Instance linking

Leif Granholm
Trimble
Senior Vice President
BIM Ambassador
Modeling Philosophy

Geospatial:
Geometry/Graphics represents a real world phenomena and is called a feature - mapping history

BIM:
Natively object oriented, conceptual model created first, object can have several geometric representations simultaneously.
Modeling Technology

STEP/EXPRESS (G)/Part 21  ->

UML/XSD/XML  ->

Semantic Web/Linked Data/RDF/OWL/SPARQL
Linked data

*The Web is evolving from a “Web of linked documents” into a “Web of linked data”... (1/2)*

Web of documents... 

Web of linked data...
Linked data

The Web is evolving from a “Web of linked documents” into a “Web of linked data”... (2/2)

- The Web started as a collection of documents published online – accessible at Web location identified by a URL.
- These documents often contain data about real-world resources which is mainly human-readable and cannot be understood by machines.
- The Web of Data is about enabling the access to this data, by making it available in machine-readable formats and connecting it using Uniform Resource Identifiers (URIs), thus enabling people and machines to collect the data, and put it together to do all kinds of things with it (permitted by the licence).

Machine-readable data (or metadata) is data in a format that can be interpreted by a computer.

2 types of machine-readable data:
- Human-readable data that is marked up so that it can also be understood by computers, e.g. microformats, RDFa;
- data formats intended principally for computers, e.g. RDF, XML and JSON.
Linked data

Defining linked data...

“Linked data is a set of design principles for sharing machine-readable data on the Web for use by public administrations, business and citizens.”

EC ISA Case Study: How Linked Data is transforming eGovernment

The **four design principles** of Linked Data *(by Tim Berners Lee)*:

1. Use Uniform Resource Identifiers (URIs) as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).
4. Include links to other URIs so that they can discover more things.
What is BIM?
BIM, software, process, technology or what?

- No practical consensus of what BIM is
- Every presenter still today includes a what is BIM section
- Most common perception of people in the industry is that BIM is software.
- Other often used: Process, Model, Populated Database describing facility

- My favorite: BIM is a Technology
BIM – Human readable docs to Machine readable Structured data

Objects describing building

Functional objects

Physical objects

Logical objects

Abstract objects

Objects describing process

• Schedules
• Resources
• Costs
• Quality assurance
• Tasks

• Work
• Approvals
• RFI:s
• Submittals
• …
Why semantic data?

- Technology development enables automatic use of data
  - Use of data in other softwares, design progression, fm, planning, you name it
  - Robots (good example ping pong playing robot)
  - Augmented reality taken to next step by Google Tango project
  - Drones used for many purposes
URI Uniform Resource Identifier

An URI string is composed of 8 components and 5 parts:

```
foo://example.com:8042/over/there?name=ferret#nose
\_/ \___________\___________\___________\__/
  |           |            |            |        |
scheme     authority       path        query   fragment
\_/ \___________\___________\__/
  |         |        |
userinfo    host     port
```

urn:example:animal:ferret:nose

The URI authority part in itself can be composed of up to 3 parts.

```
john:doe@example.com:8042
\_____\___________\__/
  |            |
userinfo   host    port
RDF Resource Definition Framework (W3C)

RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.
Resource Description Framework (RDF)

- Data model for describing “things” and their interrelations
- Consists of statements about “things” (Web resources) in the form of subject-predicate-object expressions, also known as triples

```
subject       predicate       object
Renee Miller  →  Teaches     →  CSC443
Renee Miller  →  Lives in    →  Toronto
```
Resource Description Framework (RDF)

- Data model for describing “things” and their interrelations
- Consists of statements about “things” (Web resources) in the form of subject-predicate-object expressions, also known as triples

\[
\text{subject} \rightarrow \text{predicate} \rightarrow \text{object}
\]

\[
<\text{uri}> \rightarrow <\text{uri}> \rightarrow <\text{uri}> \text{ or “literal”}
\]

\[
<\text{http://cs.toronto.edu/~miller}> \rightarrow <\text{http://xmlns.com/foaf/spec/#term_based_near}> \rightarrow <\text{http://dbpedia.org/resource/Toronto}>
\]

\[
\text{or}
\]

\[
<\text{http://cs.toronto.edu/~miller}> \rightarrow <\text{http://xmlns.com/foaf/spec/#term_based_near}> \rightarrow \text{“Toronto”}
\]

Uniform Resource Identifier (URI): a string of characters used to identify a name or a resource on the Internet.

Web Ontology Language (OWL)

Class hierarchies are meant to represent structures used in source code that evolve fairly slowly (typically monthly revisions) whereas ontologies are meant to represent information on the Internet and are expected to be evolving almost constantly. Similarly, ontologies are typically far more flexible as they are meant to represent information on the Internet coming from all sorts of heterogeneous data sources.
Linking architectural and structural models
Theoretical basis


Due to the decentralized nature of the Semantic Web, the same real-world entity may be described in various data sources with different ontologies and assigned syntactically distinct identifiers. In order to facilitate data utilization and consumption in the Semantic Web, without compromising the freedom of people to publish their data, one critical problem is to appropriately interlink such heterogeneous data. This interlinking process is sometimes referred to as Entity Coreference.
Multikernel architecture needed

• Real production applications/services usually have their own native (for specific purpose optimized) schema for data that they create/maintain/publish
• Multikernel sw is capable of using native and standard schemas in parallel for functionality.
• Data in standard schemas used directly without conversion to native schema, stored locally or online from publishing services.
• Data in native schema mainly published in standard schemas
• Linked data (semantic web technologies) used for instance level linking between datasets. Linksets can be published independently of data.
Reference model principle

- Application capable of presenting and using native objects and objects in foreign schemas (IFC) simultaneously
- Foreign information not bulk converted to native schema.
System architecture diagram

New Information consumption: Mobile apps
New Information consumption: Web apps
New Information consumption: Machines, robots, drones etc.

Open formats
- IFC
- CityGML
- GML
- KML

Existing desktop applications

Publishing platform:
- Trimble connect
- AD 360
- Nemetcheck BIM+
- Bentley Projectwise

Existing desktop applications

...