Identifiers and identification in the digital age

Graham Bell
Executive director, EDItEUR

Digital Publishing Forum
UCL, 14th May 2014
About EDItEUR

• not-for-profit membership organisation

• develops, supports and promotes metadata and identification standards for the book, e-book and serials supply chains

• based in London, but a global membership of publishers, distributors, wholesalers, subscription agents, retailers, libraries, system vendors, rights and trade associations

• also provides management services to International ISBN / ISTC / ISNI Agencies
Agenda

- principles of identification and identifiers
- ISBN – Stella Griffiths
- ISNI – Andrew MacEwan
- EIDR – Raymond Drewry
- panel Q&A session
- reception in G24 Foster Court
all metadata depends on secure identity
What is an identifier?

- simply, a ‘name’ for an entity or resource
  - can be a physical object, digital object, abstract object or concept, person, organisation, place...

- usually a simple alphanumeric string in a tightly-defined format

- unique within some namespace

- ideally, persistent and meaningless

- each entity or resource is characterised by its ‘minimum referent metadata’
set of characteristic metadata for a particular identifier within an identifier scheme

for an ISBN, includes title, author, edition, format, language, imprint, publisher, pub date...

the metadata attributes define what is unique about this particular entity or resource

if any element of the minimum set changes, then the identifier changes too

minimum referent metadata set encodes exactly what we mean by ‘same’ and ‘different’

depends on the business function of the ID
What is an identifier?

- should be backed by well-defined governance policies and registration procedures
  - to ensure authority and trustworthiness
- metadata registry can be public or private
- increasingly, identifiers are resolvable
- Linked Content Coalition
  - principles of identification
  
http://www.linkedcontentcoalition.org/
Principles of identification

Version 1.1, April 2014
Editors: Naimian Pakian & Geoffroy Voss (geoffroy.voss@rightsc.com)

An identifier is a name which is unique within its type and domain. This document comprises the LCC recommendations for the design and use of identifiers within the digital network in the content and rights supply chain. Detailed support for the recommendations is provided in the attached appendices.

These recommendations are presented as a model of best practice for identification to support the highest level of automation, interoperability, trust and accuracy within the network. They are not mandatory in the sense that none is legally or systematically enforceable for all identifier types, and failure to comply will not normally block the supply chain entirely, but make its operation more time-consuming, labour-intensive and error-prone. “The digital network” here includes, but is not limited to, the internet.

1 Entities: what should be identified?

Public, persistent identification of key supply chain entities is essential.

• Each entity which needs to be recognised distinctly in the digital network should be assigned at least one persistent public identifier so that it may be denoted unambiguously wherever that is required or useful. The entity denoted by an identifier is known as its referent.

• A public identifier is one that is accessible and recognisable by people or machines within the digital network.

• Key entities which require identifiers include each item of content (“creation”) which needs to be recognised (at whatever level of granularity is required), and each party (person or organization) who is recognised as, or claims to be, a contributor or rights holder of content or an assenter of metadata.

• It is desirable for there to be a single standard public identifier for each entity, but where multiple public identifiers exist it is sufficient that they be linked (“mapped”) in a way that enables one identifier to be automatically “translated” to another.

2 Structure: what form should an identifier take?

• The assignment of an identifier always involves some pre-determined general structure, and some element of specific value assignment. The general structure may be as simple as “a ten-digit number” or may have a number of distinct components with different functions, such as a URI prefix, a date of issue, issuer code and check character. The specific value elements may be determined by something as simple as the sequential issue of a number from a range, or as complex as the generation of a digital “fingerprint” derived from the binary structure of the referent. It is normal but not essential that the assignment of an identifier is an automated process.

An identifier may have multiple “designations”.

• The same identifier may take multiple forms or designations to fulfil different functions [for example, the ISBN has had three different designations, as a human readable 10-digit code]
Uniform Resource Identifier – two types...

- Uniform Resource Name
  - specifies the resource based on an identifier
  - doesn’t tell you where to find it

- Uniform Resource Locator
  - specifies the resource based on its location
  - brittle – breaks when location is changed
DOI provides opportunity to combine an existing non-resolvable identifier (with all the benefits of a URN) with a resolvable and actionable link (with all the benefits of a URL)

http://dx.doi.org/10.978.000/7232833

separates management of identity from management of location
A network of identifiers?

ISNI: 0000 0001 2147 9135

ISTC: Ao2-2009-0000154F-A

ISBN: 978 0 00 723283 3

ISBN-A: 10.978.000/7232833