

Learning Linked Data/Glossary

This glossary is intended to define terms as they are used within the Learning Linked Data project and especially for its [Inventory of Learning Topics](#). Additional glossaries provide more general definitions, for example:

- W3C Government Linked Data Working Group glossary: <http://www.w3.org/TR/gld-glossary/>
- 3roundstones Linked Enterprise Data (LED) glossary: http://3roundstones.com/led_book/led-glossary.html
- LinkedData.org glossary: <http://linkeddata.org/glossary>
- VIVO Project glossary: <http://www.vivoweb.org/glossary>

Domain Model

In the design of metadata, as in software design, a domain model identifies a set of “things of interest.” In the context of Linked Data, a domain model enumerates the classes of things described by metadata and specifies how those classes relate to one another.

Languages of Description

In the broadest sense, languages of description are controlled sets of data-dictionary elements, attributes, concepts, properties, classes, or thesaurus or vocabulary terms used in library and information management practice for describing things in the world or, more specifically, for describing information resources. The Learning Linked Data Project is interested in languages of description that are natively expressed with, or have been translated into, the language of RDF.

Linked Data

Linked Data is data that fits into a “cloud” of diverse data sources – whether those sources are published world-readably on the Web (Linked Open Data) or behind a corporate or institutional firewall (Linked Enterprise Data). For the purposes of the Learning Linked Data Project, Linked Data is data published in a form compatible with the Resource Description Framework (RDF) model of the World Wide Web Consortium (W3C).

Ontologies

For the purposes of the Learning Linked Data Project, ontologies are formalized conceptualizations constructed using RDF vocabularies. Because in RDF “everything is data” (see RDF Data), ontologies are also considered a (special) form of data. While there is no theoretical limit to how simple an ontology may be, more complex ontologies are typically expressed using the Web Ontology Language OWL – an extension of the RDF language that supports richly expressive constructs of a logical nature. “Ontology engineering” – the construction of complex logical structures in support of sophisticated inferencing, the process by

which new information is automatically derived from the logic of existing information – is beyond the scope of the Learning Linked Data Project.

Open World Assumption

The Open World Assumption – the technical mindset of Linked Data – says that in principle, there may always exist additional sources of data, somewhere in the world, to complement the data one has at hand. Its opposite, the Closed World Assumption, assumes that data sources are well-known and tightly controlled, as in a closed, stand-alone data silo. The Open World mindset is optimized for mashing up data from multiple sources and integrating new sources as they are discovered, whereas the Closed World mindset is optimized for validating data against specific patterns and integrity constraints in the manner of most traditional data applications. Understanding the applicability and interplay of the Open and Closed approaches is a key challenge not just for learners of Linked Data, but for designers of information systems in the Web age generally.

Resource Description Framework (RDF)

The W3C Resource Description Framework (RDF) is a generic model for expressing data – a model optimized to support the merging and aggregation of data from multiple sources. In RDF, a “resource” is any thing, physical or conceptual, that can be named using a Uniform Resource Identifier (URI). RDF conceptualizes data as webs of information nodes, or RDF Graphs. Connections between nodes are expressed using RDF’s language of RDF Triples. RDF is the foundation for a suite of related specifications referred to generically as Semantic Web standards.

RDF Data

For the purposes of the Learning Linked Data Project, RDF data is any information expressed in a form compatible with RDF and thus usable as Linked Data. Learners must grasp the principle that with RDF, “everything is data.” Descriptions of things – both of things in the world or, more specifically, of information resources – are considered “data,” but in Linked Data, the Languages of Description used to describe those things are themselves expressed with RDF and, as such, are also considered to be “data.”

RDF Graphs

Conceptually, RDF graphs are webs of resources linked by named relationships. In visual-diagrammatic terms, graphs are sets of “nodes” (circles and squares) connected using “arcs” (arrows). Each node-arc-node link of a graph corresponds to an RDF Triple. Graphs can be serialized for machine processing using one of several interchangeable concrete syntaxes. Graphs are a bit like paragraphs in natural language inasmuch they consist of multiple statements (RDF Triples), which are analogous to sentences.

RDF Triples

Conceptually, RDF Triples are statements about things – statements that are constructed according to the grammatical model of the Resource Description Framework. Like some simple sentences in natural language, triples each have a subject, a predicate, and an object which, taken together, express a complete thought. Subjects and objects may be instances of an RDF “classes” (somewhat analogous to natural-language nouns), while the predicate relationship between a subject and object is expressed with an RDF “property.” If properties and classes are the words of RDF language (see RDF Vocabularies), RDF Triples are its sentences.

RDF Vocabularies

In Semantic Web usage, RDF vocabularies are sets of “properties” and “classes” used for describing things. Learners may approach Linked Data with other types of “vocabulary” in mind, such as thesauri, which are typically used as value vocabularies. Some such value vocabularies may be expressed in RDF as RDF vocabularies, while others may be expressed as other types of RDF Data, such as SKOS Concept Schemes. If properties and classes are the words of the RDF language, vocabularies provide the RDF language with its dictionaries.

Semantic Web

Semantic Web is a vision first introduced in the late 1990s by Tim Berners-Lee, according to which the distributed and globally accessible infrastructure of the World Wide Web can be leveraged to link not just documents (Web pages), but sources of structured data. In order to realize this vision, the World Wide Web Consortium (W3C) developed Semantic Web standards such as Resource Description Framework (RDF), first published as a W3C Recommendation in 1999. Starting in 2006, the movement to publish data freely on the Web using Semantic Web standards was popularized under the banner of Linked Data.

SKOS Concept Schemes

SKOS concept schemes are sets of related concepts – subject headings, keyword tags, thesaurus entries, or the like – expressed in RDF using Simple Knowledge Organization System (SKOS), a data model for hierarchies of broader, narrower, and related concepts. While SKOS concept schemes may resemble formal class-based Ontologies, they do not consist of “classes” that can be leveraged for advanced inferencing. Rather, concepts are modeled as “instances” that can be leveraged for the ontologically less sophisticated, though eminently useful, purposes of broadening or narrowing the scope of searches. Confusingly for people coming from a library or information management background, SKOS concept schemes are not considered to be RDF vocabularies per se, but RDF Data in a more generic sense.

Uniform Resource Identifiers (URIs)

Uniform Resources Identifiers, typically constructed using the http:// syntax of Web addresses, are used to name any thing of interest (“resource”), whether physical or conceptual, accessible on the Web or not.