Achieving Resilient Communities through Open Data Communities

Tracy Kijewski-Correa
Keough School of Global Affairs
College of Engineering
University of Notre Dame

Charles Vardeman II
Center for Research Computing
University of Notre Dame
Our Relationship with Data
How did we get here?
What’s stopping us?

SOURCE

TECHNOLOGY          POLICY          CULTURE

TOOLS
CHALLENGE: The formula for discoverable data is clear, the formula to change human and organizational behavior is not.
ARCHITECT

MECHANICAL

STRUCTURAL

CONTRACTOR
How do you “view” a wall?

**APPLICATION**
- Massing
- Costing
- Etc.

**ABSTRACTION**
- “wall” = [Volume]

- Finishes
- Load Projection
- Etc.

- “wall” = [Area]

- Relative Position
- Stiffness
- Etc.

- “wall” = [Line]
CHALLENGE: We need to work with different “views”
CHALLENGE: We all “view” data differently
CHALLENGE: (1) Data is dynamic; (2) Data is distributed
How will we get there?
Cyberinfrastructure Center of Excellence Pilot

https://cicoe-pilot.org/

Ewa Deelman, USC (PI)

Co-PIs:

Anirban Mandal, RENCI

Jarek Nabrzyski, Notre Dame University

Valerio Pascucci and Rob Ricci, University of Utah

Funded by the National Science Foundation
Grant #1842042
The future of W3C/OGC Standards

Spatial Data on the Web Best Practices
W3C Working Group Note 28 September 2017

This version:
https://www.w3.org/TR/2017/NOTE-sdw-bp-20170928/

Latest published version:
https://www.w3.org/TR/sdw-bp/

Latest editor's draft:
https://w3c.github.io/sdw-bp/

Previous version:
https://www.w3.org/TR/2017/NOTE-sdw-bp-20170511/

Editors:
Jeremy Tandy, Met Office
Linda van den Brink, Geonovum
Payam Barnaghi, University of Surrey

OGC API - Features - Part 1: Core

Open Geospatial Consortium
Submission Date: 2019-07-11
Approval Date: 2019-09-09
Publication Date: 2019-10-14

External identifier of this OGC® document: http://www.opengis.net/docs/17/ogcapi-features-1.0
Additional Formats (Informative):
Internal reference number of this OGC® document: 17-006r3
Version: 1.0
Category: OGC® Implementation Standard

Editors: Clemens Portele, Fanagiotis (Peter) A. Vretanos, Charles Heazle

Copyright notice
Copyright © 2019 Open Geospatial Consortium
To obtain additional rights of use, visit http://www.opengeospatial.org/legal/
“...link your data to other data to provide context.”
Standards based on “Knowledge Graphs”

Semantic Sensor Network Ontology

W3C Recommendation 19 October 2017 (Link errors corrected 08 December 2017)

This version:
https://www.w3.org/TR/2017/REC-vocab-ssn-20171019/

Latest published version:
https://www.w3.org/TR/vocab-ssn/

Latest editor’s draft:
https://w3c.github.io/sdw/ssn/

Implementation report:
https://w3c.github.io/sdw/ssn-usage/

Previous version:
https://www.w3.org/TR/2017/PR-vocab-ssn-20170907/

Editors:
Armin Haller, Australian National University
Krzysztof Janowicz, University of California, Santa Barbara
Simon Cox, CSIRO
Danh Le Phuoc, Technical University of Berlin
Kerry Taylor, Australian National University
Maxime Lefrançois, École Nationale Supérieure des Mines

Figure 3 Overview of the SOSA classes and properties (observation perspective)
The Hazardous Situation Ontology Design Pattern

Agnieszka Lawrynowicz and Ilona Lawniczak

Institute of Computing Science, Poznan University of Technology, Poznan, Poland

Abstract. This extended abstract describes an ontology for modeling hazardous situations. We build upon static hazards and hazardous events, and on existing standards of occupational safety. We also present an example of the pattern in the occupational safety and health domain.

Key words: ontology design pattern, ODP, hazards, and health

Problem: Pattern mainly concerned with “Hazardous Event” for first response applications.
Patterns are Extensible

A Modification to the Hazardous Situation ODP to Support Risk Assessment and Mitigation

Michelle Cheatham\textsuperscript{1}, Holly Ferguson\textsuperscript{2}, Charles Vardeman II\textsuperscript{2}, and Cogan Shimizu\textsuperscript{1}

\textsuperscript{1} Wright State University
\{michelle.cheatham,cogan.shimizu\}@wright.edu
\textsuperscript{2} University of Notre Dame
\{hfergus2,cvardena\}@nd.edu

Abstract. The Hazardous Situation ontology design pattern models the consequences of exposure of an object to a hazard. In its current form, the ODP is well suited for representing the consequences of exposure after the fact, which is very useful for applications such as damage assessment and recovery planning. In this work, we present a modification to this pattern that enables it to additionally support proactive questions central to risk assessment and mitigation planning.

Keywords: hazard, ontology design pattern, risk assessment, risk mitigation

https://github.com/Vocamp/Hazard
Community Based Efforts

SWEET Ontologies

Introduction


What is SWEET?

SWEET is a highly modular ontology suite with ~6000 concepts in ~200 separate ontologies covering Earth system science.

SWEET is a mid-level ontology and consists of nine top-level concepts that can be used as a foundation for domain-specific ontologies that extend these top-level SWEET components. SWEET's own domain-specific ontologies, which extend the upper level ontologies, can provide users interested in further developing a particular domain with a solid set of concepts to get started. SWEET ontologies are written in W3C Turtle, the Simple RDF Triple Language and are publicly available under the Apache License v2.0.

EnVo

EnVo is a community ontology for the concise, controlled description of environments.

If you use ENVO, please cite:


and


SimCenter

http://sweetontology.net/stateRoleImpact/Hazard

Hazard

<table>
<thead>
<tr>
<th>property</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a></td>
<td><a href="http://sweetontology.net/propOrdinal/Impact">http://sweetontology.net/propOrdinal/Impact</a> # type</td>
</tr>
<tr>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a></td>
<td><a href="http://www.w3.org/2002/07/owl#NamedIndividual">http://www.w3.org/2002/07/owl#NamedIndividual</a> # type</td>
</tr>
<tr>
<td><a href="http://www.w3.org/2000/01/rdf-schema#label">http://www.w3.org/2000/01/rdf-schema#label</a></td>
<td>&quot;hazard&quot; @en</td>
</tr>
</tbody>
</table>
Swiss Government and Localities

National Level

- Federal Administration
- offices
- geo.admin.ch - the federal geoportal

<table>
<thead>
<tr>
<th>Current</th>
<th>geodata</th>
<th>sards</th>
<th>Thematic geoportals</th>
<th>Services</th>
<th>Research and Teaching</th>
<th>Via.geo.admin.ch</th>
</tr>
</thead>
</table>

Home page → Services → Geoservices → Linked data service

- services
- Geoservices
  - view services
  - search services
  - Inspection Services
  - Linked data service
  - INSPIRE services

Linked data service: semantically link geodata

Local Level

- Stadt Zürich
- Linked Data

Linked data uses the internet (web) to connect data that is related. The data are identified, shared and linked using the Uniform Resource Identifier (URI). The Resource Description Framework (RDF) and underlying standards such as SPARQL are used to encode and link the data.

sample queries

- Five most populous municipalities (graphic)
- Administrative units at the coordinates 7.43, 46.95
- Districts by canton and year
- All versions of the resource by URI
- Corresponding resource in Wikidata and GeoNames (community)
- Five highest stops
- Stops above 3000 meters above sea level. Sea:
- Stops with means of transport = cog railway, ship

https://tinyurl.com/rnabs83

https://ld.stadt-zuerich.ch/sparql/
Zazuko Trifid Interface provides both human and machine (JSON-LD) accessible interfaces using linked data “follow your nose” principles.

Competency Question: “What is the corresponding resource in Wikidata and Geonames”?

https://id.geo.admin.ch/query

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Name</th>
<th>WikidataURI</th>
<th>GeoNamesURI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aarburg</td>
<td><a href="http://wikidata.org/entity/Q64113">http://wikidata.org/entity/Q64113</a></td>
<td><a href="http://www.geonames.org/7285014/">http://www.geonames.org/7285014/</a></td>
</tr>
<tr>
<td>2</td>
<td>Aarau</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
<tr>
<td>3</td>
<td>Aarberg</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
<tr>
<td>4</td>
<td>Aarberg</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
<tr>
<td>5</td>
<td>Aarburg</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
<tr>
<td>6</td>
<td>Aarburg</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
<tr>
<td>7</td>
<td>Aarburg</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
<tr>
<td>8</td>
<td>Aarburg</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
<tr>
<td>9</td>
<td>Aarberg</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
<tr>
<td>10</td>
<td>Aarburg</td>
<td><a href="http://wikidata.org/entity/Q740968">http://wikidata.org/entity/Q740968</a></td>
<td><a href="http://www.geonames.org/7285018/">http://www.geonames.org/7285018/</a></td>
</tr>
</tbody>
</table>

https://ld.geo.admin.ch/sparql/
Barrier: Ability for Developers to Build “Linked-Data” Applications

Linked Data Journey through Switzerland

A journey through Switzerland's linked data landscape from the federal level to municipalities to crowd-sourced data from Wikidata. All linked together of course.
Purpose

Location Index (Loc-I) is a framework that provides a consistent way to seamlessly integrate data on people, business, and the environment. Location Index aims to extend the characteristics of the foundation spatial data of taking geospatial data (multiple geographies) which is essential to support public safety and wellbeing, or critical for a national or government decision making that contributes significantly to economic, social and environmental sustainability and linking it with observational data. Through providing the infrastructure to support cross-domain foundation data linkages and analysis will open up substantial opportunity for providing a richer set of information to develop, analyse and evaluate policy, programs and service delivery by government.

The following video covers the purpose of the project

Demonstration Project: Natural disaster forecasting capability. Linking the National Exposure Information System with the Multi-Agency Data Integration Project (MADIP) and Business Longitudinal Analysis Data Environment (BLADE)
Loc-I Hybrid Approach

https://tinyurl.com/sbq3tjq
GEOSPATIAL DATA IN THE WEB

To publish data in the Web, we first have to identify the items of interest — those concepts, topics, events, communities, organizations, people, phenomena, features, datasets, applications, and services that are relevant to us. All items of interest are called resources. These are the things whose properties and relationships we want to describe and manage in the Web of Data.

GeoPlatform.gov aggregates, indexes, and links resources of the National Spatial Data Infrastructure (NSDI) to facilitate public and cross-government sharing, discovery, and access, of managed collections of authoritative and other relevant geospatial data. These are the things we are interested in curating, finding, and exploiting. The GeoPlatform Portfolio is a collection of resources of various types, roles, and interrelationships that have been curated for rapid access and use online.

The GeoPlatform Portfolio is managed as Linked Data on the Web. Linked Data enables people and software to share structured data and other related resources on the Web as easily as they can share documents today. The term Linked Data was coined by Tim Berners-Lee in his Linked Data Web Architecture note. The term refers to a style of publishing and interlinking structured data on the Web. The basic assumption behind Linked Data is this: the value and usefulness of data increases the more it is interlinked with other data. In summary, Linked Data is simply about using the Web to create typed links between data from different sources so that relevant data can be more easily found, accessed, and exploited. This is what the GeoPlatform does.

Follow the references below for more information about the Semantic Web, Linked Open Data, and Spatial Data in the Web.

Building the Web of Data (and things)
- Tim Berners-Lee TED Talk (Video) on Linked Open Data
- A short non-technical introduction to Linked Data (another YouTube Video)

Linked Open Data (LOD) tenets and standards
- W3C Data Activities
- W3C Linked Data Standards
- Tim Berners-Lee Linked Data Design Issues
- W3C Data on the Web Best Practices

Spatial Data in the Web
- W3C Spatial Data on the Web: Use Cases and Requirements
- W3C Spatial Data on the Web Best Practices
“AI” to the rescue?

“Rich Context”

Rich Context:
support for cross-agency data stewardship,
measuring dataset impact on public policy

Paco Nathan @pacoid derwen.ai

https://tinyurl.com/vsq6q9g

https://youtu.be/UsmcK64H-sQ?t=3460
How is Resilience Data Used by Communities?

Knowledge Graph, Linked Data Approach Combined with Neural Network Based Approach

Collaboration with NOAA

- Initial focus on coastal inundation and community resilience, working with NOS
- Develop reusable dataset discovery services, so that the public and researchers can find trustworthy, high-impact data
- Identify experts who have used the data and the associated research topics, associated analytical methods and tools, and related datasets (e.g., Zillow, EPA, NASA, FEMA, etc.)
- Generalize for other federal agencies, such as USDA and NSF, as well as to international organizations, such as Deutsche Bundesbank
- Bring in AI expertise from industry and academia: KAIST, LARC, Recognaí, DLA, Primer AI, GESIS, AllenAI, etc.

Administrative Data Research Facility

*Coleridge Initiative*

*Julia Lane*, et al. NYU Wagner

- FedRAMP-compliant [ADRF framework](https://tinyurl.com/vsq6q9g) on AWS GovCloud: “public agency capacity to accelerate the effective use of new datasets”
- for research projects using cross-agency sensitive data, in US and EU – *now in use by 30+ agencies*
- cited as the first federal example of Secure Access to Confidential Data in the final report of the Commission on Evidence-Based Policymaking
- augments Data Stewardship practices; collaboration with Project Jupyter on the related data gov features

https://tinyurl.com/vsq6q9g
Integration of Knowledge Graphs into Computational Toolchains

Funded additions to Project Jupyter

Make datasets and projects top-level constructs, support metadata exchange and privacy-preserving telemetry from notebook usage:

- JupyterLab **Commenting** and real-time collab similar to Google Docs
- JupyterLab **Data Explorer**: register datasets within research projects
- JupyterLab **Metadata Explorer**: browse metadata descriptions, get recommendations through knowledge graph inference (via extension)
- **Data Registry** (original proposal)
- **Telemetry** (privacy-preserving, reports usage)

https://twitter.com/SShanabrook/status/1182442214980501505

Emerging category: watch the “AI Natives”

Projects (mostly OSS) that leverage knowledge graph of metadata about datasets and their usage:

- **Amundsen** @ Lyft
data discovery and metadata
- **Data Hub** @ LinkedIn
data discovery and lineage
- **Marquez** @ Stitch Fix
collect, aggregate, visualize metadata
- **UMS** @ Uber
manage metadata about datasets
- **Metcat** @ Netflix
data discovery, metadata service
- **Dataportal** @ Airbnb
integrated data-space (not OSS)

https://tinyurl.com/vsq6q9g
Can we get there?
DATA MAP: Critical data you consume and/or produce TODAY and hope for TOMORROW

SCORECARD: Ability to ask critical questions of your data
MAIN ROOM
Robert Bailey
Doug Bausch
Phil Beilin
Youngjun Choe
Louise Comfort
Rachel Davidson
Paolo Gardoni
Peter Herrick
Laurie Johnson
Danielle Mieler
Scott Miles
Simone Nageon de Lestang
Jean-Paul Pinelli
Omar Sediek
Jane Smith
Paul Waddell
Yang Zhang

A ROOM
Jack Baker
Matt Bussmann
Greg Deierlein
Thomas Gernay
Roger Grenier
Keith Henderson
Andrew Kennedy
Michelle Meyer
Skylar Mills
Nicole Paul
Gonzalo Pita
David Prevatt
Anne Rosinski
Erutgrul Taciroglu
Anne Wein
Yu Xiao
Missing? A-M Last Name

B ROOM
Michele Barbato
Rodrigo Costa
Wael Elhaddad
Marta Gonzalez
Sanjay Govindjee
Sara Hamideh
Ajay Harish
Marccus Hendricks
Kishor Jaiswal
Preetish Kakoty
Alex Koeberle
David McCallen
Frank McKenna
Eduardo Miranda
Pallab Mozumder
Charles Wang
Adam Zsarnóczay
Missing? N-Z Last Name
ESIP Operational Readiness Levels

**ORL 1**
- Trusted and vetted source
- Secure data transfer
- Optimized interoperability standards
- Data sharable and consumable with service availability guarantees
- Change notifications issued
- Verified and tested
- Metadata completeness

**ORL 2**
- Trusted and vetted source
- Secure data transfer
- Optimized interoperability standards
- Data sharable and consumable with service availability guarantees
- Some metadata missing

**ORL 3**
- Trusted and vetted source
- Not secure data transfer
- Mostly interoperable
- Some “down time”
- Limited metadata

**ORL 4**
- Trusted and vetted source
- Not secure data transfer
- Mostly interoperable
- Some “down time”
- In testing / development phase
- No metadata

[https://www.esipfed.org/orl](https://www.esipfed.org/orl)
The properties rdf:type and rdfs:label are used to express prov:type and prov:label, respectively.

![Diagram](https://www.w3.org/TR/prov-o/)

**Figure 1.** The three Starting Point classes and the properties that relate them. The diagrams in this document depict Entities as yellow ovals, Activities as blue rectangles, and Agents as orange pentagons. The responsibility properties are shown in pink.

[https://www.w3.org/TR/prov-o/](https://www.w3.org/TR/prov-o/)
The Linked Data version of the USGS GNIS.

A description of the source dataset from the Board on Geographic Names:

The Geographic Names Information System (GNIS) is the Federal and national standard for geographic nomenclature. The U.S. Geological Survey developed the GNIS in support of the U.S. Board on Geographic Names as the official repository of domestic geographic names data, the official vehicle for geographic names use by all departments of the Federal Government, and the source for applying geographic names to Federal electronic and printed products.

The GNIS contains information about physical and cultural geographic features of all types in the United States, associated areas, and Antarctica, current and historical, but not including roads and highways. The database holds the Federally recognized name of each feature and defines the feature location by state, county, USGS topographic map, and geographic coordinates. Other attributes include names or spellings other than the official name, feature designations, feature classification, historical and descriptive information, and for some categories the geometric boundaries.