Ontology Design Pattern-driven Linked Data Publishing

Adila Krisnadhi
Data Semantics Lab (a.k.a. DaSeLab)
Wright State University, Dayton, OH
E-mail: krisnadhi@gmail.com
GitHub: krisnadhi

2016 ESIP Summer Meeting, Durham, NC
This talk is about ...

Realizing interoperability without sacrificing (semantic) heterogeneity.
• At least mentioned/introduced in ...
  – Botts, Fredericks, Gayanilo, Rueda. “Building Semantic and Syntactic Interoperability Into EnviroSensing Systems” (Tuesday afternoon)
  – Narock. “Ontologies and the Semantic Web - An Introduction for Non-Experts” (Late Wednesday afternoon)
“Often seen, though not all are realized”

W3C Semantic Web Activity (until end of 2013)

W3C Data Activity (2014 onward)
- WG on Data on the Web Best Practices
- WG on RDF Data Shapes
- WG on Spatial Data on the Web (Joint with OGC)
- SIG on Health Care and Life Sciences
Or alternatively ...

Semantic Web

- Linked Data
- Vocabulary, Ontology
- Inferencing, Querying, etc.
LINKED DATA PUBLISHING
• Use graph data model based on RDF.
• RDF graph is a set of RDF triples.
• RDF triple consists of:
  – Subject: URI, anonymous resource
  – Predicate: URI
  – Object: URI, literal, anonymous resource.
• Serialization format: XML, Turtle, Ntriple, JSON-LD.
• A triple can express a linking between pieces of data.
• Simplicity leads to popularity.
• See also Carlos Rueda’s slides on how to triplify tabular/relational data.
State of Linked Data
How do you publish (linked) data

• Linked Data Principles:
  – Use Web identifiers: HTTP URI/IRI
  – Ensure that URIs are Web-resolvable so human AND machine can obtain further information about the things URIs represented.
    • Machine-processable description \(\rightarrow\) RDF graph/triples.
  – As much as possible link to data from other parties.

• In practice, you need to decide how to:
  – Prepare vocabulary to describe/link your data
  – Mint URIs for your data and vocabulary
    • Incl. minting resolvable URIs for the vocabulary terms if necessary.
  – Set up infrastructure to serve the data as Linked Data.
Should I mint URI for X?


• If X is instance data:
  – Do, if X comes from your own local database/source.
  – Don’t (i.e., reuse existing one), if X originates from external source you don’t maintain.

• If X is a vocabulary term:
  – Do, if there’s no known URI for X or you want to assert your own definition for X (because it does not exist, or you dislike the existing one).
    • Unless the current maintainer of definition of X agrees with your (new) definition.
  – Don’t, if you like existing defn and it fits your current AND future needs.

• In any case, if you DO decide to mint a new URI for X, you’re responsible to maintain it. ➔ URIs must be persistent!

• URIs should preferably be opaque ➔ machines should not parse or read into URI to infer anything about the referenced resource; infer from the description of the data in the graph (the RDF triples).
Other things to consider ...

• Hash URI vs. Slash URI
  – Hash URI, e.g.: http://www.w3.org/ns/prov#wasAssociatedWith
  – Slash URI, e.g.: http://data.rvdata.us/id/award/100044
    • May involve a 303 Redirect
  – see https://www.w3.org/TR/cooluris/ and https://www.w3.org/wiki/HashVsSlash
  – I personally like to use hash URI for vocabulary terms, and slash URI for data instances

• Naming convention for URIs
  – CamelCase-ing?
  – Use of ‘-’ (dash) and/or ‘_’ (underscore), etc.
Ensuring Web-resolvability in a Linked Data way

• Every lookup of a URI should return *something*.
• If a human-readable description is requested:
  – Usually indicated by *content-type* header *text/html*
  – Return HTML page.
• If a machine-readable description is requested:
  – Indicated by *content-type* header:
    application/rdf+xml, application/json, text/turtle, etc.
  – Return the appropriate serialization format.
• Easing the URI persistence: use permanent redirection through PURL service (see
  [http://www.purlz.org](http://www.purlz.org), [https://w3id.org/](https://w3id.org/))
VOCABULARY PREPARATION
Vocabulary and Ontology

• Ontology = formalized vocabulary
  – Formally, ontology = set of logical statements (axioms) involving the vocabulary terms.
  – Standardized ontology languages: RDFS, OWL
  – Rule-based language such as RIF and SWRL can also be used, though more rarely.

• Why ontologies are valuable (Janowicz, 2016)?
  – Improve discoverability of your own data (as opposed to simple keyword search)
  – Cornerstone of data publication and managing strategies
  – Improve data reproducibility (through provenance information)
  – Ease cross-repository knowledge exploration (follow-your-nose browsing)
  – Ease the detection of inconsistency in the data.
  – Enable data integration
Misconceptions about Ontology

• Misconception #1: The purpose of ontology is to agree on what the term means.
  – Correction: Its purpose is to make intended meaning explicit.

• Misconception #2: Common upper-level and (large, overarching) domain ontologies could solve the messiness of Linked Data world.
  – Correction: different and conflicting perspectives are natural in the open, so there is no way to force everyone to use the same classes and properties.

• Misconception #3: Ontology constrains the way the vocabulary terms are used.
  – Correction: Ontology employs open-world assumption and inferential semantics,
  – e.g., specifying a (global) domain restriction of a property does not constrain the property usage, instead it adds more inferences.
Where to find ontologies/vocabularies?

- LOV (Linked Open Vocabulary) site - http://lov.okfn.org/
- W3C hosts several prominent ontologies/vocabularies:
  - See http://lov.okfn.org/dataset/lov/agents/W3C
- ESIP repositories:
  - http://cor.esipfed.org/ont#
  - http://semanticportal.esipfed.org/ontologies
- OBO Foundry - http://www.obofoundry.org/
- ODP Public Catalog - http://www.gong.manchester.ac.uk/odp/html/
- NCBO Bioportal - http://bioportal.bioontology.org/
• Choosing appropriate ontologies essentially depends on what you want to do with them.
  – Does ontology X defines the terms you need? Do you like/agree with the term definitions? Is X sufficiently extendible?
  – If your needs can only be satisfied by multiple ontologies, does using them together lead to potential problems?

• “I have been told to reuse other ontologies”
  => Yes, but don’t do it at an early stage! Start first with providing your own definition; then align with existing ontologies later.
  – may lead to confusion (e.g., FOAF, Organization onto, vCard, or Schema.org?) and restrict creativity
  – May lead to endless discussion on terms (not to mention: translations)

Source: Oscar Corcho, 2014
If an ontology needs to be developed ......

- **Principle #1: Small >>> Large.**
  - Smallness usually implies simplicity

- **Principle #2: Modular >>> monolithic.**
  - Easier to use as building blocks.
  - Highly extendible
  - Easily understandable

- **Principle #3: Be aware of multiple perspectives. Strike a balance between fostering interoperability vs. allowing semantic heterogeneity.**
  - e.g., street is a connection between two places, but also a separation that cuts a habitat into pieces.

- **Principle #4: Add human-readable annotations**
  - Improve understandability.
Ontology Design Pattern (ODP)

• Is a good candidate w.r.t earlier principles
• ODP: reusable solution of a recurrent modeling problem
• Content ODPs (aka knowledge patterns): ODP corresponding to a core notion in a particular domain.
  – Cover a wide range of domains or application areas.
  – Be extensible to allow additional details; minimal ontology commitments fostering reuse.
  – Be self-contained to a degree where they can be used on their own.
  – Supports multiple granularities.
  – Provide an axiomatization beyond mere surface semantics.
  – Have various hooks to well-known ontologies / patterns.
Variant of Semantic trajectory pattern (Hu, et al., 2013). Axiomatization is also important part of the pattern, but not displayed here. Consult the OWL encoding at http://w3id.org/daselab/onto/trajectory
• Data providers A, B, and C, each with their own local ontologies, but use semantic trajectory pattern as a core component.
• A: data about (pedestrian) human mobility captured using smartphones, other mobile devices, and social media.
• B: data about cars, buses, taxis, trucks, and so forth.
• C: sparse GPS-based wildlife tracking data from Californian mountain lions.
• Federated query example: detect spots where wildlife crosses highways or enters human settlements.
Cruise at R2R
My not-so-well-designed Cruise pattern

geosparql:WktLiteral

my:asWKT

my:Trajectory

my:Cruise

my:Agent

my:Role

my:providesRole

rdfs:subClassOf

my:Person

my:Organization

my:Vessel

my:traversedBy

my:undertakenBy

my:fundedBy

my:Award
Next steps

• Fill in the logical axiomatization of the pattern.
  – Use ontology editors, e.g., Protégé
• Prepare human-readable HTML documentation.
  – E.g., use LODE, Parrot, etc.
• Make both the pattern and the documentation available online according the pattern URI (may need to set up content negotiation)
• Start populate the pattern with data (virtual or warehousing-style).
PUBLISHING AGAINST THE PATTERNS
Local schemas to pattern mapping

- Mappings can be expressed as rules / SPARQL Construct queries / OWL axioms [live demo running SPARQL queries on the R2R and BCO-DMO SPARQL endpoints]

- **R2R**:
  - gl:Cruise (x) -> my:Cruise(x)
  - gl:isUndertakenBy(x,y) -> my:isUndertakenBy(x,y)
  - r2r:hasAward(x,y) -> my:fundedBy(x,y)
  - etc.

- **BCO-DMO**:
  - odo:Cruise(x) -> my:Cruise(x)
  - odo:ofPlatform(x,y) -> my:isUndertakenBy(x,y)
  - odo:Cruise(x), prov:associatedWith(x,y), odo:Project(y), odo:hasAward(y,z), odo:GrantAward(z) -> my:Cruise(x), my:fundedBy(x,z), my:Award(z)
  - etc.
Interoperability through the pattern

• We can make data available according to the pattern.
  – Possible even without physically persistently housing the data.
  – Mapping rules are needed (expressible in SPARQL).

• R2R and BCO-DMO do not have to annotate their data using vocabulary terms in the pattern directly.

• Federated query can also be posed in any of the two repositories’ endpoints, assuming the corresponding repository can read the mapping.
Conclusion

• ODPs can act as interoperability bridge, or as a glue, without sacrificing the local heterogeneity from each data source.
• There is no need to force everyone to use the same class and properties, as one can map/align local schemas/data models to the ODPs.
  – Helped by the fact that ODPs are small and modular.
• ODPs open a way to publish Linked Data more cheaply since the costly endeavor of developing overarching upper level and domain ontologies can be avoided.
Data Semantics Lab

Pascal Hitzler
Professor, Lab Co-Director

Michelle Cheatham
Asst. Professor, Lab Co-Director

Other Members:
• Postdoc: Adila Krisnadhi
• 8 PhD Students
• A few Master’s students and visiting researchers

Web: http://www.daselab.org
Twitter: @DaSeLab
FB: https://www.facebook.com/daselab
References

1) O. Corcho, “Ontology Engineering for and by the masses: are we already there?”. Keynote Talk at EKAW 2014.


3) K. Janowicz, “Modeling Ontology Design Patterns with Domain Experts – A View From the Trenches”. In: (2)

4) K. Janowicz, A. Gangemi, P. Hitzler, A. Krisnadhi, V. Presutti, “Introduction: Ontology Design Patterns in a Nutshell”. In: (2)

Adila Krisnadhi is supported by the National Science Foundation under the award 1440202 “EarthCube Building Blocks: Collaborative Proposal: GeoLink - Leveraging Semantics and Linked Data for Data Sharing and Discovery in the Geosciences.”
Thank you!

Special thanks to:
Adam Shepherd (BCO-DMO) & Bob Arko (R2R)