

Identifiers

Key points

- What makes something an identifier?
- What do they identify?
- What are the rules?


Identifiers

- Unique
- Consistent
- Persistent

Identifiers must....

- be unique within domain (private db; web)
- be consistent (identifier must always ID the same thing; DO NOT RE-USE!)
- be persistent (must live as long as thing it identifies)
- be in a standard format

Language ambiguity


dog = 

Name ambiguity

Tolkien, J. R. R. (John Ronald Reuel), 1892-1973



Internationalization (i18n)

dog (lang=en) hund (lang=de)
 perro (lang=sp)
 chien (lang=fr)

IDa87nn3

Rules for identifiers

1. Unique... within a domain

Bob has account 43729 at Bank of America 111143279

Mary has account 43729 at Chase 222243279

Rules for identifiers

1. Unique... within a domain

Karen has a web site with file `index.html` `kcoyle.net/index.html`

Google has a web site with file `index.html` `google.com/index.html`

Rules for identifiers

2. Consistent.

Same identifier always identifies the same thing.

Note on “Consistent”

Same thing may get more than one identifier – this happens naturally in the creation of metadata. It’s not a huge problem as long as you have a way of saying that:

A = B

So that you can bring together the identifiers for the same thing.

Rules for identifiers

3. Persistent.

The identifier does not change.

Smith, John, 1946-2009

Rules for identifiers

4. In a standard format

urn:ISBN:1-46492-149-0

http://id.loc.gov/authorities/names/n89613425

} Web

Identifiers: what they identify

- Easier to attach an identifier than understand what it actually identifies.
- ISBN – identifies publisher’s product
- LCCN – identifies LC-created metadata; \= ISBN even though may have very similar metadata to publisher’s
- DOI – identifies item in DOI system, but may link to a general sales page

Identifiers should ...

- link to information about who created and supports them
- lead to information about the thing identified

LC Control No.: 48008547
 LCCN Permalink: <http://lcn.loc.gov/48008547>
 Type of Material: Book (Print, Microform, Electronic, etc.)
 Personal Name: [Polo, Marco, 1254-1323](#)
 Uniform Title: [English, 1948 \[from old catalog\]](#)
 Main Title: The adventures of Marco Polo,
 Published/Created: New York, John Day Co., 1948.
 Description: p. cm.
 CALL NUMBER: [G370 .P72 1948a](#)

Identifier readability


- Opaque: no meaning to the identifier: (just a string)
 - 93IG9484KF
- Readable: makes sense to a human:
 - http://en.wikipedia.org/wiki/J_K_Rowling
- Can be both: system can add readable bit to opaque identifier:
 - http://openlibrary.org/authors/OL23919A/J_K_Rowling

Exercise: Identifiers

three questions:


1. What does it identify?
2. In what context is the identifier unique?
3. What could be done to make it unique “globally” on the Web?

UPC code



ISBN

ISBN-13: 978-0-8352-0000-4
ISBN-10: 0-8352-0000-0



9 780835 200004

E-mail address

johnnie1234@gmail.com

Telephone number

- Tel : +60 6 082 442000 | Fax:6 082 449944 | Email:librarian [at] sarawak [dot] gov [dot] my
- +1 917 275 6975

Telephone Codes



International calling codes sorted by first digit

Telephone codes

02 Singapore
082 Kuching (Sarawak)



Area Code Map for New York Area

Single Area Code Multiple Area Codes Area Code Boundary