

Uniform Resource Identifiers In the CIMI Context

Harold Solbrig

Mayo Clinic

Glossary

In this discussion, we use:

- **URI – A Uniform Resource Identifier.**
 - **Some resources, such as web pages, pdf files, mp3's, etc. are digital resources. When this is the case, you can put the URI into a browser..**
- **URL – A Uniform Resource Locator. A URI that, when plugged into a browser, returns a digital resource**
- **(URN) – A URI that isn't a URL. (included for completion)**
- **(IRI) – A URI with a different alphabet. (Not discussed further)**

Brief Introduction

URI – Uniform Resource Identifier

- **Uniform** – single, well understood syntax
- **Resource** –
 - pretty much anything
 - TBL’s definition is circular “anything that might be identified by a URI”
 - (more in a minute)
- **Identifier** – references a unique resource (whatever that me be).

Uniformity

Uniform Syntax:

<scheme name> : <hierarchical part> [? <query>] [# <fragment>]

- Scheme name – see http://en.wikipedia.org/w/index.php?title=URI_scheme
- urn:oid:2.16.840.1.113883.6.96
- <ftp://john:password@mayo.edu:/etc/passwd>
- urn:uuid:D15FABD5-EEBB-4A97-92DB-79584641D3D9
- ...

Scheme names are registered, documented and determine subsequent syntax (!)

The key is that URIs are **SELF CONTAINED** – no need for a second attribute specify format or context

Ready for http and the semantic web!

Resource Identifier

- 1) A resource has identity.
- 2) Leibniz's *identity of indiscernibles* – if all the properties are common, “they” / “it” is the same thing
 - a) We need to separate “essential” (rigid) properties from “accidental” (non-rigid)
 - i) All properties are rigid – identifier is not necessary. The properties *are* the identity
 - ii) All properties are non-rigid – not a resource. Out of scope
 - iii) Some properties rigid, some non-rigid...

Resource (continued)

... some rigid, some non-rigid

- **Strictly speaking, identifiers aren't necessary – the identifier IS the collection of rigid properties..**
- **... But we frequently don't record (or can't even determine) the complete set of rigid properties – we just know they exist.**

Resource (continued)

3) Uniqueness requirement:

- A resource identifier must reference a single unique resource, be it...
 - A person (Alice Jones)
 - A class / category (Automobile, Mammal)
 - A collection (Sharp things made of steel, genes/proteins/bicycle parts)
 - Imaginary things (Unicorns)
 - ...
- Note that this does not imply that a resource can only have one unique identifier... this is desirable but highly unlikely.

Resource Identifier (continued)

4) A resource and its name/description are not identical ... basic result of Leibniz, meaning that they cannot have the same identifier without:

a) violating the “Uniform” requirement, in that the identifier requires context and, as such, is no longer uniform or...

b) Violating the uniqueness requirement.

So What

So... why are you telling us this???

Because:

- 1) The W3C (used to tell us different), arguing that a URI should also be a URL – that we could find Cabernet in the wine ontology...**
- 2) The “C” word... which we should never use because it implies that the description and the thing being described *are* the same thing.**

Why URI's

The Uniform Resource Locator (URI) is the identifier scheme underlying the the Semantic Web

- Resource Description Framework (RDF) data model is “subject-predicate-object”
 - Subject – URI or Blank Node
 - Predicate – URI
 - Object – URI or Literal or Blank Node
- If you want to make an assertion about something in RDF, you have to give it a URI
- If you want to *share* information about something, you need to
 - Come up with a URI equivalence cross map (not good)
 - Agree to use the same URI for the same thing (!)

The Problem (continued)

Unless the interested communities come up with a standardized approach to creating URI's for healthcare terminology ...

... we will be facing the 1990's XML Element all over again.

The Problem

Appendicitis in SNOMED-CT:

http://www.ihtsdo.org/SCT_74400008- perl OWL rendering

<http://purl.oclc.org/snomed/sct#id-74400008> - SNOMED CT in SKOS

<http://purl.bioontology.org/ontology/SNOMEDCT/74400008> - BioPortal

<urn:oid:2.16.840.1.113883.6.96.74400008> – One solution to the OLD problem

<urn:oid:2.16.840.1.113883.6.96:74400008> – another

... many more can be uncovered...

Resources in the CIMI World

Ontological resources – the stuff that instances of CIMI models are about

- **Patients, specimens, locations, anatomical parts, measurements, assessments**
 - **“entities”**
- **Descriptions of these things:**
 - **“Entity description”**
- **Organized description systems**
 - **“Code System” (ontologies, concept systems, ...**
- **Collections of descriptions in organized systems**
 - **“Code System Version”**

Resources in the CIMI World (continued)

Modeling Resources

- **Value Set Definition** – a set of rules that, when resolved against a code system version produce a...
- **Value Set Resolution** – a collection of identifiers that reference resources and, ideally, can be used to locate corresponding descriptions of said resources
- **Value Set** – provenance about a collection of definitions that change over time
- **Concept Domain** – aka “Data Element Concept” – a field in a message, a column in a table
- **Concept Domain Binding** – Association of concept domain with a value set in a given context

Resources in the CIMI World (continued)

Modeling Resources

- **Sources – people and organizations**
- **Formats – XML, TSV, PDF, DOC, ...**
- **Syntax/semantics – OWL / ADL / XML Schema / UML ...**
- **Languages – German, French, ...**
- **(many others)**

**Note: None all of the
URI's below are finalized**

Proposal

In order of precedence:

1) Use official publisher's URI:

- <http://snomed.info/id/900000000000380005> - SNOMED CT Core
- <http://snomed.info/id/900000000000380005/version/20120731> - SNOMED CT Core July 2012 Version
- <http://snomed.info/id/900000000000380005> - SNOMED CT Core
- <http://snomed.info/id/74400008> - Appendicitis (Entity)
- <http://snomed.info/id/447565001> - Simple Reference Set (Resolved Value Set)
- <http://id.who.int/icd/release/9/WHO> - ICD-9
- <http://id.who.int/icd/release/9/CM/2010> - ICD-9-CM 2010
- <http://id.who.int/icd/release/9/CM/540.9> - Appendicitis
- <http://id.who.int/icd/release/10/CM> - ICD-10-CM
- <http://sig.biostr.washington.edu/fma3.0#> - FMA

2) Use community accepted URI:

- <http://www.w3.org/2004/02/skos/core#> - SKOS
- <http://purl.org/dc/terms> - Dublin Core Terms
- <http://purl.org/dc/terms/publisher> - Publisher
- <http://www.w3.org/2001/XMLSchema#> - XML Schema
- <http://www.w3.org/2001/XMLSchema#int> - XML Schema Integer

Proposal (continued)

In order of precedence:

3) Use the RCUI for code system:

- <http://id.nlm.nih.gov/cui/C1136323> (LOINC)
- <http://id.nlm.nih.gov/cui/C1136323/35952-1> (Resection of Appendix - narrative)

4) Use the VCUI for code system version:

- <http://id.nlm.nih.gov/cui/C3260726> (LOINC v 238)

5) Use HL7 OID as URN

- urn:oid:2.16.840.1.113883.5.1 - HL7 AdministrativeGender
- <http://id.hl7.org/codesystem/AdministrativeGender/code/M>
(Male in AG)

5) Use established URI when appropriate

- <http://www.omg.org/spec/MNT#> - Mime types
- <http://www.omg.org/spec/LNG#> - ISO 1766

CIMI URIs

CIMI will need to create URI's for

- Identifiable model artifacts**
- Non standard (and standard?) Data types**
- CIMI Value set Definitions / Resolution**
 - Issue – CIMI could use IHTSDO Simple Refset identifier and mechanism even when contents weren't SNOMED**
- CIMI Value Sets (abstract)**
 - Code system not known or context specific?**

Dereferencing URI's

ALWAYS perform one level of indirection:

F(uri) → URL

Generalizable into: f(id) → URL

Nothing wrong with identity function

f(X) → X

(e.g. f(<http://snomed.info/id/12345>) → <http://snomed.info/id/12345>)

<http://myserver.org/cts2/codesystembyuri?uri=http://snomed.info/id/12345>

Summary

- Proposal above is a “straw man” – a first cut to try to get things going... but the sooner we settle the better.
- URI’s should be “meaning opaque” – different when things are different, not different when they aren’t ... (don’t version everything just because 1% changes...)
- Provenance should be visible:
 - <http://snomed.info/>
 - <http://id.nlm.nih.gov/>
 - -VS-
 - urn:oid:2.16.840.1.113883.13.190
 - urn:uuid:F580628E-EA11-4DD4-BB3F-0B54499D977B
 - <http://purl.org/bioterms/mglt>

Summary

- **Assembly (and disassembly) of namespace/code should be rule based**
 - **http: over urn: over custom:**
 - **Favor “/” over ‘#’ in DNS schemes**

- **Only one identifier change the things that change**

- <http://id.who.int/icd/release/9/2008/A00.0>
- <http://id.who.int/icd/release/9/2010/A00.0>

This says that A00.0 is different (!)

- <http://id.who.int/icd/release/9/code/A00.0>
- <http://id.who.int/icd/release/9/code/A00.0/m2>

This does too, but only impacts those that do change

Note from Lloyd McKenzie

- You're sending a code over the wire. It is a the job of the receiver to determine the meaning of the code once received.

Meaning is determined by dereferencing a code in context – not intended to be fully assigned in the code itself

References

- SNOMED CT URI Specification
- SNOMED CT URI Guide
- <http://www.w3.org/TR/cooluris/#semweb> - section on *URIs for Real-World Objects*
- <http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf>
- <http://www.w3.org/DesignIssues/LinkedData.html>
- ISSUE-14 <http://www.w3.org/2001/tag/group/track/issues/14>
- Choosing between 303 and Hash
<http://www.w3.org/TR/cooluris> - choosing
- <http://informatics.mayo.edu/cts2> - link point for CTS2 Specification (look at resource model)

Special Acknowledgement

**Michael Lawley – editor of SNOMED CT
URI guides for research and references**