM DIGITAL LIBRARY

Feedback   Report a problem   Satisfaction survey

Going digital: a look at assumptions underlying digital libraries

Full text   📄 Pdf (220 KB)

Source   **Communications of the ACM** archive
Volume 38 , Issue 4 (April 1995) table of contents
Pages: 77 - 84
Year of Publication: 1995
ISSN:0001-0782

Authors   David M. Levy   Xerox PARC, Palo Alto, CA
Catherine C. Marshall   Texas A&M Univ., College Station

Publisher   ACM Press   New York, NY, USA

Additional Information: abstract   references   citings   index terms   collaborative colleagues   peer to peer

Tools and Actions: Discussions   Find similar Articles   Review this Article
Save this Article to a Binder   Display in BibTex Format

DOI Bookmark: Use this link to bookmark this Article: http://doi.acm.org/10.1145/205323.205346
What is a DOI?

Digital Object Identifiers (DOIs)

- DOIs are a standardised implementation of the handle concept.

- Handles/DOIs are URIs that refer to digital objects while URLs are URIs that refer to network services.

- Handle/DOI resolution can be performed transparently using a browser plug-in.

Going digital - Netscape

File   Edit   View   Go   Bookmarks   Tools   Window   Help

doi:10.1145/205323.205346
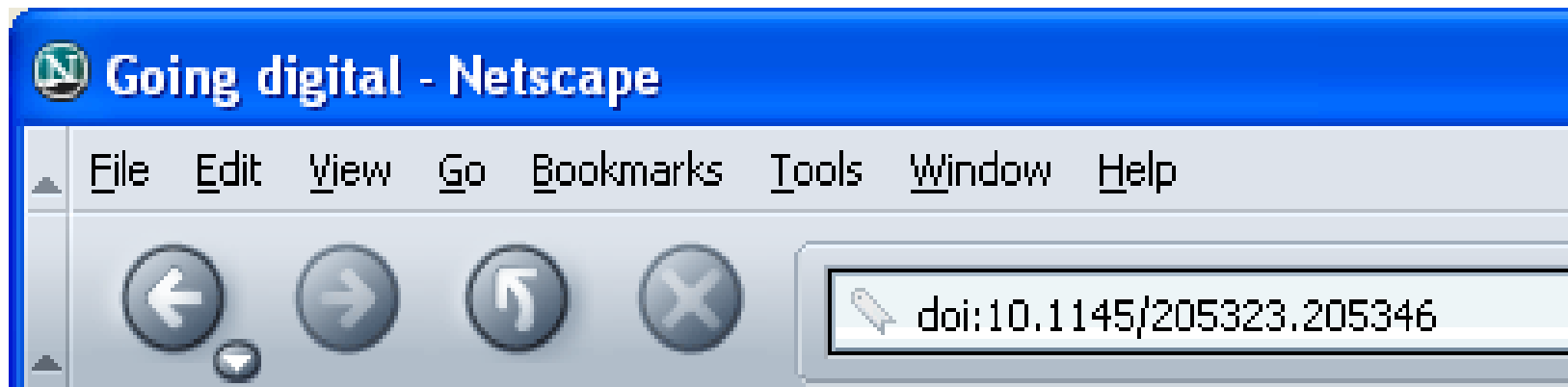
Other repository models

- FEDORA (Flexible Extensible Digital Object and Repository Architecture) defines a generic interface to manage digital objects at a lower layer in an information system.

- SODA (Smart Objects Dumb Archive) packages digital objects into buckets containing the data along with the code to mediate access, display the objects, enforce rights, etc.

Institutional Repositories

- "Green Route" for **Open Access**
- Archiving of research-related documents.
- User interface to locate and access documents (Web-based).
- Administration interface to maintain archive.
- Ability for users to submit documents.
- Ability for authorities to review and accept submissions.

Criteria for Software

- Preservation – does the software support long-term maintenance of documents?
- Security – how can we be certain that the system cannot be circumvented?
- Stability – will it die when we need it most at the end of year/semester?
- Interoperability – will it connect into other systems such as the library ILS or the NDLTD Union Catalog?
- Standards-compliance – what does it adhere to?
- Cost – does it?
- Hardware – what do we need to run the software?
- Support – how much staffing do we need to run it?

EPrints

- Software to archive electronic pre-prints of journal/conference proceedings, but easily adaptable to any OA need.
- Perl Web server scripts and Mysql database, plus a few Perl libraries.
- Needs a bit of effort for initial installation and customisation but has a slick user interface.
- Active user community in many different areas, especially Open Access.
- Modifying the software can be tricky.

EPrints 2 Interface

ETD Individuals   Home || About || Browse || Search || Register || User Area || Help

NDLTD

Year: 2002

Number of records: **9**

Niefer, Dr. Inge Andrea (2002) *Analysis of the visitors' profile of the islands Ilha do Superagüi e Ilha do Mel - Marketing as an instrument for sustainable tourism.* UNSPECIFIED, Federal University of Paraná, Brazil.

Mellert, Dipl.-Phys. Karolin (2002) *Aufbau einer Interferenzlithografie-Anlage zur Herstellung photonischer Kristalle.* UNSPECIFIED, University of Bonn, Germany.

Eraßme, Dr. Rolf (2002) *Der Mensch und die 'Künstliche Intelligenz' Eine Profilierung und kritische Bewertung der unterschiedlichen Grundauffassungen vom Standpunkt des gemäßigten Realismus.* UNSPECIFIED, www.rwth-aachen.de.

Reyes Rodríguez, PhD Ricardo (2002) *DEVELOPMENT AND DIFFERENTIATION OF THE VERTEBRATE PITUITARY GLAND.* UNSPECIFIED, University of La Laguna.

Gennari, R. (2002) *Mapping Inferences. Constraint Propagation and Diamond Satisfaction.* UNSPECIFIED, Institute of Logic, Language and Information (ILLC), Universiteit van Amsterdam.

Meier, Dr. Michael (2002) *Returning Science to the Scientists. Der Umbruch im STM-Fachzeitschriftenmarkt durch Electronic Publishing.* UNSPECIFIED, Ludwig-Maximilinas-Universität München.

Petit, Dr Frédéric (2002) *Reverberation Chamber Modeling Using Finite-Difference Time-Domain Method.* UNSPECIFIED, University of Marne la Vallée.

EPrints 3 Interface

DSpace

- DSpace was developed by Hewlett-Packard for MIT to manage its institutional repositories, but it is open source and free for anyone else to use.

- Has the backing of a professional software development company, and an open source community.

- Based on Java servlets, Postgres database, Java libraries.

- Widespread use for many different purposes.

- Easiest package to run on Windows (besides Greenstoone)!

DSpace Interface

Manakin – DSpace Customisation

Manakin

◇ DSpace Home

DSpace Manakin is Live

Welcome to our new Manakin interface to the DSpace digital repository of My University.

More exciting news to appear here.

Search DSpace

Enter some text in the box below to search DSpace.

[   ] Go

Communities in DSpace

Select a community to browse its collections.

Search DSpace

[   ]
Go

Advanced Search

Browse

- **All of DSpace**
  - ◇ Communities & Collections
  - ◇ Titles
  - ◇ Authors
  - ◇ Subjects
  - ◇ By Dates

My Account

- Login
- Register

Fedora

- Fedora is a digital asset management system.
- It provides only a repository with a Web Services interface – other tools need to be layered on Fedora.
- It is argued that it has the best architecture and is most scalable of all systems.
- Fez and Vital/Valet are some IR tools built on top of Fedora.
- Fez is new and shows much promise, but still quite unproven.

Fez + Fedora

Greenstone

- Well-known open source digital library software, funded largely by UNESCO for digital libraries in third-world countries.

- Good support for compression and full-text indexing of documents.

- Alas, it does not support submission of documents – only works for reasonably static collections

  - Not really suitable for ongoing management of documents.

  - Research has been done on making it IR-aware - next version (4.0?) may be different …

Greenstone

Research 1/2

- Import metadata/files into DSpace
  - Student assignment to migrate metadata/content.
  - Based completely on OAI-PMH interface.
  - All 15 groups replicated basic EPrints functionality in DSpace with same data set.
- Higher-level services to enhance basic services provided by EPrints/DSpace/Greenstone
  - Ongoing work into component-based digital libraries…
- DSpace<->EPrints<->Greenstone **Interoperability**
  - Ongoing work in/among different groups.
- Study of usability of IR software
  - to make it easier for more widespread adoption

Research 2/2

- New architectures for digital libraries:
  - Components (ODL, SODA, OpenDLib, Dienst)
  - Grid computing (DILIGENT)
  - Cluster computing (FDL)
  - Peer-to-Peer computing (OCKHAM)
- **Preservation**:
  - Lots of Copies Keeps Stuff Safe
  - Trustworthy Digital Objects
- **Rights Management**:
  - OAI-Rights
  - Creative Commons SA/…

References

- Kahn, Robert and Robert Wilensky (1995) "A Framework for Distributed Digital Object Services", CNRI. Available
- Maly, Kurt, Michael L. Nelson and Mohammed Zubair (1999) "Smart Objects, Dumb Archives: A User-Centric, Layered Digital Library Framework", in D-Lib Magazine, Vol. 5, No. 3, March 1999. Available http://www.dlib.org/dlib/march99/maly/03maly.html
- Fedora
  - http://www.fedora.info/
- EPrints
  - http://www.eprints.org/
- DSpace
  - http://www.dspace.org/
- Greenstone
  - http://www.greenstone.org