Panel Title: Linked Data -- Enabling Standards and Other Approaches

Linked Data and Identifiers

Sam Oh
Professor, Sungkyunkwan University, Seoul, Korea
samoh21@gmail.com

ISO TC46/SC9 (Identification & Description), Chair
ISO JTC1/SC34 (Doc Description and Processing Languages), Chair
DCMI Oversight Committee Member

2010 ASIST, Pittsburgh, USA
Talk Outline

• Introduction to Linked Data
• Linked Data and ISO Topic Maps
• Linked Data and ISO Identifiers
• Essential Features of OWL2
Introduction to **Linked Data**

Graham Moore, NetworkPlanet
The Web is changing...

From publishing HTML created for consumption by people to DATA for the consumption by machines
Linked Data: Key Principle 1

URLs are the names of things
For People, products, events, THINGS, any thing.

http://www.networkedplanet.com/people/gra
http://www.networkedplanet.com/products/webplatform
http://www.networkedplanet.com/company/networkedplanet
Linked Data: Key Principle 2

• URLs resolve to ‘Linked Data’ representations in the form of RDF/XML

@prefix person: <http://www.networkedplanet.com/people/>
@prefix company: <http://www.networkedplanet.com/company/>
@prefix model: <http://www.networkedplanet.com/model/>

person:gra foaf:friend person:kal .
The data returned is ‘something that will be possibly of interest’
Linked Data: Key Principle 3

The data returned contains links to other data on the web

prefix @np “http://www.networkedplanet.com/”
np:people/gra np:def/worksfor np:company/networkedplanet

=> http://www.networkedplanet.com/company/networkedplanet
Dereferencing URIs

• Best practice is to deliver an HTML page for humans to understand the ‘thing’ and representations for machines using RDF/XML
Modelling

• One of the biggest challenges of Linked Data is deciding what data to expose and what ‘model’ to define

A model / schema / ontology defines what kind of data will be exposed.

E.g. Person, works-for, Company, has-product, Product
Modelling

Linked Data is only about exposing data and not updating it.

Therefore, the process is about choosing how to expose the raw data.
URI Reuse

• It’s important to try and use existing identifiers

• Mostly in terms of types and properties of models

• But also links between data set entities
Linked Data &
ISO JTC1/SC34 (WG3: Topic Maps)
Subject Identifier and Subject Indicator

A subject is identified via a URL
The URL is called a subject identifier

The URL is the address of a document
That document provides a human-interpretable indication of the identity of the subject
The document is called a subject indicator

Humans use the indicator
By inspecting the document one can be sure that the identifier does not refer to, say, Giacomo’s grandfather Domenico (who was also a composer of operas)

Computers use the identifier
Simple comparison of string values: Identical values mean that the subject is the same

http://psi.onotopia.net/composer/puccini

Giacomo Puccini
Italian composer, b. Lucca 22nd Dec 1858, d. Brussels, 29th Nov 1924. Best known for his operas, of which Tosca is one of the most popular and well-known.
Principles of merging in Topic Maps

• **In Topic Maps, every topic represents some subject**

• **The collocation objective requires exactly one topic per subject**
  
  - When **two topic maps** are merged, topics that represent the same subject should be merged to a **single topic**
  
  - When **two topics** are merged, the resulting topic has the union of the characteristics of the two original topics

...and the resulting topic has the union of the original characteristics
Linked Data & ISO Identifiers

ISO TC46/SC9

(Identification and Description)
The ISBN is the identification system for each product form or edition of a monographic publication published or produced by a specific publisher.

The ISBN is applicable to monographic publications (or their individual sections or chapters where these are made separately available) and certain types of related products that are available to the public.
ISO 3297: IS Serial Number

• The ISSN is a standard code for the unique identification of serials and other continuing resources.

• The ISSN provides a unique identifier for a specific serial or other continuing resource in a defined medium.

• The ISSN describes a mechanism, the “linking ISSN (ISSN-L)” that provides for collocation or linking among the different media versions of the same continuing resource.
ISO 21047: IS Text Code

• The ISTC provides the **efficient identification of textual works**.

• The ISTC provides a means of uniquely and persistently identifying textual works in information systems and of facilitating the exchange of information about those works between authors, agents, publishers, retailers, libraries, rights administrators and other interested parties, on an international level.
ISO 3901: IS Recording Code

- The ISRC defines and promotes the use of a standard code for the unique identification of recordings.

- The ISRC may be applied to audio recordings and music video recordings regardless of whether they are in analogue or digital formats.

- The ISRC shall not be used for the numbering of audio or audiovisual carriers (e.g. compact discs or videocassettes).

- Audiovisual recordings, other than music video recordings produced in conjunction with an audio recording, are excluded from the scope of the ISRC. Such audiovisual recordings should be assigned an ISAN in accordance with ISO 15706.
ISO 15707: IS Musical Work Code

• The ISWC specifies a means of uniquely identifying a musical work.

• The ISWC standardizes and promotes internationally the use of a standard identification code so that musical works can be uniquely distinguished from one another within computer databases and related documentation and for the purposes of collecting societies involved in the administration of rights to such works.

• The ISWC identifies musical works as intangible creations. It is not used to identify manifestations of or objects related to a musical work. Such manifestations and objects are the subject of separate identification systems, such as ISRC for sound recordings, ISMN for printed music, and ISAN for audiovisual works.
ISO 15706: IS Audiovisual Number

• The ISAN establishes and defines a voluntary standard numbering system for the unique and international identification of audiovisual works.

• An ISAN identifies an audiovisual work throughout its life and is intended for use wherever precise and unique identification of an audiovisual work would be desirable.

• An ISAN is applied to the audiovisual work itself. It is not related to the physical medium of such an audiovisual work, or the identification of that medium.
ISO 27729: IS Name Identifier

An example of how ISO identifiers and others can work together
ISO 27729: IS Name Identifier

- The ISNI identifies “Public Identities used publicly by parties involved throughout the media content industries”
- In the ISNI system, parties may be natural, legal or fictional.
ISO 27729: IS Name Identifier

Individual person (Party) vs Public Identity (Name)

Stefani Germanotta vs Lady Gaga
ISO 27729: IS Name Identifier

Natural Person Metadata set:

• ISNI Number
• Name
• Date of Birth
• Place of Birth
• Reference resource
• Class / Role / URI
ISO 27729: IS Name Identifier

Example:

- ISNI 1234 6834 9573 0495
- Lady Gaga
- 28 March 1986
- New York, USA
- “Poker Face”
- Musical_Work / Author / www.ipi.net/isni?1236483
- Sound_Record / Performer / www.ipda.org/isni?1236483
ISO 27729: IS Name Identifier

Technical Architecture
ISO 27729: IS Name Identifier

A Bridge Identifier

Public Layer

ISNI

Private Layer

IPI

IPN

VIAF
ISO 27729: IS Name Identifier

Technical Architecture

1. Registration Agencies
2. Central ISNI DB
3. VIAF
4. External Databases
5. Verification Agencies
6. Resources databases
7. External databases
8. Manual verifications
International Standard Name Identifier

Technical Architecture

• **OCLC will act as the Assignment Agency**
  - Interface with Registration Agencies
  - Cross-repertoire matching
  - ISNI Numbers allocation

• **CISAC will act as a Verification Agency for:**
  - Musical Works Creators (Phase 1)
  - Audio-Visual Works Creators (phase 2)
  - Visual Arts Creators (phase 3)
International Standard Name Identifier

Initial Load
Performed by OCLC on behalf of ISNI-IA

• Cross match of
  – The VIAF database.
  – The IPI database.
  – The IPDA database.
  – The Bowker “Books in Print” Database.
  – The ProQuest Researcher Database.
  – The British Library / JISC Names project Database
  – The ALCS, Prolitteris and any other IFRRO members’ database.

If the same record appears in 2 databases an ISNI is assigned.
Introduction

The International Standard Name Identifier (ISNI) is a draft ISO Standard (ISO 27729) whose scope is the identification of Public Identities of parties; that is, the identities used publicly by parties involved throughout the media content industries in the creation, production, management, and content distribution chains. The ISNI system uniquely identifies Public Identities across multiple fields of creative activity. The ISNI provides a tool for disambiguating Public Identities that might otherwise be confused. ISNI is not intended to provide direct access to comprehensive information about a Public Identity but can provide links to other systems where such information is held.

The Syntax

An ISNI is made up of 10 decimal digits, the last one being a check character.

Examples: ISNI 1422 4568 3573 0476
Digital Object Identifier (DOI)

- DOIs are first class identifiers which can be used in any data model.
- DOIs are Linked Data friendly, since they can be expressed as URIs
creation types

artefact

manifestation

has instance

item

is abstracted to

is expressed in

abstraction

Concepts  "I conceived it"

is fixed in

expression

actions "I did it"

perceived

spatio-temporal

identifiers include...
isbn
upc/ean
umid
doi

atoms, bits  "I made it"

identifiers include...
iswc
isan
issn
pii
doi

Metadata Schema Diagram 11

Version 5.0 ©<indecs> June 2000
Big Picture

International Standard Book Number
Big Picture

International Standard Serial Number
International Standard Text Code

• Consortium between:
  » CISAC
  » Nielsen
  » Bowker
  » IFRRO

• Starts operations in Q1 2009
Big Picture

International Standard Record Code

» identifies “records”
» approx. 12 million ISRCs
» no common database

» CISAC member of the ISRC revision working group
Big Picture

International Standard Music Work Code

- Identification of intangible musical creations
- Central database: CSI
- 18 Millions ISWC assigned

» CISAC is the International Agency
Big Picture

Linking ISRC and ISWC

ISSN

ISTC

ISBN

CSI

ISWC

ISRC

41
Big Picture

International Standard Audio Visual Number

- identifies audio visual works, records,...
- 800,000 ISAN assigned
- common database

- CISAC one of the funding members of ISAN
- ISANIA is the international agency
Big Picture

Linking ISRC and ISAN

ISSN

ISTC

ISBN

CSI

ISWC

ISRC

IDA

ISAN

ISAN
Interested Parties Identifier

» identification of interested parties (in royalties distribution)
» not an ISO standard, but highly used by collective management societies

» 2.5 Millions IPIs assigned
International Standard Name Identifier

- identification of “public identities”
- standard draft
- business model to be defined in 2009

- CISAC one of the 5 founding members
- forecast of 10 million ISNI to be registered in 2010
Overview: Essential Features of OWL2

Ian Horrocks
Oxford University Computing Laboratory
What is an Ontology?
What is an Ontology?

A model of (some aspect of) the world
What is an Ontology?

A model of (some aspect of) the world

- Introduces vocabulary relevant to domain, e.g.:
  - Anatomy
What is an Ontology?

A model of (some aspect of) the world

- **Introduces vocabulary relevant to domain, e.g.:**
  - Anatomy
  - Cellular biology
What is an Ontology?

A model of (some aspect of) the world

- **Introduces vocabulary relevant to domain, e.g.**:
  - Anatomy
  - Cellular biology
  - Aerospace
What is an Ontology?

A model of (some aspect of) the world

- **Introduces vocabulary relevant to domain, e.g.:**
  - Anatomy
  - Cellular biology
  - Aerospace
  - Dogs
What is an Ontology?

A model of (some aspect of) the world

- Introduces **vocabulary relevant to domain, e.g.:**
  - Anatomy
  - Cellular biology
  - Aerospace
  - Dogs
  - Hotdogs
  - ...
What is an Ontology?

A model of (some aspect of) the world

- **Introduces vocabulary relevant to domain**
- **Specifies meaning of terms**
  
  Heart is a muscular organ that is part of the circulatory system
What is an Ontology?

A model of (some aspect of) the world

- **Introduces vocabulary relevant to domain**
- **Specifies meaning of terms**
  Heart is a muscular organ that is part of the circulatory system
- **Formalised using suitable logic**

\[
\forall x. [\text{Heart}(x) \rightarrow \text{MuscularOrgan}(x) \land \\
\exists y. [\text{isPartOf}(x, y) \land \\
\text{CirculatorySystem}(y)]]
\]
The Web Ontology Language OWL

• Motivated by **Semantic Web activity**
  Add meaning to web content by annotating it with terms defined in ontologies

• Developed by **W3C WebOnt working group**
  – Based on earlier languages
    RDF, OIL and DAML+OIL
  – Became a recommendation on 10 Feb 2004

• Supported by **tools and infrastructure**
  – APIs (e.g., OWL API, Thea, OWLink)
  – Development environments (e.g., Protégé, TopBraid Composer)
  – Reasoners & Information Systems (e.g., Pellet, HermiT, Quonto)

• Based on a **Description Logic (SHOIN)**
Experience with OWL

• OWL playing **key role** in increasing number & range of applications
  – eScience, eCommerce, geography, engineering, defence, ...
  – E.g., OWL tools used to identify and repair errors in a medical ontology:
    “would have led to missed test results if not corrected”

• Experience of **OWL in use** has identified restrictions:
  – on expressivity
  – on scalability

  These restrictions are problematic in some applications

• **Research** has now shown how some restrictions can be overcome

  W3C  OWL WG has updated OWL accordingly

  Result is called OWL 2
OWL 2 in a Nutshell

- **Extends OWL** with a small but useful set of features
  - That are needed in applications
  - For which semantics and reasoning techniques are well understood
  - That tool builders are willing and able to support

- **Adds profiles**
  - Language subsets with useful computational properties

- **Is fully backwards compatible with OWL:**
  - Every OWL ontology is a valid OWL 2 ontology
  - Every OWL 2 ontology not using new features is a valid OWL ontology

- **Already supported by popular OWL tools & infrastructure:**
  - Protégé, HermiT, Pellet, FaCT++, OWL API
What’s New in OWL 2?

Four kinds of new feature:

• **Increased expressive power**
  - **Qualified** cardinality restrictions, e.g.:
    persons having two friends who are republicans
  - **Property chains**, e.g.:
    the brother of your parent is your uncle
  - **Local reflexivity** restrictions, e.g.:
    narcissists love themselves
  - **Reflexive, irreflexive, and asymmetric** properties, e.g.:
    nothing can be a proper part of itself (irreflexive)
  - **Disjoint** properties, e.g.:
    you can’t be both the parent of and child of the same person
  - **Keys**, e.g.:
    country + license plate constitute a unique identifier for vehicles
What’s New in OWL 2?

Four kinds of new feature:

• **Extended Datatypes**
What’s New in OWL 2?

Four kinds of new feature:

• **Extended Datatypes**
  – Much wider range of XSD Datatypes supported, e.g.:
    
    Integer, string, boolean, real, decimal, float, datetime, ...


What’s New in OWL 2?

Four kinds of new feature:

• **Extended Datatypes**
  - Much wider range of XSD Datatypes supported, e.g.:
    Integer, string, boolean, real, decimal, float, datetime, ...
  - User-defined datatypes using facets, e.g.:
    ```
    max weight of an airmail letter:
    xsd:integer maxInclusive "20"^^xsd:integer
    ```
What’s New in OWL 2?

Four kinds of new feature:

- **Extended Datatypes**
  - Much wider range of XSD Datatypes supported, e.g.:
    - Integer, string, boolean, real, decimal, float, datetime, ...
  - User-defined datatypes using facets, e.g.:
    - max weight of an airmail letter:
      - `xsd:integer maxInclusive "20"^^xsd:integer`
    - format of Italian registration plates:
      - `xsd:string xsd:pattern "^[A-Z]{2} [0-9]{3}[A-Z]{2}$"`

![Airmail Letter](image1.jpg)

![Italian Registration Plate](image2.jpg)
What’s New in OWL 2?

Four kinds of new feature:

- **Metamodelling and annotations**
  - Restricted form of metamodelling via “punning”, e.g.:
    - `SnowLeopard subClassOf BigCat` (i.e., a class)
    - `SnowLeopard type EndangeredSpecies` (i.e., an individual)
  - Annotations of axioms as well as entities, e.g.:
    - `SnowLeopard type EndangeredSpecies` (**“source: WWF”**)
  - Even annotations of annotations
What’s New in OWL 2?

Four kinds of new feature:

• **Syntactic sugar**
  – Disjoint unions, e.g.:
    Element is the **DisjointUnion** of Earth Wind Fire Water 
    i.e., Element is equivalent to the union of Earth Wind Fire 
    Water Earth Wind Fire Water are pair-wise disjoint 
  – Negative assertions, e.g.:
    Mary is not a sister of Ian 
    21 is not the age of Ian
Alternative Syntaxes

- Normative exchange syntax is **RDF/XML**

```xml
<owl:Class rdf:about="#Heart">
  <owl:equivalentClass>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <rdf:Description rdf:about="#MuscularOrgan"/>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#isPartOf"/>
          <owl:someValuesFrom rdf:resource="#CirculatorySystem"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </owl:equivalentClass>
  <rdfs:subClassOf rdf:resource="#&owl;Thing"/>
</owl:Class>
```
Alternative Syntaxes

- Normative exchange syntax is **RDF/XML**
- Functional syntax mainly intended for language spec

```plaintext
EquivalentClasses(Heart
   ObjectIntersectionOf(ObjectSomeValuesFrom(isPartOf CirculatorySystem)
   MuscularOrgan))
```
Alternative Syntaxes

- Normative exchange syntax is RDF/XML
- Functional syntax mainly intended for language spec
- XML syntax for interoperability with XML toolchain

```xml
<EquivalentClasses>
  <Class URI="Heart"/>
  <ObjectIntersectionOf>
    <Class URI="MuscularOrgan"/>
    <ObjectSomeValuesFrom>
      <ObjectProperty URI="isPartOf"/>
      <Class URI="CirculatorySystem"/>
    </ObjectSomeValuesFrom>
  </ObjectIntersectionOf>
</EquivalentClasses>
```
Alternative Syntaxes

- Normative exchange syntax is RDF/XML
- Functional syntax mainly intended for language specification
- XML syntax for interoperability with XML toolchain
- Manchester syntax for better readability

```
Class: Heart
EquivalentTo: MuscularOrgan
    that isPartOf CirculatorySystem
```
Profiles

• **OWL 2 defines three different tractable profiles:**
  - **EL**: polynomial time reasoning for schema and data
    • Useful for ontologies with large conceptual part
  - **QL**: fast (logspace) query answering using RDBMs via SQL
    • Useful for large datasets already stored in RDBs
  - **RL**: fast (polynomial) query answering using rule-extended DBs
    • Useful for large datasets stored as RDF triples
Concluding Remarks

• The more identifiers are used, the better links will be made available among data.

• We should provide both machine and human-understandable description when an identifier is dereferenced.

• ISO identifiers provide different identification schemes for works, expressions, and manifestations that can be useful in enhancing the quality of linked data.