

# Identifiers and Repositories



hussein suleman  
uct cs honours 2009

# 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

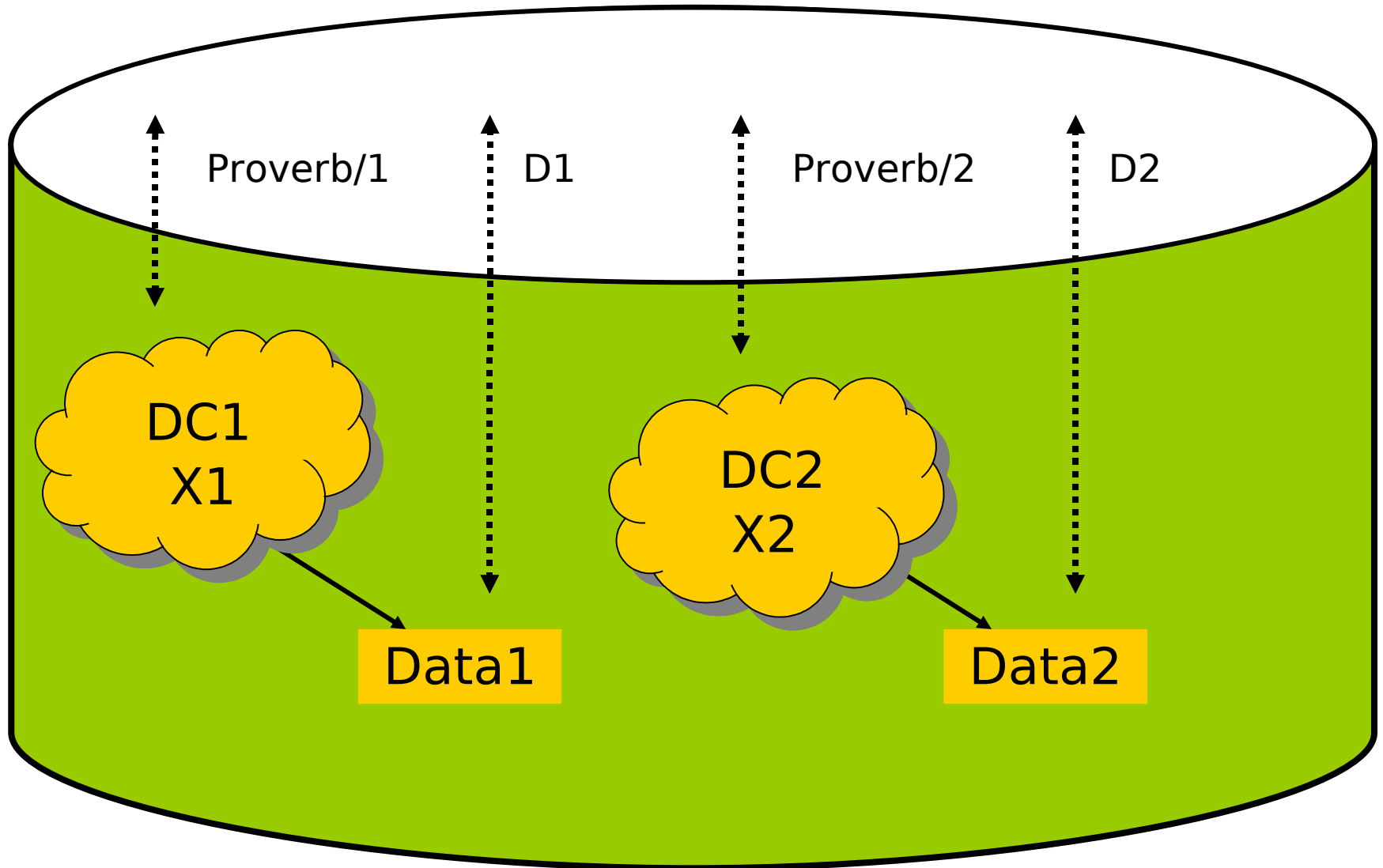
---

| Identifier | MetadataX          | DC  |
|------------|--------------------|---|
| Proverb/1  | Id: P1<br>Data: D1 | Source: Africa<br>Type: Proverb<br>Identifier: D1             |
| Proverb/2  | Id: P2<br>Data: D2 | Creator: Albert Einstein<br>Type: Quotation<br>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

---



[Subscribe](#) (Full Service) [Register](#) (Limited Service, **Free**) [Login](#)

**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](#)

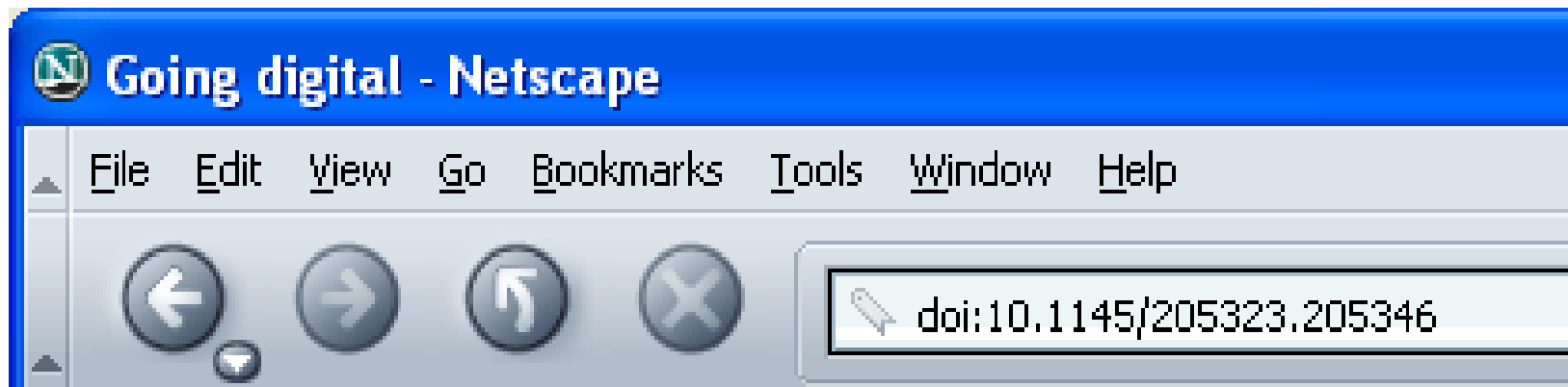
**DOI Bookmark:** Use this link to bookmark this Article: <http://doi.acm.org/10.1145/205323.205346>  
[What is a DOI?](#)

---

# 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.



# 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 Superagui 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-Maximilians-Universität München.](#)

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



# EPrints 3 Interface

## Hussein's Space



[Home](#) | [About](#) | [Browse by Year](#) | [Browse by Subject](#)

[Login](#) | [Create Account](#)

### test document

suleman, hussein (2006) *test document*. PhD thesis, UCT.



[MS Word](#)

### Abstract

test abstract

|                       |  |
|-----------------------|--|
| <b>Item Type:</b>     | Thesis (PhD)   |
| <b>Subjects:</b>      | <a href="#">Q Science &gt; QA Mathematics &gt; QA75 Electronic computers. Computer science</a> |
| <b>ID Code:</b>       | 1  |
| <b>Deposited By:</b>  | Unnamed user with email <a href="mailto:hussein@cs.uct.ac.za">hussein@cs.uct.ac.za</a>         |
| <b>Deposited On:</b>  | 22 Apr 2007 00:27  |
| <b>Last Modified:</b> | 22 Apr 2007 00:56  |

# 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 Greenstone)!

# DSpace Interface



[About DSpace](#)



Search DSpace:

Go

[Advanced Search](#)

[Home](#)

Browse

[Communities & Collections](#)

[Titles](#)

[Authors](#)

[By Date](#)

Sign on to:

[Receive email updates](#)

[My DSpace authorized users](#)

[Edit Profile](#)

[Help](#)

[About DSpace](#)

[DSpace at MIT](#) >

[MIT Press](#) >

[MIT Press Out of Print Books](#) >

## Browsing MIT Press Out of Print Books by Title

Jump to: [0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

or enter first few letters:

Go!


Showing items 1-21 of 74.

[Next page](#)

| Date of Issue | Title  | Authors   |
|---------------|--|---|
| 1986          | <a href="#">Actors : a model of concurrent computation in distributed systems</a>                          | Agha, Gul A.  |
| 1972          | <a href="#">The American business corporation; new perspectives on profit and purpose.</a>                 | Goldston, Eli, 1920- comp. ; Morton, Herbert Charles; Ryland, G. Neal |
| 1971          | <a href="#">The analysis of feedback systems</a>   | Willems, Jan C.   |
| 1973          | <a href="#">An annotated bibliography on microwaves: their properties, production, and applications to</a> | Goldblith, Samuel A. ; Decareau, Robert V.                            |

# Manakin – DSpace Customisation

Login

  
**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.

### Search DSpace

[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



## Fez Digital Repository

Fez (Login) [Browse Communities](#) [Help](#)

### Search

 [help](#)

Use [Advanced Search](#) to search by multiple criteria

### Browse

[by Communities & Collections](#) | [by Subject](#) | [by Year](#) | [by Author](#) | [by Latest additions](#)

### Recently Added Items

(No publicly viewable items published in the last week)

**Fez** is Hussein's Space's institutional digital repository for publications, research, and teaching materials. Deposited material covers a very wide range of subjects and disciplines. The repository holds the full text of many peer-reviewed published articles and conference papers, as well as book chapters, theses and other forms of written research from HS academic staff and postgraduate students. It contains working and discussion papers, technical reports and some preprints. Research datasets, sound and image collections, multimedia, software and other forms of research data are also held. Learning objects and other course-related materials are welcome.

**Fez** provides free, searchable access to openly available HS research, publications and teaching materials and manages their long-term archiving. HS staff and students must log in to deposit materials.

**Fez** is running on **Fez** software. **Fez** is part of a wider project entitled [eScholarshipUQ](#), which is the local testbed for the [Australian Partnership](#)

# 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

Titles - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://nzdl.sadl.uleth.ca/cgi-bin/library?e=d-00000-00---off-0aginfo--00-0

HOME HELP PREFERENCES

titles

search titles subjects organizations

English

A-B-C-D-E

- Agricultural and rural development policy in Latin America. New directions and new challenges. (FAO Agricultural Policy and Economic Development Series - 2) (1997)
- Agricultural market information service union of Myanmar (1999)
- Animal feeding and food safety. (FAO Food and Nutrition Paper - 69) (1998)
- Asia Pacific forestry sector outlook: focus on coconut wood. (1997)
- Basic finance for marketers. (Marketing and Agribusiness Texts - 1) (1997)
- Bioenergy for development. Technical and environmental dimensions (1994)
- Biological farming in Europe - Agriculture biologique en Europe (REUR Technical Series 12) (1990)
- Carbohydrates in human nutrition. (FAO Food and Nutrition Paper - 66) (1998)
- Case studies of people's participation in watershed management in Asia (1996)
- Cereal policies review, 1995-97. (1998)
- CERES No. 144 - Low-input farming: merits and limits (1993)
- CERES No. 148 - Working out the links: labor in sustainable agriculture (1994)

Done



# 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>