Ontology Design Patterns for Large-Scale Data Interchange and Discovery

Pascal Hitzler
DaSe Lab for Data Semantics
Wright State University
http://www.pascal-hitzler.de
Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies

Chapman & Hall/CRC, 2010

Choice Magazine Outstanding Academic Title 2010 (one out of seven in Information & Computer Science)

http://www.semantic-web-book.org
Semantic Web journal

• EiCs: Pascal Hitzler
  Krzysztof Janowicz

• Funded 2010; going strong.

• We very much welcome contributions at the “rim” of traditional Semantic Web research – e.g., work which is strongly inspired by a different field.

• Non-standard (open & transparent) review process.

• http://www.semantic-web-journal.net/
Data Semantics (DaSe) Lab
Wright State University, Dayton, Ohio, USA

Directors: Michelle Cheatham & Pascal Hitzler

PhD students: Reihaneh Amini
David Carral
Amit Joshi
Nazifa Karima
Adila Krisnadhi
Raghava Mutharaju
Stella Sam
Kunal Sengupta
Cong Wang

Master students: Ashley Coleman
Todd Huster
Kylyn Magee
Brooke McCurdy
Current focus topics:

- ontology modeling
- ontology design patterns
- ontology and data alignment
- data and information integration
- use of formal semantics
- semantic web languages
- logical foundations
- efficient reasoning algorithms
- applications in the sciences and elsewhere
Ontologies?
Ontology represents general domain knowledge

e.g. every publication has an author

Reconciling OWL and Rules
Knorr, Hitzler, Maier
ECAI 2012

Data e.g. on Websites
The ontology hype

- Large, well-thought-out ontologies (foundational/domain/etc).

- “You just have to get your formal definitions right, and a lot of the rest will just fall into place.”
The ontology hype

• “You just have to get your formal definitions right, and a lot of the rest will just fall into place.”

  – This does not even work for
    • scientists
    • wanting to share and reuse scientific data
    • through well-kept data repositories

  – So how is this supposed to work for the web at large?
Multiple perspectives

• Try to find a universal definition for
  – Forest
  – Mountain
  – City
  – River
  – Etc.

• The stronger our ontological commitments, the more we loose reusability.

• We need to accept that conceptualizations are often very local, resulting in “micro-ontologies”.
Multiple perspectives

Two ontologies.
Left: transportation domain
Right: agriculture domain

We cannot simply equate a:Canal and b:Canal!
Multiple perspectives

\[
\begin{align*}
\text{symmetric(a:hasSpouse)} \\
\exists a:hasSpouse.a:Female &\sqsubseteq a:Male \\
\exists a:hasSpouse.a:Male &\sqsubseteq a:Female \\
a:hasWife(a:john, a:mary) &\qquad b:Male(a:john) \\
b:Female(a:mary) &\qquad b:Male(b:david) \\
\end{align*}
\]

\[
\text{symmetric(b:hasSpouse)} \\
b:hasSpouse(b:mike, b:david) \\
b:Male(b:david) \\
b:Male(b:mike) \\
b:Female(b:anna) \\
\]

\[a:Male \sqcap a:Female \sqsubseteq \bot\]
The well-done ontologies

- Brittle
- Expensive
- Sometimes unintuitive
- Unwieldy
- Single-perspective
- Difficult to reuse

- Work in some contexts.
- Work if a lot of central control is imposed.
- Need a lot of manpower to create.
Pre-LOD Semantic Web

- Large, monolithic ontologies
- Sophisticated ontology languages

Scientific Hypothesis:
These will solve your data and information management problems

Remember that scientific progress is fundamentally about falsification, not verification 😊
Linked Data?
The linked data counter-hype

• “Ontologies don’t work, let’s just link data”

• “Okay, with a little bit of ontologies on top.”

• “The Linked Data Web is the true Semantic Web.”
Geoindexed Linked Data – courtesy of Krzysztof Janowicz
http://stko.geog.ucsb.edu/location_linked_data
Using Linked Data is tricky

“Nancy Pelosi voted in favor of the Health Care Bill.”
Alignment? Integration?

R2R:

foaf:Person
  type
  person/101396
    name
      "Smith, John"

BCO-DMO:

foaf:Person
  type
  Person_752
    name
    familyName
    givenName
      "John Smith"
      "John"
      "Smith"
Absence of schema?

Copernicus lunar crater located on earth – courtesy of Krzysztof Janowicz http://stko.geog.ucsb.edu/location_linked_data (missing reference coordinate system)

Copernicus is a lunar impact crater named after the astronomer Nicolaus Copernicus, located in eastern Oceanus Procellarum. It is estimated to be about 800 million years old, and typifies craters that formed during the Copernican period in that it has a prominent ray system.

Contents
- Characteristics
- Names
- Satellite
- craters
- See also
- References
- External links

**Characteristics**

Copernicus is visible using **binoculars**, and is located slightly northwest of the center of the Moon's Earth-facing hemisphere. South of the crater is the Mare Insularum, and to the south-south-west is the crater Reinhold. North of Copernicus are the Montes Carpatus, which lie at the south edge of Mare Imbrium. West of Copernicus is a group of dispersed lunar hills. Due to its relative youth, the crater has remained in a relatively pristine shape since it formed.

The circular rim has a discernible hexagonal form, with a **terraced** inner wall and a 30 km wide, sloping **rampart** that descends nearly a kilometer to the surrounding **mare**. There are three distinct terraces visible, and arc-shaped **landslides** due to slumping of the inner wall as the crater debris subsided.

Most likely due to its recent formation, the crater floor has not been flooded.

- Quality of schema and documentation.
- Level of reuseability.
Example from Linked MDB

Film

hasActor

Actor

hasName

xsd:string
Problem!

Sesame Street hasActor Actor (class) hasActor “Jim Henson” hasName

Muppet Show hasActor Actor (class) plays Kermit
plays Ernie plays
The linked data counter-hype

• “Ontologies don’t work, let’s just link data”

• “Okay, with a little bit of ontologies on top.”

• But then we don’t even know how to effectively query over multiple linked datasets (without using a lot of manpower to manually integrate them).

• It seems rather obvious that we need to get ontologies into the picture, but how to do it while avoiding the drawbacks of strong ontological commitments?
So What Now?
Ways forward?

How to establish a flexible conceptual architecture using data and ontological modeling?
“An ontology design pattern is a reusable successful solution to a recurrent modeling problem.”

So-called \textit{content patterns} usually encode specific abstract notions, such as process, event, agent, etc.
Ontology Design Patterns

• Bottom-up homogenization of data representation.

• Avoidance of strong ontological commitments.

• Avoidance of standardization of specific modeling details.

• Well thought-out patterns can be very strong and versatile, thus serve many needs.

We are currently establishing many geo-patterns in a series of hands-on workshops, the GeoVoCamps, see http://vocamp.org/
Ontology Design Patterns

“Horizontal” alignment via patterns
Example: The NSF GeoLink Project
EarthCube: Developing a Community-Driven Data and Knowledge Environment for the Geosciences

“concepts and approaches to create integrated data management infrastructures across the Geosciences.”

“EarthCube aims to create a well-connected and facile environment to share data and knowledge in an open, transparent, and inclusive manner, thus accelerating our ability to understand and predict the Earth system.”
EarthCube GeoLink project

Targeting data sharing and discovery in the Earth Sciences.

LDEO: Robert Arko, Suzanne Carbotte, Kerstin Lehnert
WHOI: Cynthia Chandler, Peter Wiebe, Lisa Raymond, Adam Shepherd
UCSB: Mark Schildhauer, Krzysztof Janowicz, Matt Jones, Yingjie Hu

Ocean Leadership: Douglas Fils
Marymount Univ: Thomas Narock
WSU: Pascal Hitzler, Michelle Cheatham, Adila Krisnadhi
UMBC: Tim Finin

~$1.9M, 2 years duration
GeoLink setup

User Interface: Faceted Search

GeoLink Patterns

mappings

R2R  BCO-DMO  IEDA  LTER  IODP  ...

November 2014 – EKAW 2014 – Linköping – Pascal Hitzler
“An ontology design pattern is a reusable successful solution to a recurrent modeling problem.”

So-called *content patterns* usually encode specific abstract notions, such as process, event, agent, etc.

Patterns provide modular, reusable, replaceable, pieces.

By agreeing on *reuse of generic patterns* (but *leaving the relationships* between the patterns to a specific assembly for a special purpose), we can have *reuse while preserving heterogeneity*.
Semantic Trajectories

[Hu, Janowicz, Carral, Scheider, Kuhn, Berg-Cross, Hitzler, Dean, COSIT2013]
Semantic Trajectories
Semantics in OWL

\[\exists \text{atTime}.OWL\text{-Time}:\text{Temporal Thing} \sqcap \exists \text{hasLocation}.\text{Position}\]
\[\sqcap \exists \text{hasFix}' . \text{SemanticTrajectory}\]

\[\text{Segment} \sqsubseteq \exists \text{startsFrom}.\text{Fix} \sqcap \exists \text{endsAt}.\text{Fix}\]
\[\square \leq 1 \text{startsFrom}.\top\]
\[\square \leq 1 \text{endsAt}.\top\]

\[\text{Segment} \sqsubseteq \exists \text{hasSegment}' . \text{SemanticTrajectory}\]

\[\text{startsFrom}' \circ \text{endsAt} \sqsubseteq \text{hasNext}\]
\[\text{hasNext} \sqsubseteq \text{hasSuccessor}\]
\[\text{hasSuccessor}' \circ \text{hasSuccessor} \sqsubseteq \text{hasSuccessor}\]
\[\text{hasNext}' \sqsubseteq \text{hasPrevious}\]
\[\text{hasSuccessor}' \sqsubseteq \text{hasPredecesor}\]
Semantics in OWL

\[
\begin{align*}
\text{Fix} \subseteq & \neg \exists \text{endsAt}. \text{Segment} \sqsubseteq \text{StartingFix} \\
\text{Fix} \subseteq & \neg \exists \text{startsFrom}. \text{Segment} \sqsubseteq \text{EndingFix} \\
\text{Segment} \sqsubseteq & \exists \text{startsFrom}. \text{StartingFix} \sqsubseteq \text{StartingSegment} \\
\text{Segment} \sqsubseteq & \exists \text{endsAt}. \text{EndingFix} \sqsubseteq \text{EndingSegment} \\
\text{SemanticTrajectory} \sqsubseteq & \exists \text{hasSegment}. \text{Segment} \\
\text{hasSegment} \circ \text{startsFrom} \sqsubseteq & \text{hasFix} \\
\text{hasSegment} \circ \text{endsAt} \sqsubseteq & \text{hasFix} \\
\exists \text{hasSegment}. \text{Segment} \sqsubseteq & \text{SemanticTrajectory} \\
\exists \text{hasSegment}^- . \text{SemanticTrajectory} \sqsubseteq & \text{Segment} \\
\exists \text{hasFix}. \text{Segment} \sqsubseteq & \text{SemanticTrajectory} \\
\exists \text{hasFix}^- . \text{SemanticTrajectory} \sqsubseteq & \text{Fix}
\end{align*}
\]
Oceanographic Cruise

- RepositoryObject
- Event
- Vessel
- Segment
- CruiseInformationObject
- Trajectory

Relations:
- originatesFrom
- rdfs:subClassOf
- isUndertakenBy
- isTraversedBy
- hasSegment
- hasTrajectory
- isDescribedBy
Roles (Cruise as Event)
Cruise Trajectories

Diagram of Cruise Trajectory relationships with entities such as Vessel, Cruise, Trajectory, Position, Segment, Fix, Port, and Attribute. The diagram shows properties like hasTrajectory, hasSegment, hasFix, hasLocation, nextFix, atTime, and atPort.
Cruise Trajectories

\[
\begin{align*}
\text{Cruise}(x) \land \text{hasTrajectory}(x, y) \\
\land \text{hasSegment}(y, z) \land \text{isTraversedBy}(z, v) \\
\rightarrow \text{isUndertakenBy}(x, v)
\end{align*}
\]
Cruise trajectory

\[ \text{Cruise}(x) \land \text{hasTrajectory}(x, y) \]
\[ \land \text{hasSegment}(y, z) \land \text{isTraversedBy}(z, v) \]
\[ \rightarrow \text{isUndertakenBy}(x, v) \]

\[ \text{Cruise} \equiv \exists \text{cruise}. \text{Self} \]
\[ \text{cruise} \circ \text{hasTrajectory} \circ \text{hasSegment} \circ \text{isTraversedBy} \]
\[ \sqsubseteq \text{isUndertakenBy} \]
Cruise Trajectories

\[ \text{Fix}(x) \wedge \text{hasAttribute}(x, \text{portStopArrival}) \]
\[ \wedge \text{atPort}(x, y) \wedge \text{hasSpatialFootprint}(y, z) \]
\[ \wedge \text{hasLocation}(x, w) \rightarrow \text{locatedIn}(w, z) \]
Cruise trajectory

\[ \text{Fix}(x) \land \text{hasAttribute}(x, \text{portStopArrival}) \]
\[ \land \text{atPort}(x, y) \land \text{hasSpatialFootprint}(y, z) \]
\[ \land \text{hasLocation}(x, w) \rightarrow \text{locatedIn}(w, z) \]

\[ \text{Fix} \land \exists \text{hasTrajectory} \cdot \{\text{portStopArrival}\} \equiv \exists \text{fixps}\cdot \text{Self} \]
\[ \text{hasLocation}^\neg \circ \text{fixps} \circ \text{atPort} \circ \text{hasSpatialFootprint} \]
\[ \sqsubseteq \text{locatedIn} \]
Patterns as interchange format

- Aggregated data can be “pulled back” along the same mappings, if desired.
- Since the patterns are very generic, there is no loss of information by using them as interchange format.

GeoLink Patterns

- R2R
- BCO-DMO
- IEDA
- LTER
- IODP
- ...

mappings
Ways forward

- Establish a flexible conceptual architecture using data and ontological modeling.
- A principled use of patterns, including
  - the development of a theory of patterns and
  - the provision of a critical amount of central patterns may provide a primary path forward.
Some central questions

- ODPs as subject of study
- Understanding generic versus specific modeling in patterns.
- Developing pattern languages and tools
- Understanding and formalizing relationships between patterns, and making systematic use of it: ecosystems of patterns
- Evaluating the added value of patterns for ontology-based tasks or applications, e.g. ontology alignment, linked data visualization, information integration, ...
Thanks!

www.oceanlink.org
www.geo-link.org
References


References

References

References


References
