

Identifiers and Repositories

hussein suleman
uct cs honours 2007

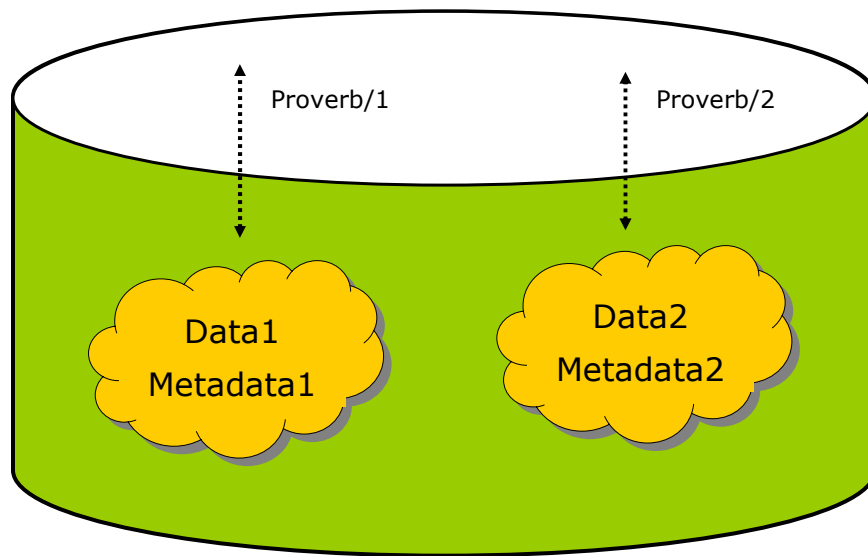
Features of (Good) Identifiers

- 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 1

Identifier	Metadata	Data
Proverb/1	Source: Africa Type: Proverb	When an old man dies, a library burns down
Proverb/2	Author: Albert Einstein Type: Quotation	Imagination is more important than knowledge

Conceptual Model 1

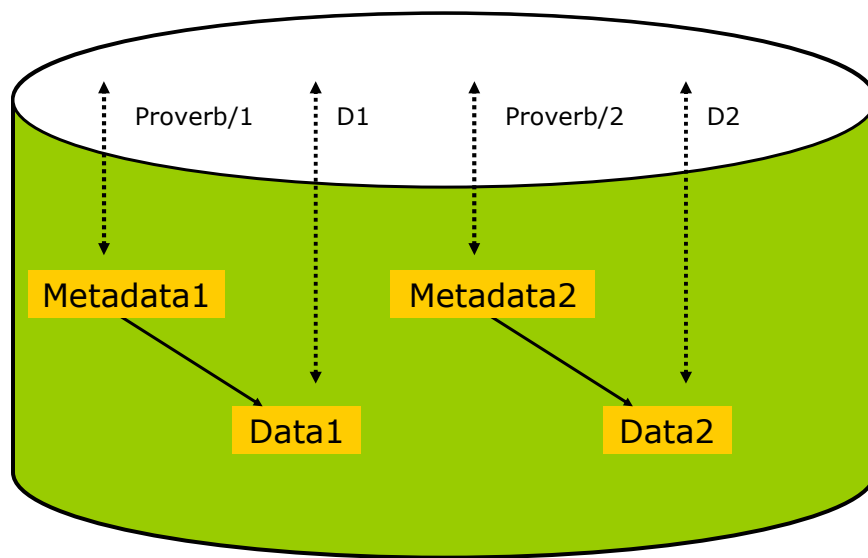


Example Repository 2

Identifier	Metadata
Proverb/1	Source: Africa Type: Proverb Identifier: D1
Proverb/2	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 2

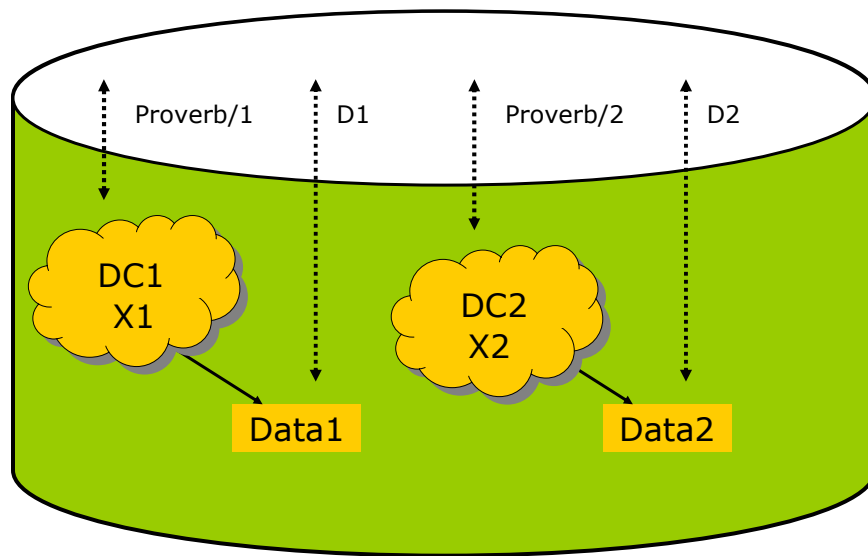


Example Repository 3

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 3



Repository 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

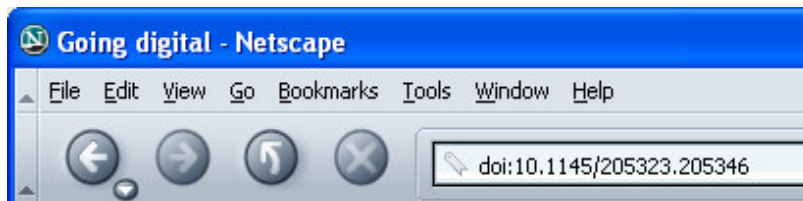
- A handle server stores the association between handles and physical locations of objects.
- Handle servers follow a DNS model:
 - they are distributed and replicated
 - there are global and local servers
 - handles may be cached locally after being resolved to minimise resolution traffic
 - management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

Handle Example

The screenshot shows the ACM Portal website interface. At the top, there is a navigation bar with the ACM logo and the word 'PORTAL' in large letters, with 'University of Cape Town' underneath. To the right of the logo are links for 'Subscribe (Full Service)', 'Register (Limited Service, Free)', and 'Login'. Below the logo is a search bar with the text 'Search:' and two radio buttons: 'The Guide' (selected) and 'The ACM Digital Library'. A 'SEARCH' button is to the right of the search bar. Below the search bar is a green banner with the text 'THE ACM DIGITAL LIBRARY' and links for 'Feedback', 'Report a problem', and 'Satisfaction survey'. The main content area displays the article title 'Going digital: a look at assumptions underlying digital libraries'. Below the title are links for 'Full text' (with a PDF icon) and '220 KB'. The 'Source' section lists 'Communications of the ACM archive', 'Volume 38, Issue 4 (April 1995)', 'Pages: 77 - 84', 'Year of Publication: 1995', and 'ISSN:0001-0782'. The 'Authors' section lists 'David M. Levy' (Xerox PARC, Palo Alto, CA) and 'Catherine C. Marshall' (Texas A&M Univ., College Station). The 'Publisher' is 'ACM Press' (New York, NY, USA). Below the article information are links for 'Additional Information' (abstract, references, citations, index terms, collaborative colleagues, peer to peer) and 'Tools and Actions' (Discussions, Find similar Articles, Review this Article, Save this Article to a Binder, Display in BibTex Format). At the bottom, the 'DOI Bookmark' section provides a link to bookmark the article: 'http://doi.acm.org/10.1145/205323.205346' and a link for '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.
 - see <http://www.fedora.info/>
- ❑ 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.

References

- Kahn, Robert and Robert Wilensky (1995) "A Framework for Distributed Digital Object Services", CNRI. Available <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>
- 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>