ITU - Telecommunication Standardization Sector
Focus Group on Identity Management: Geneva 13-16 February 2007

Source: Tertius Ltd: Dr. Norman Paskin
(International DOI Foundation and ISO TC46/SC9)

Title: Content Industry Standards Activities

Purpose: For information

Summary

This document accompanies a presentation of the same title to be made at the FG IdM meeting in Geneva 13-16 February 2007.

There is considerable activity within a number of standards development bodies on transacting digital content on digital networks. This requires management of entities including content resources of all forms (physical, digital, abstract); parties (people and organisations) and licenses (permissions specific to users), and their interoperability, and is thus within scope of the ITU FG IdM.

Content industry standards activities are extending their earlier focus on numbering schemes into party identification, licensing, data modelling, and fundamental principles such as interoperability, internet registries, and ontologies. The need for first class naming and the need for semantic interoperability of existing identifier and metadata schemes have led some standards activities to converge on some practical applications of these fundamental principles. Identity management discussions can learn from and use these techniques.

References

Key reference:
"Identifier Interoperability: A Report on Two Recent ISO Activities"
http://www.dlib.org/dlib/april06/paskin/04paskin.html

Supplementary reference:
"Naming And Meaning: Key To The Management Of Intellectual Property In Digital Media"
http://www.doi.org/topics/060922IPDM_China_Paskin_preprint.pdf

Content Industry Standards Activities

A consistent approach to all forms of inter-related entities (parties, resources, licenses, etc) is now recognised as necessary. Identity management is not restricted to individual people, but may include all these, and may include organisations, avatars, pseudonyms etc. Note that the term "identifier" is overloaded and may be used to mean either

- *numbering schemes*: registries, normally with central control, commitment (e.g. ISBN, EAN bar codes, IANA, ITU phone numbering plans etc); or
- *syntax specifications*: normally little central control, few structured attributes, low barriers to entry (e.g. URI (URL); MPEG-21 DII)

In addition there are confusions in terminology as some practical systems use both schemes and specifications (e.g. DOI); there are interactions between schemes and specifications; and "identifier" may mean either the overall system or a specific unique label.

Relevant work in standards development includes:

- ISO content identifiers (http://www.collectionscanada.ca/iso/tc46sc9/)
- DOI (Digital Object Identifier) (http://www.doi.org)

- Music, publishing and licensing content sectors (http://www.editeur.org ; http://www.ddex.net, http://www.the-acap.org/, etc)
- ISO/IEC JTC1 MPEG 21 (http://www.chiariglione.org/mpeg/)
- Party identifiers (e.g. http://www.interparty.org)
- Web-related identifiers such as infoURI and openURL (www.niso.org)

ISO Technical Committee *ISO/TC 46 SC9* manages identifiers familiar to the content and digital library communities, including the International Standard Book Number (ISBN); International Standard Serial Number (ISSN); International Standard Recording Code (ISRC); International Standard Music Number (ISMN); International Standard Audio-visual Number (ISAN) and the related Version identifier for Audio-visual Works (V-ISAN); and the International Standard Musical Work Code (ISWC). Currently it is developing standard identifier proposal for parties (ISPI), and abstract texts (ISTC), as well as standardisation of the Digital Object Identifier system. ISO TC46/SC9 has a current working group on "Identifier interoperability".

The *Digital Object Identifier* system is an application of digital identification on the internet (currently over 26 million such identifiers are in use from over 2000 different users assigners); the DOI is managed by the non-profit membership-based International DOI Foundation.

The *music supply chain* has a number of activities, some of which resulted from a recent Music Industry Integrated Identifiers Project. CISAC co-ordinates a music industry information system (member-based) including the long-established *IPI = Interested Party Identifier* and recent *MWLI = Musical Works Licence Identifier*. DDEX (Digital Data Exchange) has developed messaging standards for music industry chain and its own Party ID. IFPI maintains *GrId = Global Release Identifier*.

The *publishing supply chain* uses ONIX (Online information exchange), EDItEUR/EDIFACT & XML/EDI standards etc maintained by Editeur, an international umbrella body for book industry standards development which works closely with ISBN International and others. ONIX is developing standards for licensing and for multimedia, both of which require a rich semantic interoperability, including *ONIX for Licensing Terms*, and working also with the Digital Library Federation's Electronic Resource Management Initiative (ERMI) to enable standardised statement of usage rights linked with digital resources. It has also developed the shared "RDA/ONIX Framework" to cover cataloguing, digital archiving and preservation projects that have similar requirements to precisely define carrier and content forms.

The recently launched *Automated Content Access Protocol* project is to develop a technical framework which will allow publishers to provide permissions information (relating to access and use of their content) in a form in which it can be recognised and where necessary interpreted by a search engine "crawler", with the aim of enabling a search engine operator (and perhaps, ultimately, any other user) systematically to comply with a policy or licence. It is "being developed as an industry standard by the publishing industry, working with search engines and other technical and commercial partners". ACAP management state that "ACAP would not perceive itself as an identity management project although, of course, in many ways it is…we will certainly have to do something about licence identification and the whole issue of party identity and authentication."

The *MPEG-21 multimedia framework* provides a growing suite of standards for transaction/use relationships between two parties requiring authorization and value exchange of digital items, within the context of a wider set of technical standards for digital interactions. Of particular note are the fundamental "Digital Item Identifier" and the Part 6 "Rights Data Dictionary" for semantic interoperability based on a contextual event-based, extensible, data model http://iso21000-6.net/ . Part 5 "Rights Expression Language" can identify Principals, and Part 15 "Event Reporting" is intended to enable owners of content to receive information about what has happened to their stuff.

Party identifiers are of growing importance. There are some industry-specific standards and activities (e.g. the CIS Interested Party Identifier, and recent work in the scholarly publishing sector on author disambiguation). End-user identification has been seen mainly as an issue of authentication (e.g. Athens, Shibboleth), but identification of individual and corporate persons is a major issue for rights management (and authority control in

libraries). Parties are more than just persons, e.g. organisations, personae, pseudonyms, avatars…The need for a system linking party identifier schemes was a focus of the Interparty project.

The Internet community has been through some debate and confusion regarding *internet identifier specifications*. Confusions seem to centre on a conflation of "indication of the location of the end point", and an "indication of identity"; and on differing views of whether DNS should be optional or required for resolution. The so-called "contemporary point of view" of the URI working group aims at reconciliation, though there are still some different views compared to those coming from practical ontology applications: semantic web work may close this gap. Meanwhile, related work specific to information industries has been developed through NISO, notably the Open URL and "info" URI specifications.

Emerging common themes and approaches

Although this summary of activities may suggest widespread and disparate effort, there are some common themes emerging and common approaches being taken amongst some (though not all) these activities, which offer some insight and guidance in identity management:

(1) A common view (and fundamental data model) of interoperability requirements for identifiers and metadata*, arising from the <indecs> (interoperability of data in e-commerce) view.*

Indecs provided the following principles:

- *Unique Identification*: every entity should be uniquely identified within an identified namespace.
- *Functional Granularity*: it should be possible to identify an entity whenever it needs to be distinguished [first class]
- *Designated Authority*: the author of an item of metadata should be securely identified.
- *Appropriate Access*: everyone requires access to the metadata on which they depend, and privacy and confidentiality for their own metadata from those who are not dependent on it.
- *Definition of metadata*: An item of metadata is a relationship that someone claims to exist between two referents (description)

It delivered a generic data model of e-commerce all types of intellectual property which led, through other projects, to a contextual ontology architecture. This detailed extensible data model supports *semantic interoperability* in many of the developments noted above including:

- ISO MPEG-21 Rights Data Dictionary
- DDEX Digital Data Exchange standards
- ONIX messaging schemas
- ONIX for Licensing Terms, Repertoire and Distribution
- Digital Library Federation ERMI communication of licence terms
- DOI Data Dictionary
- RDA/ONIX common framework
- ACAP: Automated Content Access Protocol
- Rightscom's OntologyX

and is consistent with other major ontology work such as FRBR, ABC-Harmony, OWL, CIDOC CRM, etc.

(2) ***A common view (and fundamental data model) of identifier resolution,*** *to support internet registries and distributed resolution through first class naming and appropriate functional granularity.*

Pre-eminent among these, the *Handle System* offers an ideal choice to provide resolution for all identifiers. DOI is a prime current implementation of this, though schemes that don't wish to use DOI could use a separate handle implementation. Among other advantages, existing numbering schemes may be made actionable and interoperable by converting to handle form; and linking data values in handle records can express relationships. DOI is currently working with ISBN on such an approach.