

IDENTIFIERS AND LICENSING INFORMATION

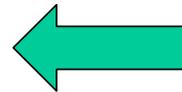
Overview and introduction

WIPO
Seminar on Rights Management Information
Geneva, September 17 2007

Norman Paskin

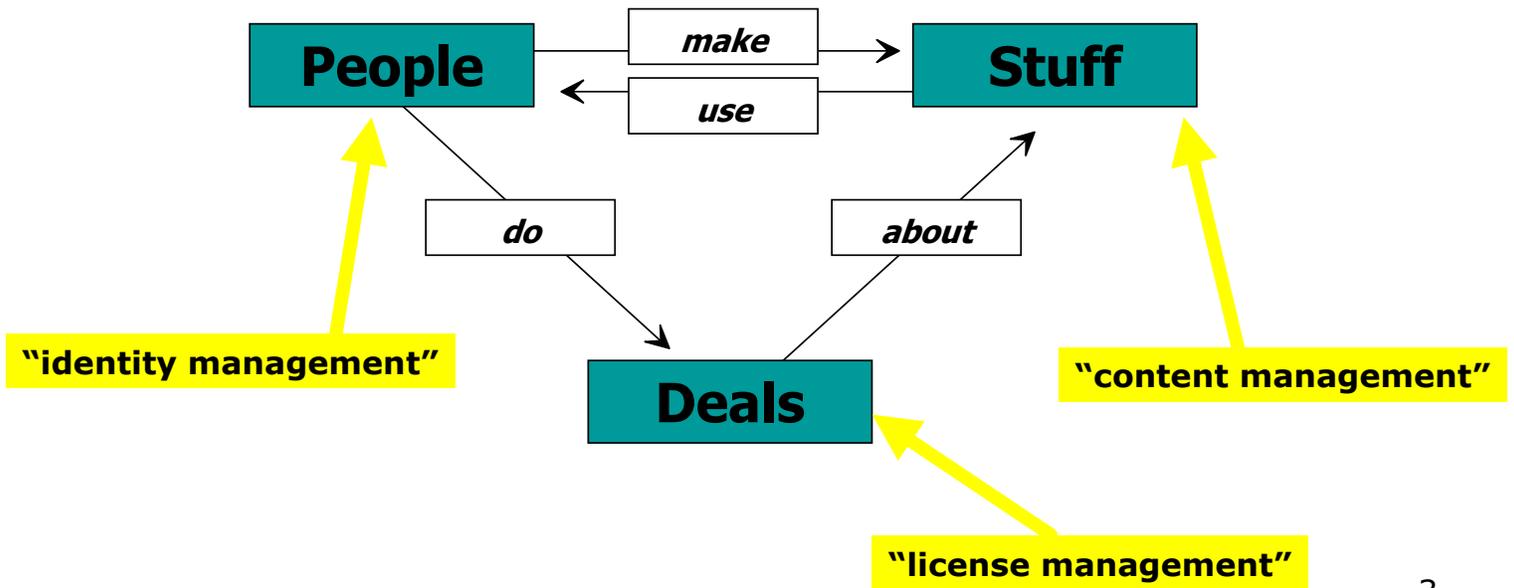
Outline

- Why “identifiers” as the keynote?
- What do we need identifiers to do?
- Major relevant identifier activities
- How can they work together?
- Conclusions



Rights Management Information

- DRM: Technical Protection Measures which use RMI
- But: simple management WITHOUT technical protection also needs RMI
- What is being managed for any rights purpose has to be identified
- We need to accommodate existing and new identifier schemes
- A consistent approach to all kinds of inter-related entities is necessary:



Describing rights using data

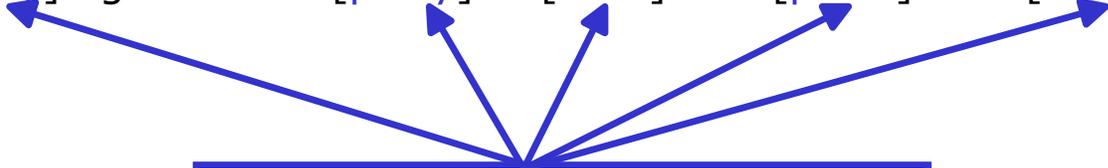
Primary rights events (claims, deals) are described using pieces of data from all these domains:

Rights Statement ("claim"):

[party] owns [right] in [creation] in [time] and [place]

Rights Agreement ("deal"):

[party] agreed with [party] in [time] and [place] that [event]



Pieces of "rights metadata" used in each rights statement are things which need to be identified

Describing rights using data

Primary rights events (claims, deals) are described using pieces of data from all these domains:

Rights Statement ("claim"):

[party] owns [right] in [creation] in [time] and [place]

Rights Agreement ("deal"):

[party] agreed with [party] in [time] and [place] that [event]

Creations typically have standard identifiers, which may have associated structured data, or which may act as keys to get this data

Other pieces of data also need standard identifiers (time, party..)

Describing rights using data

Secondary rights events (licences) are also described using pieces of data:

Permission: [party] can [verb] [amount] to [creation] at [time] in [place].

Prohibition: [party] can't [verb] to [creation] at [time] in [place].

Requirement: [party] must [verb] [amount] to [creation/party] at [time] in [place].

Rights Transfer: [party] can [grant right] to [party] in [creation] at [time] in [place].

Describing rights using data

Pieces of "rights metadata" used
in each rights declaration

Permission: [party] can [verb] [amount] to [creation] at [time]
in [place].

Prohibition: [party] can't [verb] to [creation] at [time] in [place].

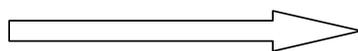
Requirement: [party] must [verb] [amount] to [creation/party]
at [time] in [place].

Rights Transfer: [party] can [grant right] to [party] in [creation]
at [time] in [place].

What are these pieces of "rights metadata"?

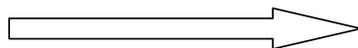
A mix of data from many sources:

1 Rights "events"



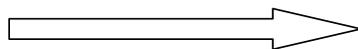
Statements, agreements, transfers, permissions, prohibitions, requirements, assertions, approvals...

2 Descriptive metadata



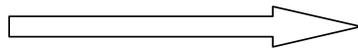
Creations, creation types, contributor roles, user roles, tools, classifications, measures ...

3 Legal terms



Rights, persons, companies, intellectual property, jurisdictions ...

4 Financial metadata

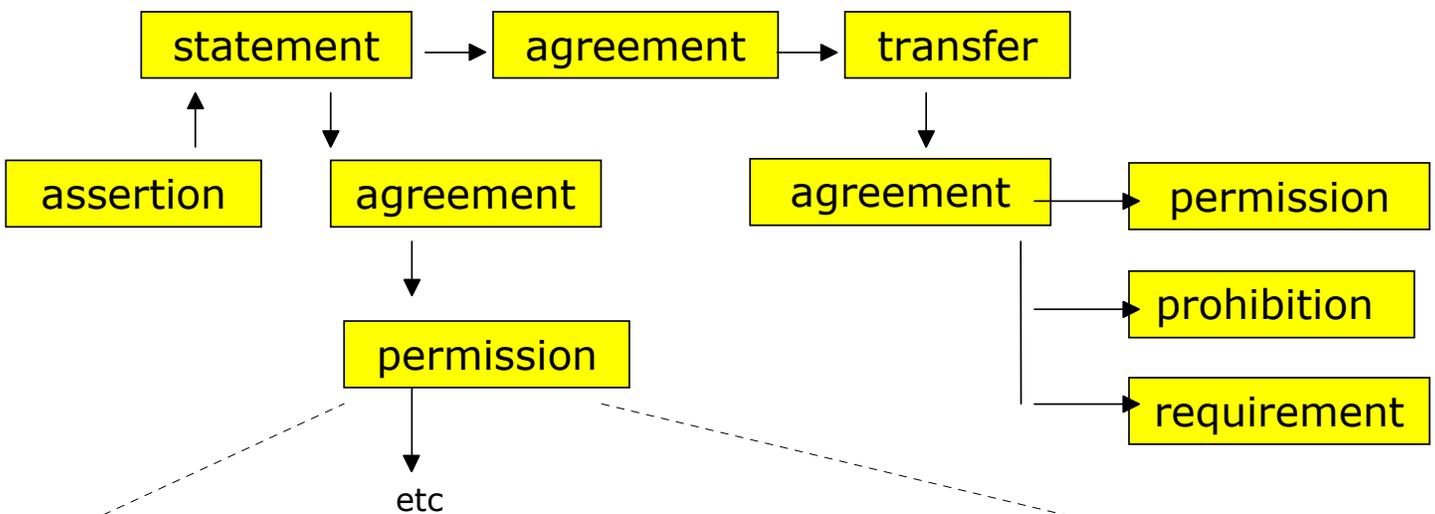


Terms, currencies, conventions...

These sets of "rights metadata" are standardized and maintained in different places.

Distributed rights management

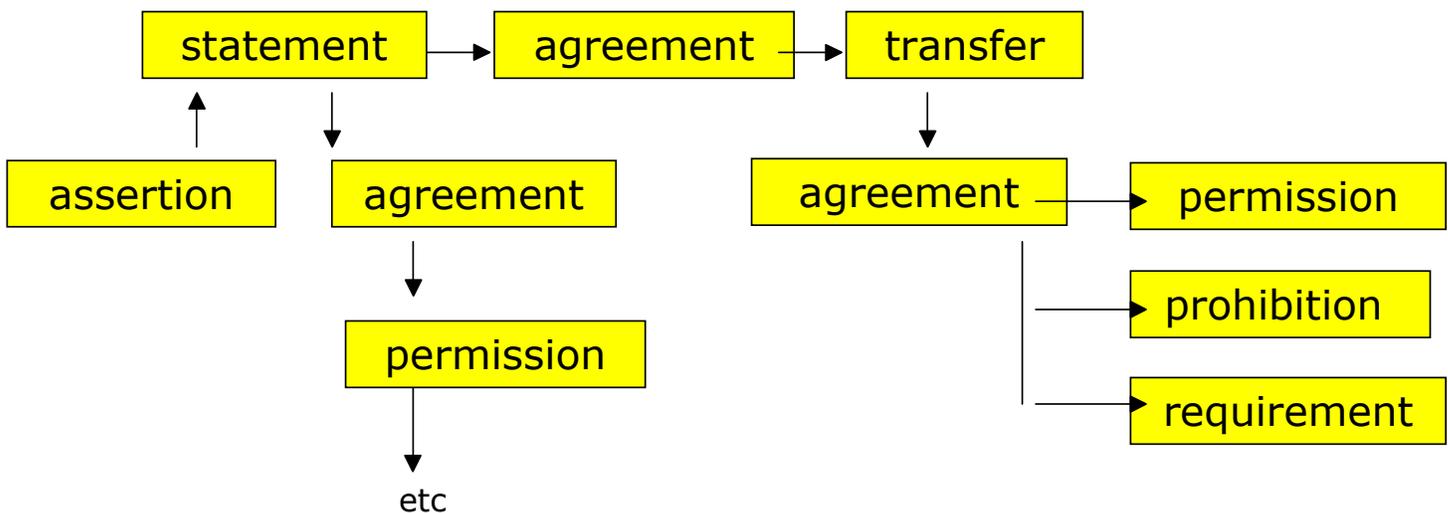
This **mix of data from many sources** is used in many different places by different people in chains of rights events:



[party] can [verb] [amount] to [creation] at [time] in [place].

Compound entity can be expanded to reveal more data

Distributed rights management

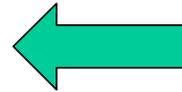


Each of these is an information object:

- which needs to be identified (and may be a compound object);
- which may need to **link** to or **use** information objects in **other** databases;
- which should be interoperable

Outline

- Why “identifiers” as the keynote?
- What do we need identifiers to do?
- Major relevant identifier activities
- How can they work together?
- Conclusions



A “pointer” is not enough

- Suppose I find an identifier which resolves to:
 - a *pdf version of Defoe’s “Robinson Crusoe” issued by Norton.*
- Is it an identifier of:
 - All works by Daniel Defoe?
 - The work “Robinson Crusoe”?
 - The Norton edition of “Robinson Crusoe”?
 - The pdf version of the Norton edition of.... ?
 - The pdf version of that held on this server...?
- Most items of interest are compound objects, simultaneously embodying several referents
 - Multiple identifiers may be necessary (cf music CDs)
- Precisely what is being named?
- You must know (say) **WHAT is being identified**
 - Identifiers assigned in one context may be encountered, and may be re-used, in another place or time - without consulting the assigner. You can’t assume that your assumptions made on assignment will be known to someone else.

Granularity

- **Granularity:** the extent to which a collection of information has been subdivided for purposes of identification (e.g. a collection; a book; tables and figures)
 - *Functional Granularity:* it should be possible to identify an entity whenever it needs to be distinguished
- Your functional granularity may not be my functional granularity:
 - A wants to distinguish “this book in any format”, but B wants to distinguish “the pdf version” from “the html version”, etc”
- “It is a fundamental of almost any statistic that, to produce it, something, somewhere has been defined and identified. Never underestimate how much nuisance that small practical detail can cause. First, it has to be agreed what to count.... In maths numbers seem hard, pristine and bright, neatly defined around the edges. In life, we do better to think of something murkier and softer”
 - “*The Tiger That Isn’t: Seeing Through a World of Numbers*” (2007) Blastland & Dilnot
- You must know (say) **PRECISELY WHAT** is being identified

First class naming

- Many of the items we manage should be treated as “First-class objects”
- First class = having an identity independent of any other item.
 - A key concept of Digital object architecture (e.g. Handles)

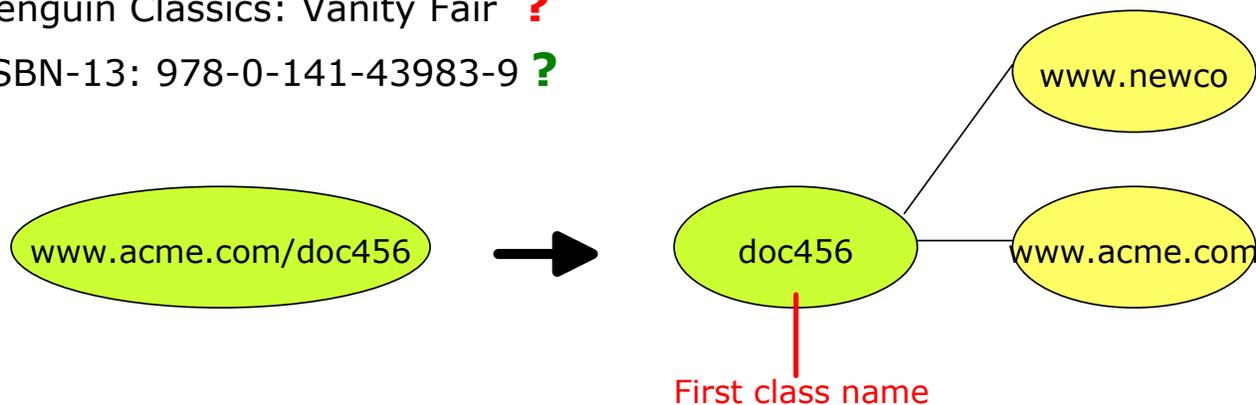
www.acme.com/document456 ?

Document456 ?

Vanity Fair ?

Penguin Classics: Vanity Fair ?

ISBN-13: 978-0-141-43983-9 ?



Context

- Example 1 “the change [in making music DRM free] is not technically complicated. What is complicated is the many **different versions** required to service digital stores, from iTunes to mobile phone downloads. **There are only three basic file formats in use - AAC, MP3 and WMA – but details like bit rates and the metadata identifiers are different for each store. There are 63 variants for mobile devices alone, and overall there are hundreds.**”
 - Scott Cohen, founder “The Orchard” quoted
<http://technology.guardian.co.uk/weekly/story/0,,2065072,00.html>
- Example 2 Person A finds an identifier: it is of an item A already has a library license for. Person B does not have a library license for it. *To where should that identifier resolve?* (“The appropriate copy problem”)
- **Context of use will determine what other things need to be identified**

Infrastructure

Any naming scheme must specify:

- **syntax**: the permissible string of bits for an the identifier;
- the scheme that determines how those bits are **resolved** to some entity; and
- The assumptions for determining **how to interpret anything** that may be found by this process.

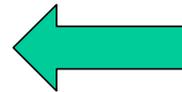
(John Sowa <http://ontolog.cim3.net/forum/ontolog-forum/2007-04/msg00030.html>)

Any naming scheme needs to be implemented:

- **application rules**
- **social infrastructure**
- **technical infrastructure**

Outline

- Why “identifiers” as the keynote?
- What do we need identifiers to do?
- Major relevant identifier activities
- How can they work together?
- Conclusions



ISO content “identification numbering”

<http://www.collectionscanada.ca/iso/tc46sc9/>
Information and Documentation - Identification and Description

ISO 2108	International Standard Book Numbering (ISBN)
ISO 3297	International Standard Serial Number (ISSN)
ISO 3901	International Standard Recording Code (ISRC)
ISO 10957	International Standard Music Number (ISMN)
ISO 15706	International Standard Audiovisual Number (ISAN)
ISO 15706-2	Version identifier for Audiovisual Works (V-ISAN)
ISO 15707	International Standard Musical Work Code (ISWC)
ISO 21047	International Standard Text Code (ISTC)

**Defining metadata now a requirement for each identifier scheme:
entities must be described as well as named**

Some current ISO TC46/SC9 activities

- **International Standard Name Identifier (ISNI)**
 - ISO Project 27729
 - “a new international identification system for the parties (persons and corporate bodies) involved in the creation and production of content entities”.
 - Work on the ISPI (now ISNI) project began in August 2006
- **Digital Object Identifier (DOI) System**
 - ISO/WD 26324
 - To standardise the existing DOI system (syntax is already a national US standard, NISO Z39.84)
 - One application of the *Handle System*
 - adds to it additional features – social and technical infrastructure, policies, metadata management
- **Identifier Interoperability working group**
 - Informal group
 - To consider what steps are necessary to improve interoperability of existing and future ISO TC46/SC9 identifiers
 - “Identifier Interoperability: a report...” <http://www.dlib.org/dlib/april06/>

Music supply chain

CISAC = Int. Confederation of Societies of Authors and Composers

- Co-ordinates a music industry information system (member-based)
- **IPI = Interested Party Identifier** ("which John Williams?")
- Long established system
- Recent **MWLI: Musical Works Licence Identifier***

DDEX = Digital Data Exchange*

- <http://www.ddex.net>
- Messaging standards for music industry chain
- Modelled on earlier publishing industry efforts (ONIX) etc
- Has its own **Party ID** (<http://ddex.net/evaluation/licenceform.html>)

GrId = Global Release Identifier

- for digital tracks etc*.

* Spun out from Music Industry Integrated Identifiers Project (Mi3p)

Publishing supply chain

ONIX = Online information exchange

<http://www.editeur.org/>

- Editeur: International umbrella body for book industry standards development
- ONIX is developing standards for *licensing* and for *multimedia*, both of which require a rich semantic interoperability,
 - ONIX for Licensing Terms: need for **license terms** to be expressed in standard processable format
 - Digital Library Federation Electronic Resource Management Initiative (ERMI) working with NISO and EDItEUR to enable standardised statement of **usage rights linked with digital resources**
- RDA (Resource Description and Access – new AACR); shared “RDA/ONIX Framework for resource categorisation”
 - <http://www.dlib.org/dlib/january07/dunsire/01dunsire.html>
 - Cataloging, Digital Archiving and Preservation projects have similar requirements

Automated Content Access Protocol

- The ACAP project
<http://www.the-acap.org/>
- Recently launched; wide participation
- “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”,
- “the availability or otherwise of [standard methods of identification of content, licenses, systems and business partners](#) are key issues for ACAP. Identification is crucial for [authentication of systems and partners](#) as well as for location of content and licenses.”*

MPEG 21 (ISO/IEC 21000)

Moving Picture Experts Group (MPEG) working group of ISO/IEC

- Builds on MPEG standards MPEG 1,2,4,7..
- MPEG 21: The "Multimedia Framework"

18 standards under various categories:

- "Digital Item" Identification
- Intellectual Property Management and Protection
- Terminals and Networks
- Digital Item Management and Usage
- Digital Item Representation
- Event Reporting

Includes:

- Digital Item Identifier (specification)
- MPEG 21 Rights Data Dictionary
- MPEG 21 Rights Expression Language

Principles used in some proposed DRM systems:

- Moebius (CISAC)
- Digital Media Project (DMP)

Identifying “parties”

- People, organisations, etc
- Some industry-specific standards
 - e.g. CIS IPI system (availability only to membership)
 - Current publisher work on author and institute disambiguation
 - PLUS (Picture licensing: will issue IDs for each party, license and image)
- Impractical to identify everybody
 - Privacy and economic concerns
- End-user identification mainly an issue of *authentication*
 - *Binding of an identifier to a person*
- A major issue for rights (and authority control in libraries)
- Parties are more than just persons
 - Organisations, personae, pseudonyms, avatars...
- <indec> proposed a “directory of parties” linking person identifier schemes

- ISO: ISNI (simple registration number)
- ITU: Identity Management Focus Group (wider issues)
- Others: http://www.ituwiki.com/index.php?title=Living_List_of_Identity_Management_Forums.
- ITU/ISO/IEC/EU FIDIS workshop Lucerne Sept 30
 - <http://www.itu.int/md/T05-TSB-CIR-0163/en>

Web-related identifiers

- URI, URL and URN

- Not sophisticated enough alone for rights management
- Additional techniques: PURLs, RDF, SW, ARK, Handle, etc

- Related work specific to information industries through NISO:*

- Open URL

A syntax to create web-transportable [packages of metadata and/or identifiers](#) about an information object.

- Not an identifier, but a complementary technology for appropriate redirection of identifier resolution

- e.g. in use with URLs, Digital Object Identifiers (DOI)

- "info" URI Registry

Turn legacy identifiers (e.g. info:lccn/2002022641) into URLs

- IETF RFC 4452: The "info" URI Scheme for Information Assets with Identifiers in Public Namespaces. <http://info-uri.info/>

Abstract identifiers

- ITU/ISO “Object identifiers”: OIDs
 - First class name schemes
 - IETF Domain Name System emerged from same 1970’s effort
 - Software, telephone number schemes, broadcasting, etc
 - Abstract: i.e. no specification for defining metadata or resolution
 - e.g. 1.2.276.0.76.3.1.8 = “siemens medical solutions”
 - OID construct allows for network objects, “human objects”
 - <http://www.alvestrand.no/objectid/>
- Many, many other schemes with similar abstract or partial functionality
 - GUID, UUID, etc (e.g. http://en.wikipedia.org/wiki/Category:Universal_Identifiers)
 - Not useful for Rights Management without additional functionality (description, resolution, etc).
- Many, many other schemes for specific purposes
 - National Bibliography Numbers, etc
 - All of which could be part of some rights expressions

Possible needs for WIPO activities

·Limitations and exceptions

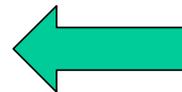
- public domain content;
- specific jurisdiction exceptions: e.g. orphan works, compulsory public licence schemes
 - e.g. CC licences may allocate work to public domain
 - CC licences also need (as with any other expression) to identify relevant entities from a multitude of sources
- no separate identification scheme exists (or is needed): **an exception rule is an attribute of an entity, like any other**

·Technical expression of limitations and exceptions

- technology can help, but **full automation of licence by jurisdiction is very unlikely**
 - pointer to the appropriate “legal namespace”
 - possible simple codification of this right (e.g. ACAP?)
- information about these could be available (attached to / redirected from) a licence identifier (or resource identifier)
- “appropriate copy” technologies could be helpful: “appropriate jurisdiction for this entity in this context”

Outline

- Why “identifiers” as the keynote?
- What do we need identifiers to do?
- Major relevant identifier activities
- How can they work together?
- Conclusions



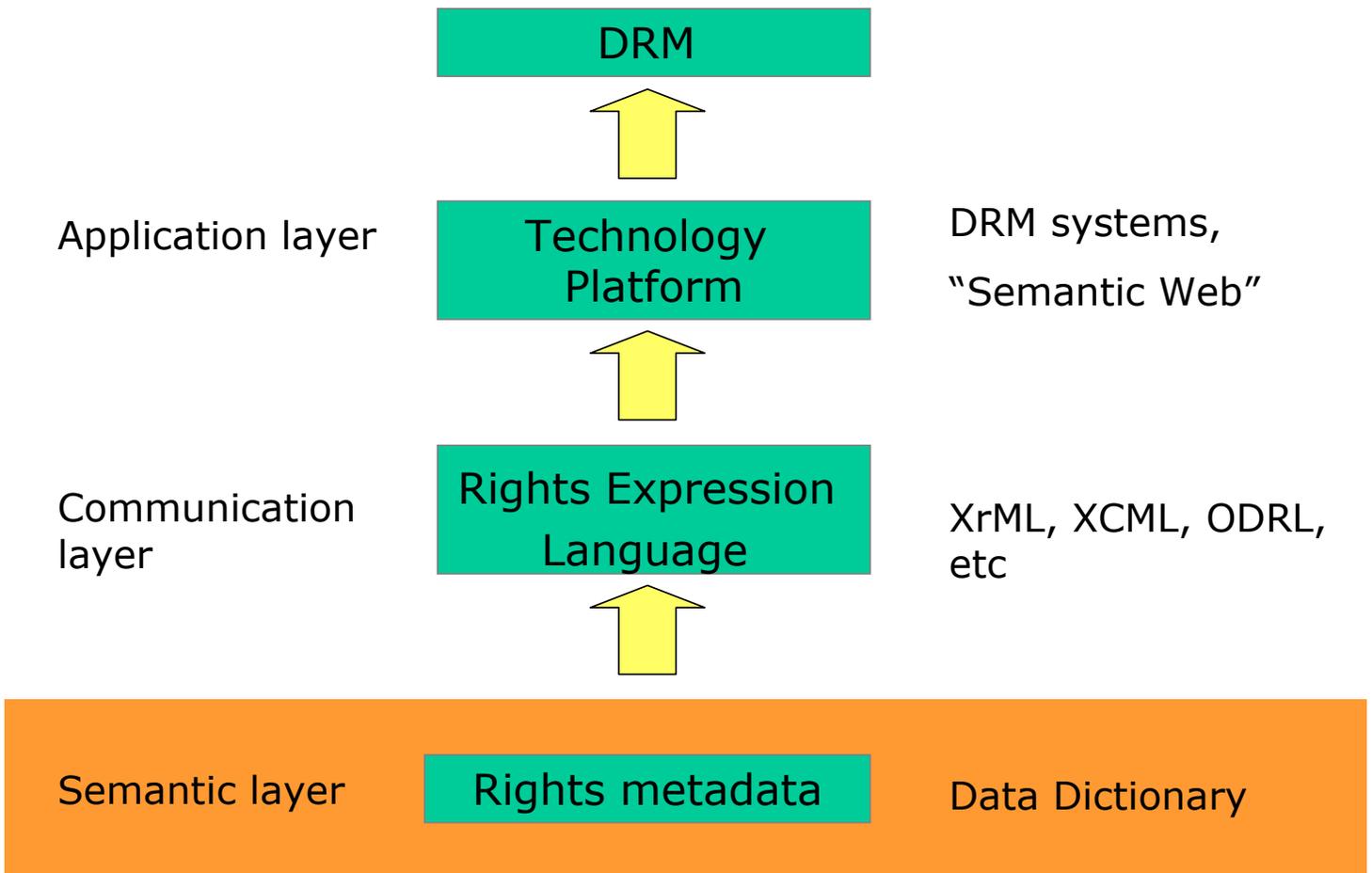
Meaning

- Assigning metadata to a referent, to enable **semantic interoperability**
 - “say what the referent is” (what is the thing that is identified)
- **Semantic:**
 - Do two identifiers from different schemes actually denote the same referent?
 - If A says “owner” and B says “owner”, are they referring to the same thing?
 - If A says “released” and B says “disseminated”, do they mean different things?
- **Interoperability:** the ability for identifiers to be used in services outside the direct control of the issuing assigner
 - Identifiers assigned in one context may be encountered, and may be re-used, in another place or time - without consulting the assigner. You can’t assume that your assumptions made on assignment will be known to someone else.
 - **Persistence** = “interoperability with the future”

Tools

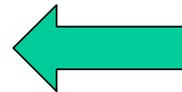
- *Basis:* "Interoperability of Data in E-Commerce Systems" (indecs) : <http://www.indecs.org> 1998-2000
- *Led to Contextual Ontology approach - used in:*
- ISO MPEG-21 [Rights Data Dictionary](http://iso21000-6.net/) (<http://iso21000-6.net/>)
- Digital Object Identifier (DOI) (<http://www.doi.org>)
- DDEX [digital data exchange - music industry](http://ddex.net/) (<http://ddex.net/>)
- ONIX: [Book industry \(+\) messaging schemas](http://www.editeur.org) (www.editeur.org)
- Digital Library Federation - [communication of licence terms \(ERMI: ONIX for licensing terms\)](#)
- Rightscom's OntologyX - [licensee of output, plus own work on tools](http://www.rightscom.com) (www.rightscom.com)
- *informs development of ACAP - Content Access* (<http://www.the-acap.org/>)
- *Can be applied to any other scheme*

Provide a common base semantic layer to build on



Outline

- Why “identifiers” as the keynote?
- What do we need identifiers to do?
- Major relevant identifier activities
- How can they work together?
- Conclusions



Conclusions

- Objects to be identified may be abstract, physical or digital
 - e.g. creations, resources, agreements, people, organisations...
 - need for many new identifiers: Parties, Licences, etc ...
- Your functional granularity may not be my functional granularity: *A wants to distinguish "this book in any format", but B wants "the pdf version, the html version...."*
 - Need to enable different identifiers to work together
 - Through common frameworks
- Most digital objects of interest embody several referents.
 - Compound objects
- Things can be identified at various levels of granularity.
 - An identifier string alone is not enough.
 - You need to say what you are identifying.
- Context of use of an identifier may vary
 - Click to "get the thing that is right for me"
 - Simple resolution of an identifier may not be enough.

Conclusions

- Heterogeneous ecology of identifiers and services
 - Unlikely to be “one rights management service”
 - One identifier may be linked to many services
 - Even if not directly linked, it may be usable in other services
- Services using an identifier may be offered by multiple providers
 - Some may be more definitive than others
 - “Resolution” shades into “query”
- Each registration authority for an identifier scheme must retain autonomy and precedence in determining rules for usage within its own scheme or community.
 - Many early applications will be silos; interoperability is not needed (and may not be desired)
 - New applications will reach across silos (mash ups etc); new silos will appear. As such services grow and become many, interoperability makes sense

Conclusions

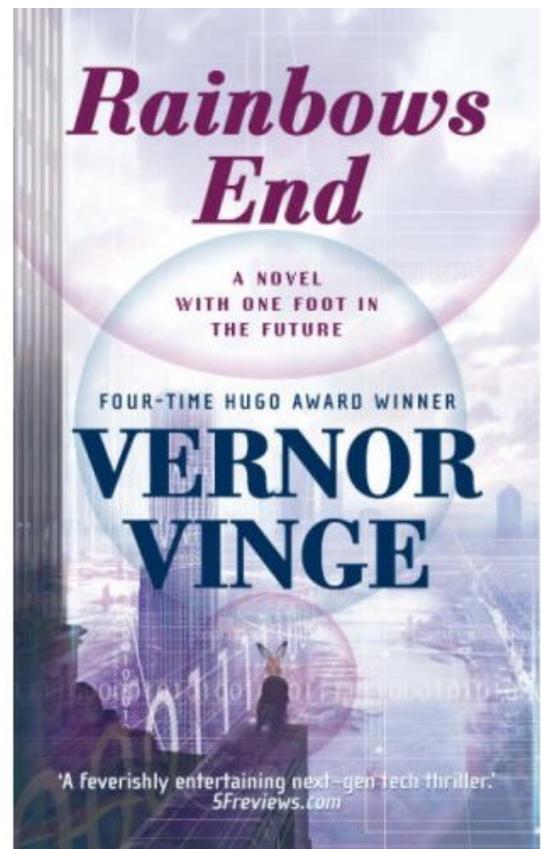
- Content industry standards activities are extending their old focus on numbering schemes
 - into party identification, licensing, data modelling, and fundamental principles
 - interoperability, internet registries, ontologies
- Management of **identifiers** and **metadata**
 - = **Naming** and **meaning**
- Need for first class naming, granularity, etc
 - infrastructure for *extensible distributed* services for using **names** to manage, locate and disseminate **objects**
 - appropriate administrative granularity
 - appropriate social infrastructure
- Need for semantic interoperability
 - **Contexts, roles, relationships** e.g. Contextual ontology (<indec>):
 - functional granularity
 - mapping of existing metadata schemes

A view of the (possible) future...

- “They couldn't get any video, and .. had suffered a 3030 error. Xiu had looked that up; “3030” was a catchall code for a system deadlock caused by licensing conflicts.”

Vernor Vinge

“Rainbow’s End: a novel with one foot in the future” (2006)
(<http://vrinimi.org/rainbowsend.html>)



IDENTIFIERS AND LICENSING INFORMATION

Overview and introduction

WIPO
Seminar on Rights Management Information
Geneva, September 17 2007

Norman Paskin