SimCenter NHERI

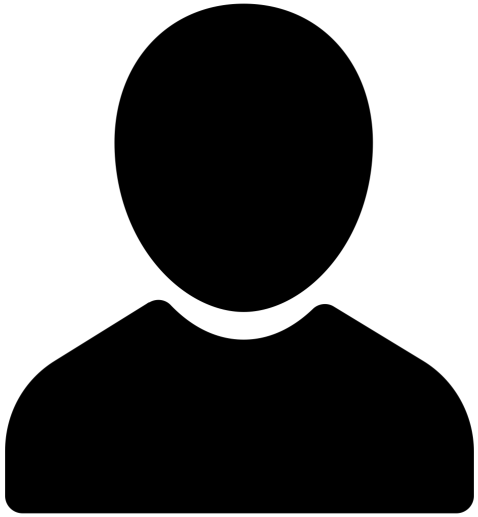
Center for Computational Modeling and Simulation

Achieving Resilient Communities through Open Data Communities

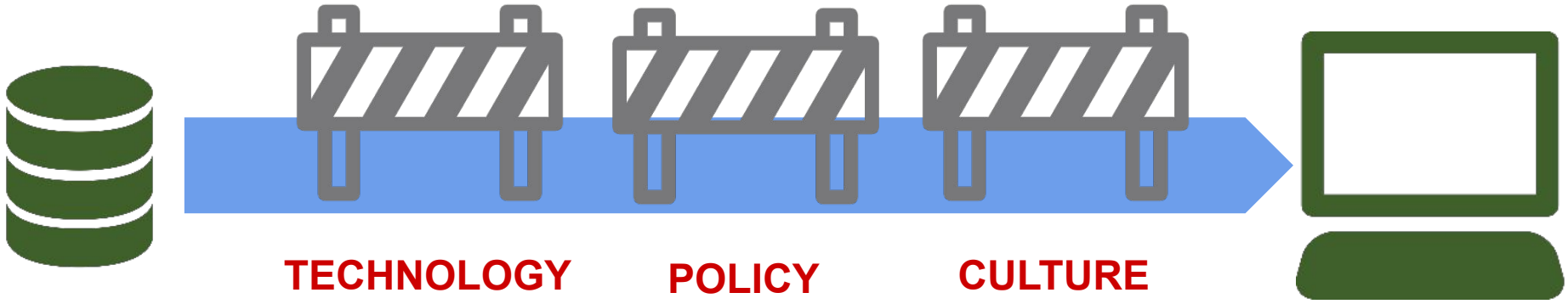
Tracy Kijewski-Correa	Charles Vardeman II
Keough School of Global Affairs College of Engineering University of Notre Dame	Center for Research Computing University of Notre Dame

Our Relationship with Data

How did we get here?



What's stopping us?



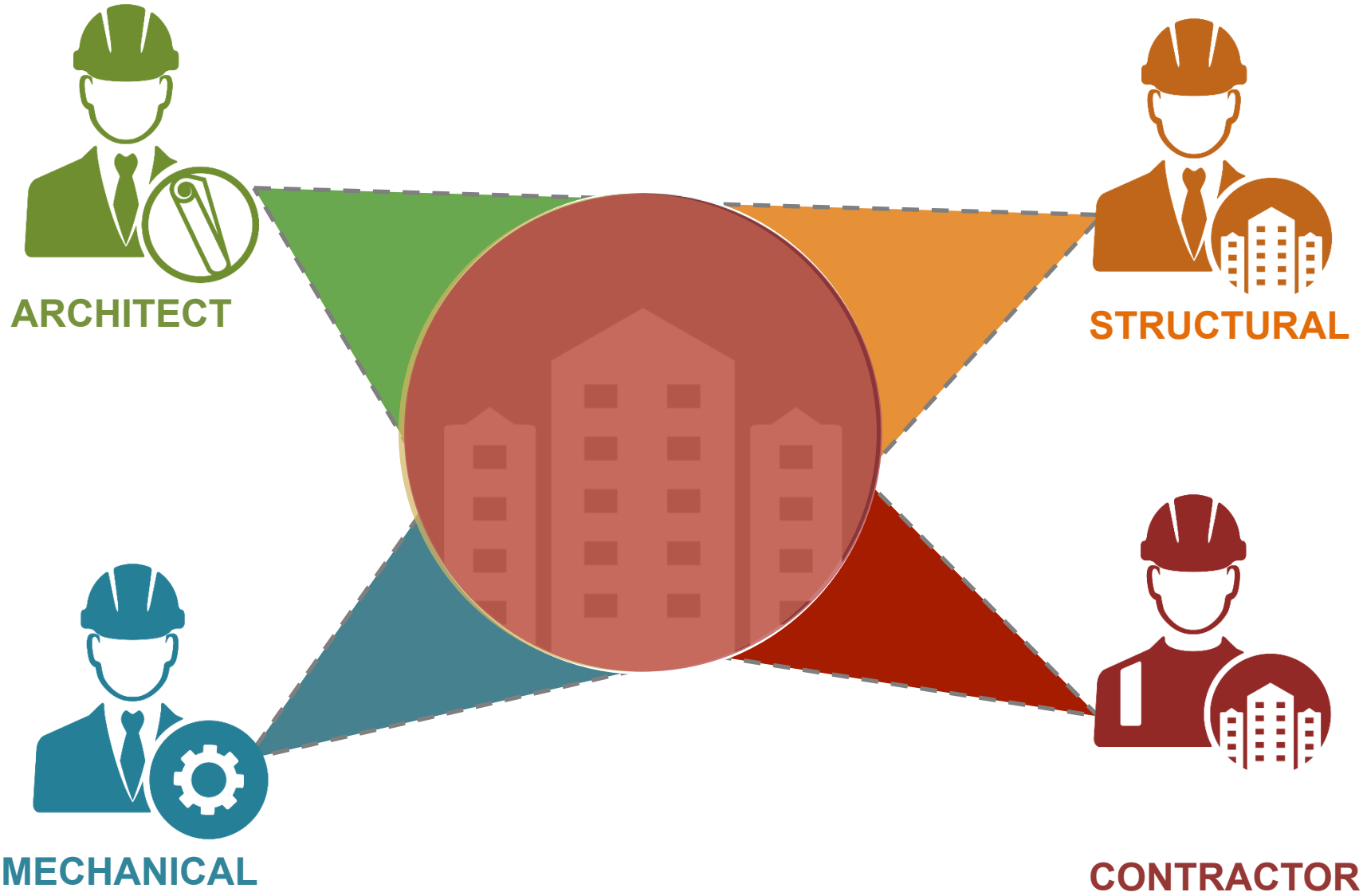
SOURCE

TOOLS

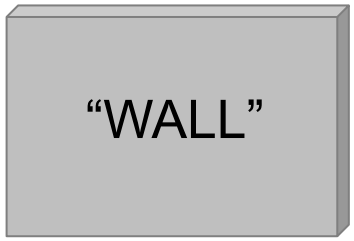
OPEN DATA FORMULA

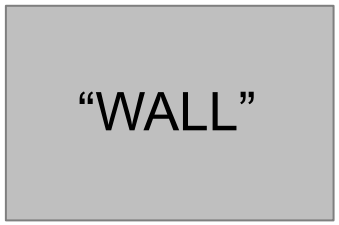
SOURCE + SCHEMA + STANDARDS + INTERFACES + TOOLS

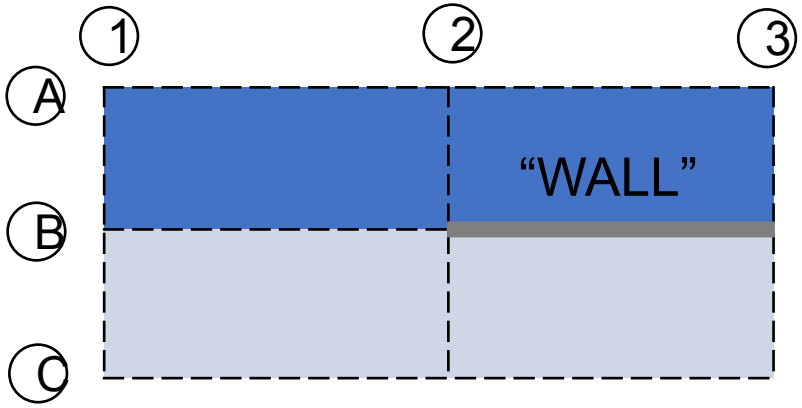
CHALLENGE: The formula for **discoverable data** is clear, the formula to change **human and organizational behavior** is not.


ARCHITECT
STRUCTURAL
MECHANICAL
CONTRACTOR

How do you “view” a wall?

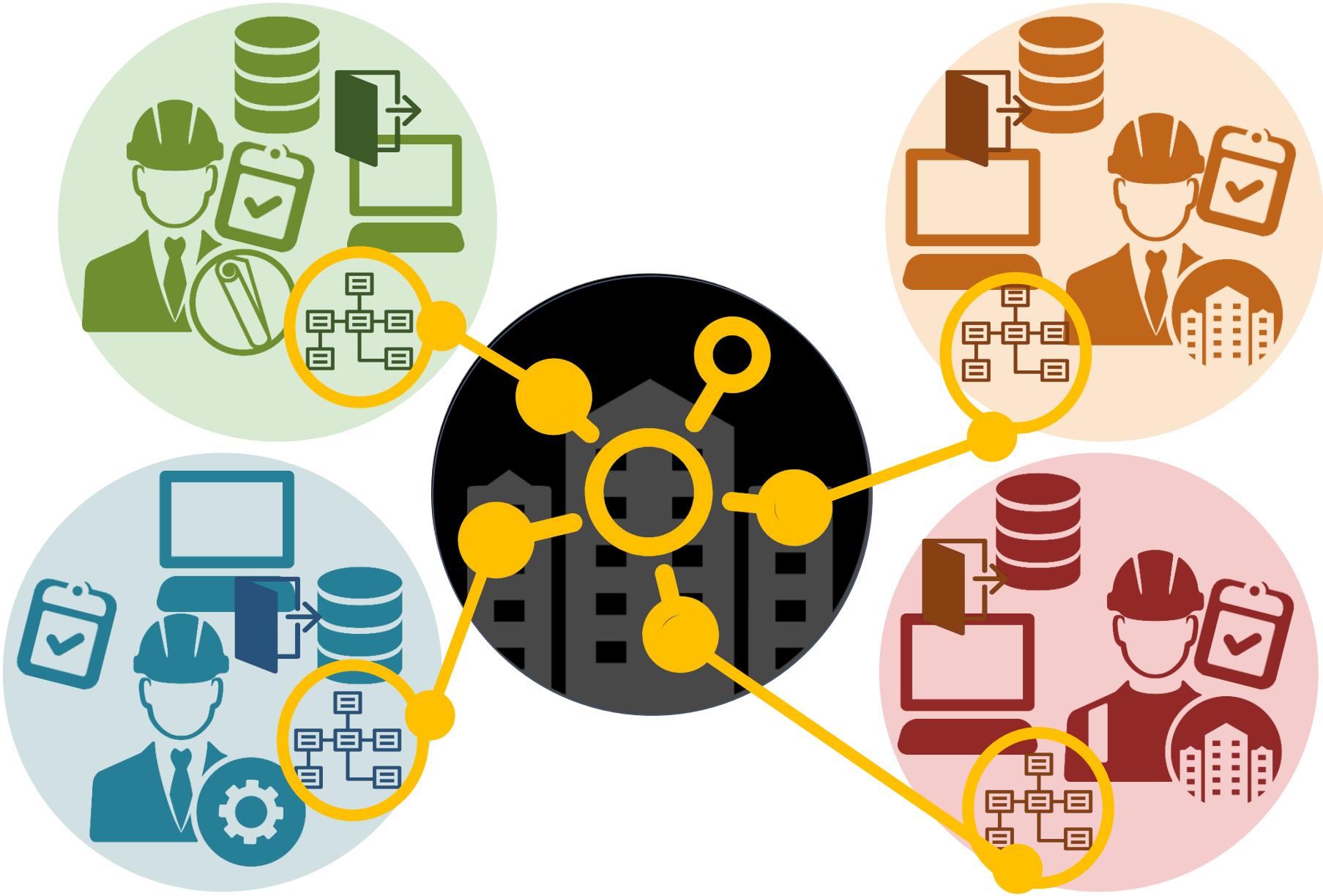




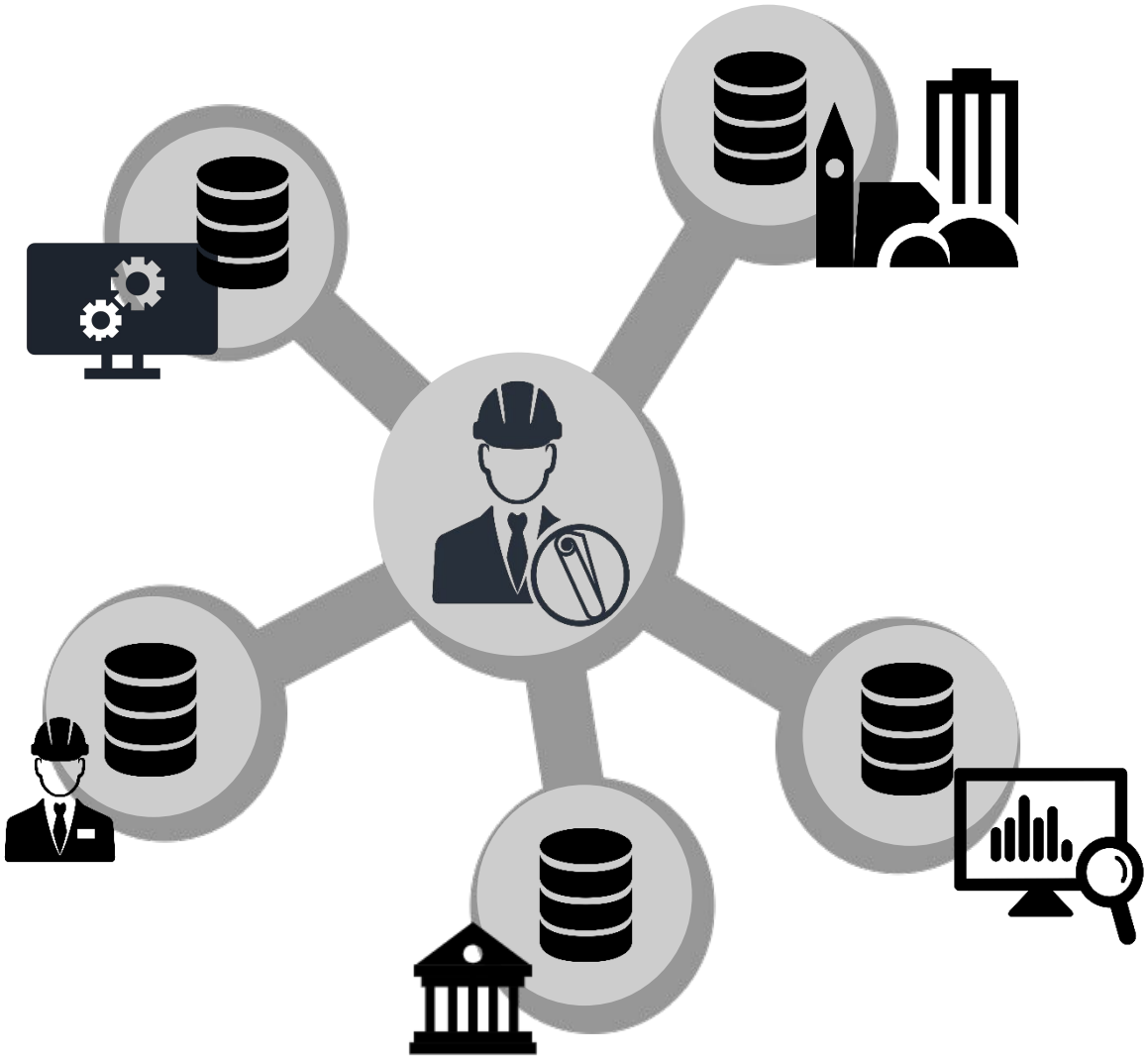


APPLICATION	ABSTRACTION
Massing Costing Etc.	“wall”= [Volume]
Finishes Load Projection Etc.	“wall”= [Area]
Relative Position Stiffness Etc.	“wall”= [Line]

CHALLENGE: We need to work with different “views”

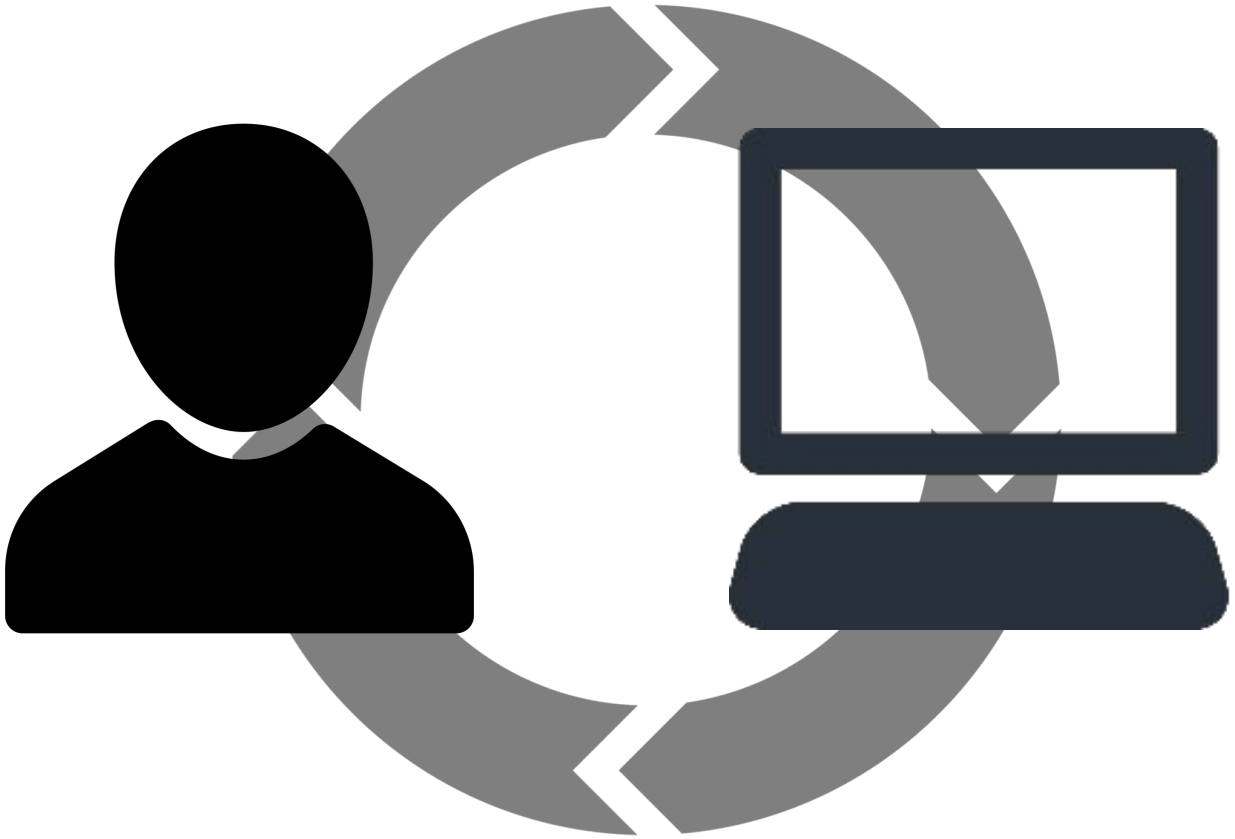


CHALLENGE: We all “view” data differently



CHALLENGE: (1) Data is **dynamic**; (2) Data is **distributed**

How will we get there?



CI CoE PILOT

Cyberinfrastructure Center of Excellence Pilot

https://cicoe-pilot.org/

Funded by the National Science Foundation Grant #1842042

Ewa Deelman, USC (PI)

Co-PIs:

Anirban Mandal, RENCI

Jarek Nabrzyski, Notre Dame University

Valerio Pascucci and **Rob Ricci**, University of Utah

USC Viterbi School of Engineering Information Sciences Institute

UNIVERSITY OF NOTRE DAME

THE UNIVERSITY OF UTAH

CENTER FOR APPLIED CYBERSECURITY RESEARCH

renci

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® Making location count.

W3C

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/doc/IS/ogcapi-features-1/1.0

Additional Formats (informative):

Internal reference number of this OGC® document: 17-069r3

Version: 1.0

Category: OGC® Implementation Standard

Editors: Clemens Portele, Panagiotis (Peter) A. Vretanos, Charles Heazel

OGC API - Features - Part 1: Core

Copyright notice

W3C/OGC Best Practices

"...link your data to other data to provide ***context***."

Best Practice 1: Use globally unique persistent HTTP URIs for Spatial Things

Use stable HTTP URIs to identify Spatial Things, re-using commonly used URIs where they exist and it is appropriate to do so.

Why

To publish spatial data on the Web, we need to stitch the Spatial Things and their corresponding entities into the Web's information space; contributing to the *Web of data*. First: [WEBARCH] Good Practice: Identify with URIs states that "agents should provide URIs as identifiers for resources". Second: the 5 Star Data scheme states: "★★★★ use URIs to denote things, so that people can point at your stuff".

Resources identified with HTTP URIs can be specified as the target of links within the Web's global information space, enabling information to be related, combined and referred to. This is the fundamental basis of 5★ Linked Data: "★★★★★ link your data to other data to provide context".

The HTTP URIs used to identify Spatial Things need to be stable or persistent so that relationships that link them to other resources don't break.

Intended Outcome

Spatial Things become part of the Web's global information space enabling them be linked with other Spatial Things and other resources and for those links to be durable. In other words, spatial data becomes part of the Web of Data.

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 Barbar
Simon Cox, CSIRO
Danh Le Phuoc, Technical University of Berlin
Kerry Taylor, Australian National University
Maxime Lefrançois, École Nationale Supérieure des Mine

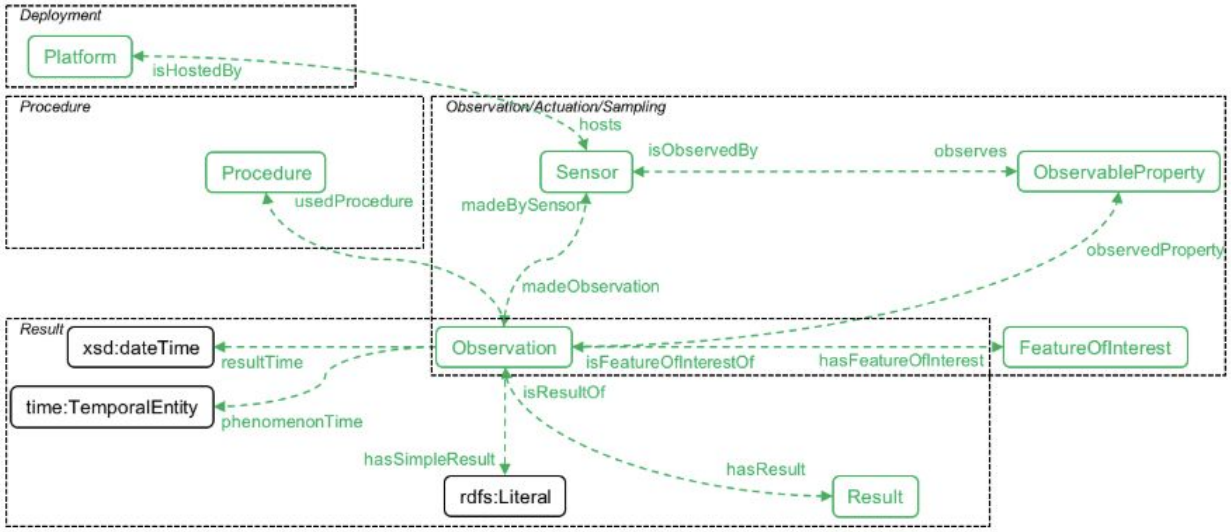


Figure 3 Overview of the SOSA classes and properties (observation perspective)

Reusable Knowledge Graph Fragments

The Hazardous Situation Ontology Design Pattern

Agnieszka Ławrynowicz[1] and Ilona Ławniczak[1]

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

Abstract. This extended abstract describes an ontol
for modeling hazardous situations. We build upon stat
hazards and hazardous events, and on existing standa
of occupational safety. We also present an example of
the pattern in the occupational safety and health don

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

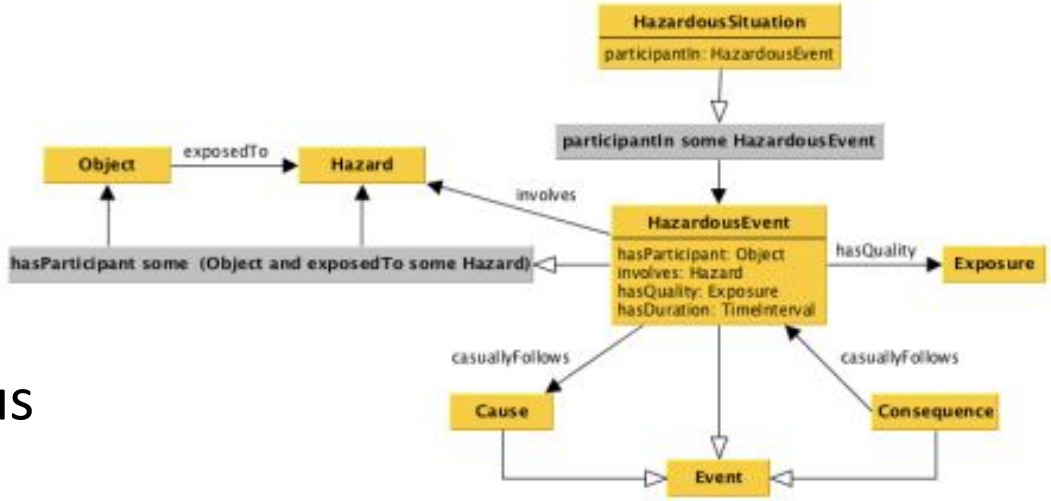


Fig. 1. The Hazardous Situation Ontology Design Pattern

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

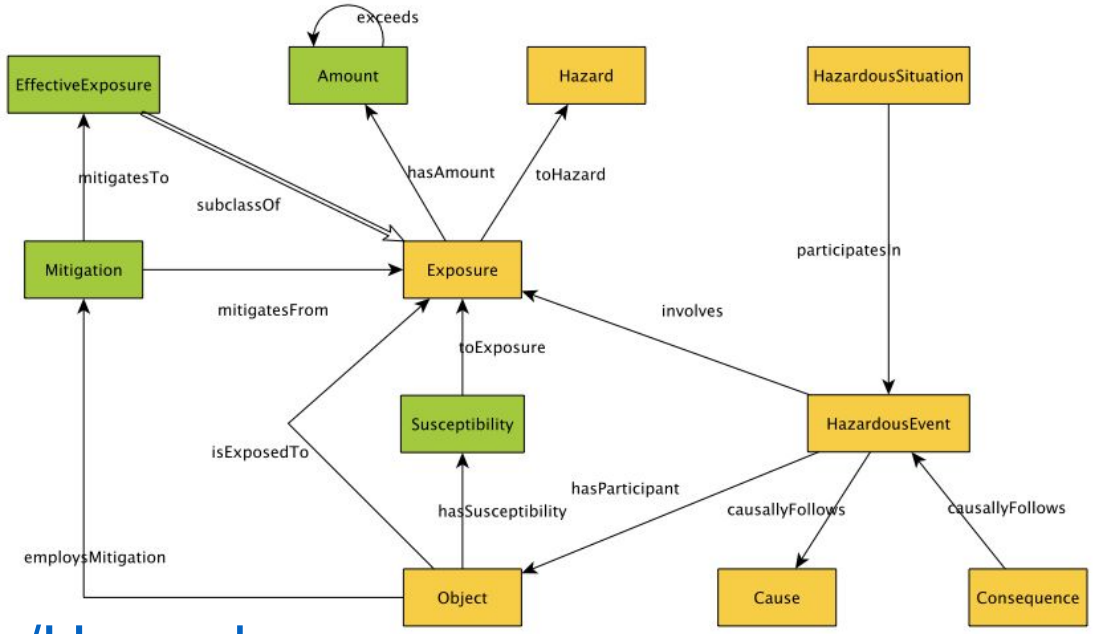
Michelle Cheatham[1], Holly Ferguson[2], Charles Vardeman II[2], and Cogan Shimizu[1]

[1] Wright State University
{michelle.cheatham,cogan.shimizu}@wright.edu
[2] University of Notre Dame
{hfergus2,cvardema}@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.

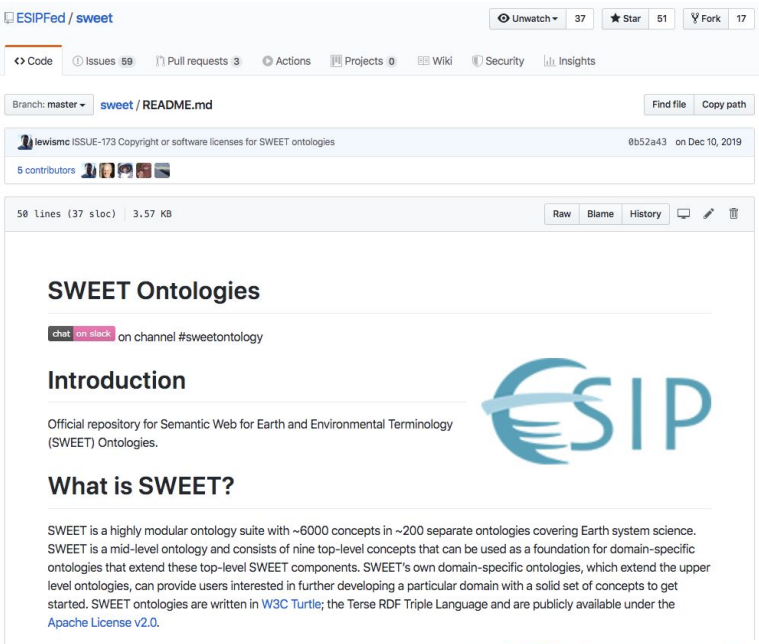
Keywords: hazard, ontology design pattern, risk assessment, risk mitigation

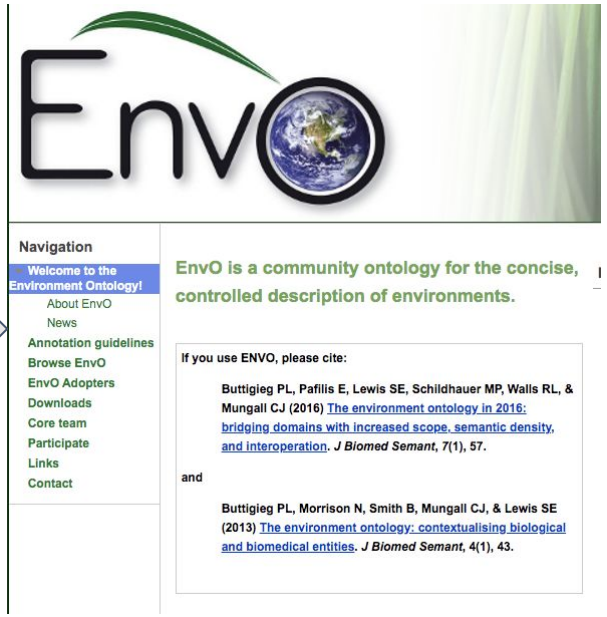
Modeling with Domain Experts using the "Vocamp" Methodology

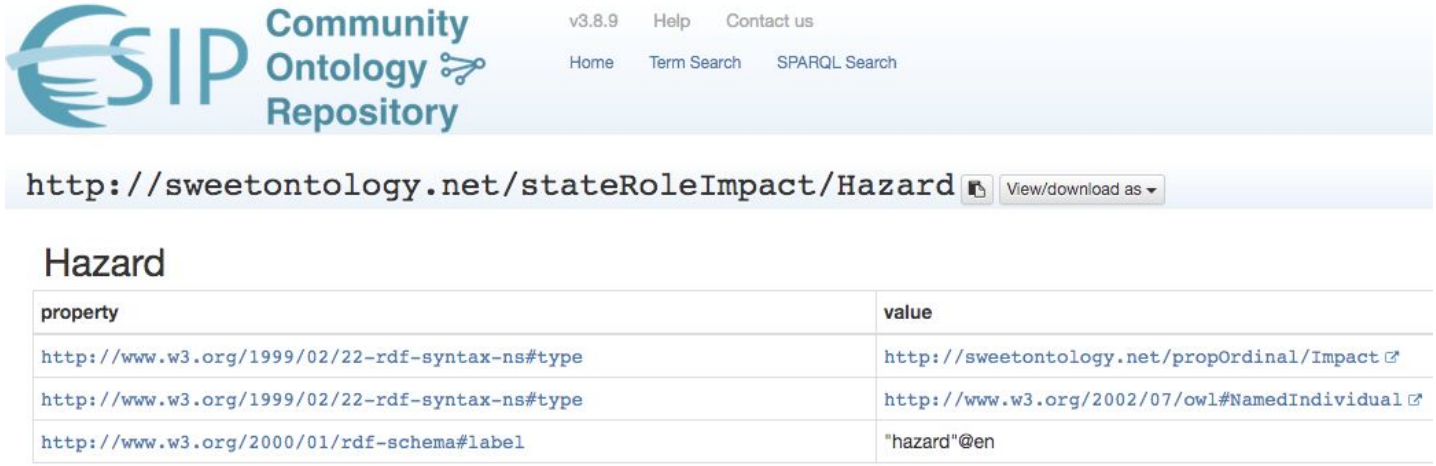


https://github.com/Vocamp/Hazard

Community Based Efforts

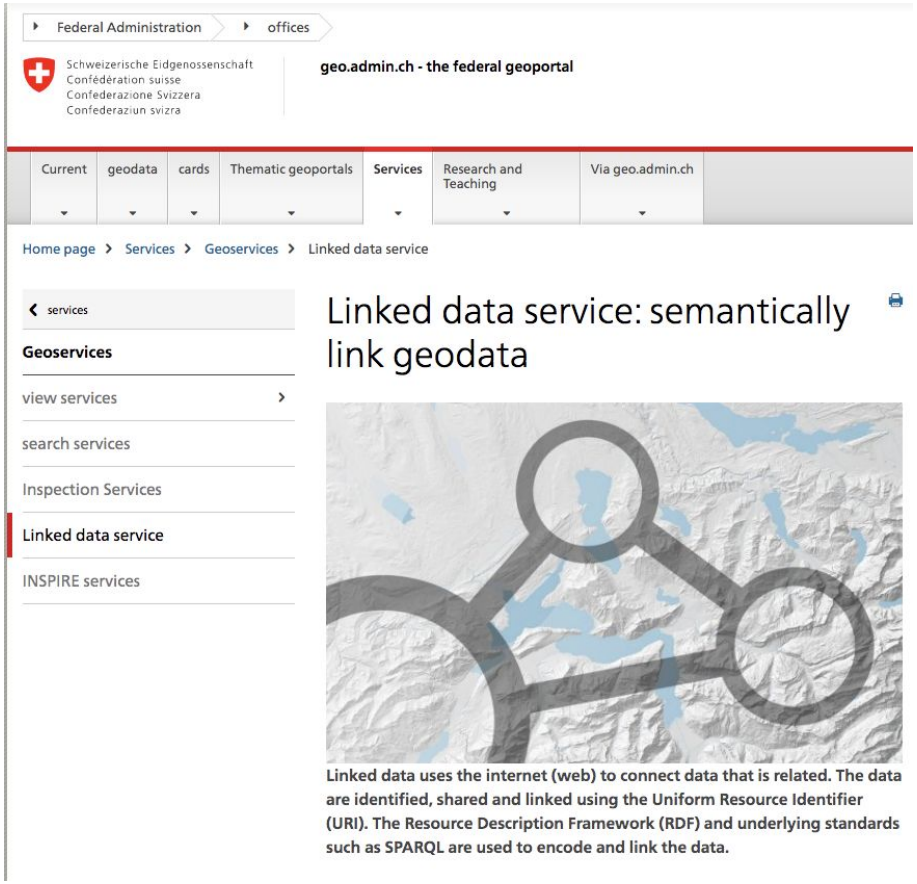






Swiss Government and Localities

National Level



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

Local Level

Stadt Zürich Linked Data

https://ld.stadt-zuerich.ch/sparql/

Open-Source Linked Data Interfaces

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://ld.geo.admin.ch/query

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wikibase: <http://wikiba.se/ontology#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX fn: <http://www.w3.org/2005/xpath-functions#>
PREFIX gn: <http://www.geonames.org/ontology#>
PREFIX schema: <http://schema.org/>

SELECT ?Municipality ?Name ?WikidataURI ?GeoNamesURI WHERE {
    ?Municipality a <http://schema.org/AdministrativeArea> . #specify only the non-versioned entries.
    ?Municipality <http://schema.org/name> ?Name .
```

Table | Response | Pivot Table | Google Chart | Geo | </>

Showing 1 to 50 of 2,278 entries (in 3.162 seconds) Search: Show 50 entries

	Municipality	Name	WikidataURI	GeoNamesURI
1	https://ld.geo.admin.ch/boundaries/municipality/4551	Aadorf	http://www.wikidata.org/entity/Q69131	http://sws.geonames.org/7285012/
2	https://ld.geo.admin.ch/boundaries/municipality/4001	Aarau	http://www.wikidata.org/entity/Q14274	http://sws.geonames.org/7285013/
3	https://ld.geo.admin.ch/boundaries/municipality/301	Aarberg	http://www.wikidata.org/entity/Q64113	http://sws.geonames.org/7285014/
4	https://ld.geo.admin.ch/boundaries/municipality/4271	Aarburg	http://www.wikidata.org/entity/Q64165	http://sws.geonames.org/7285015/
5	https://ld.geo.admin.ch/boundaries/municipality/321	Aarwangen	http://www.wikidata.org/entity/Q69222	http://sws.geonames.org/2661874/
6	https://ld.geo.admin.ch/boundaries/municipality/4221	Abtwil	http://www.wikidata.org/entity/Q69219	http://sws.geonames.org/7285017/
7	https://ld.geo.admin.ch/boundaries/municipality/5621	Aclens	http://www.wikidata.org/entity/Q30986	http://sws.geonames.org/7285018/
8	https://ld.geo.admin.ch/boundaries/municipality/5048	Acquarossa	http://www.wikidata.org/entity/Q68358	http://sws.geonames.org/7285019/
9	https://ld.geo.admin.ch/boundaries/municipality/561	Adelboden	http://www.wikidata.org/entity/Q68994	http://sws.geonames.org/7285020/
10	https://ld.geo.admin.ch/boundaries/municipality/1051	Adliqenswil	http://www.wikidata.org/entity/Q14563	http://sws.geonames.org/7285021/

Bern

https://ld.geo.admin.ch/boundaries/municipality/351:2016

a http://www.opengis.net/ont/geosparql#Feature

type	Feature
area	5162.0 (HA)
isVersionOf	351
issued	2016-01-01 (date)
name	Bern
validUntil	2016-12-31
featureCode	A.ADM3
parentADM1	2:2016
parentADM2	246:2016
parentCountry	CH:2016
population	130015
defaultGeometry	351:2016
hasGeometry	351:2016
bfsNumber	351


```
[ {
  "@graph" : [ {
    "@id" : "http://classifications.data.admin.ch/municipality/6082",
    "@type" : [ "https://gont.ch/PoliticalMunicipality", "https://gont.ch/Municipality" ],
    "http://purl.org/dc/terms/identifier" : [ {
      "@type" : "http://www.w3.org/2001/XMLSchema#integer",
      "@value" : "6082"
    } ],
    "http://www.w3.org/2000/01/rdf-schema#seeAlso" : [ {
      "@id" : "https://ld.geo.admin.ch/boundaries/municipality/6082"
    }, {
      "@id" : "http://dbpedia.org/resource/Ayent"
    } ],
    "http://www.w3.org/2002/07/owl#sameAs" : [ {
      "@id" : "http://www.wikidata.org/entity/Q70167"
    } ],
    "https://gont.ch/id" : [ {
      "@type" : "http://www.w3.org/2001/XMLSchema#integer",
      "@value" : "6082"
    } ],
    "https://gont.ch/municipalityVersion" : [ {
      "@id" : "http://classifications.data.admin.ch/municipalityversion/11666"
    } ]
  } ],
  "@id" : "https://linked.opendata.swiss/graph/eCH-0071"
} ]
```


https://ld.geo.admin.ch/sparql/#

Barrier: Ability for Developers to Build "Linked-Data" Applications

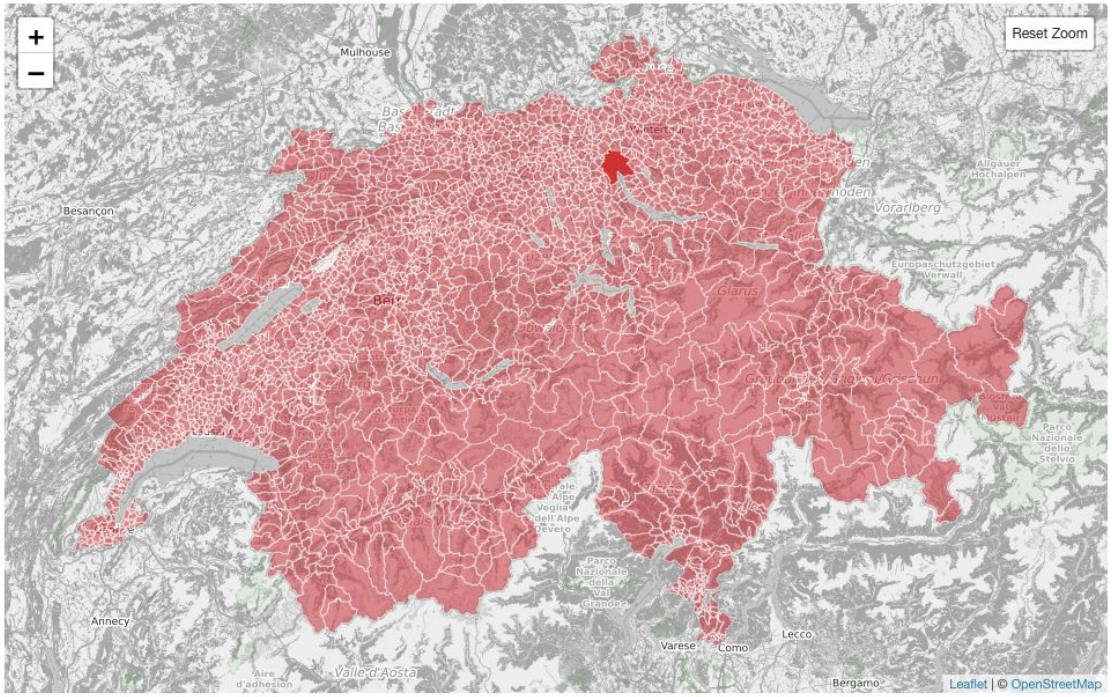
https://observablehq.com/@mmznrstat/dinacon2019

Matthias Mazenauer • www.statistik.zh.ch

Published Oct 16, 2018

Linked Data Journey through Switzerland

A journey through switzerlands linked data landscape from the federal level to municipalities to crowd sourced data from wikidata. All linked together of course.



Ruben Verborgh

blog publications articles teaching contact

Designing a Linked Data developer experience

Making decentralized Web app development fun.

While the Semantic Web community was fighting its own internal battles, we failed to gain traction with the people who build apps that are actually used: *front-end developers*. Ironically, Semantic Web enthusiasts have failed to focus on the *Web*; whereas our technologies are delivering results in specialized back-end systems, the promised intelligent end-user apps are not being created.
Within the Solid ecosystem for decentralized Web applications, Linked Data and Semantic Web technologies play a crucial role. Working intensely on Solid the past year, I realized that designing a fun *developer experience* will be crucial to its success. Through dialogue with front-end developers, I created a couple of JavaScript libraries for easy interaction with complex Linked Data—without having to know RDF. This post introduces the core React components for Solid along with the LDflex query language, and lessons learned from their design.

28 December 2018

https://tinyurl.com/yb86cvn6

Australian Gov't Loc-I

http://locationindex.org/

locationindex

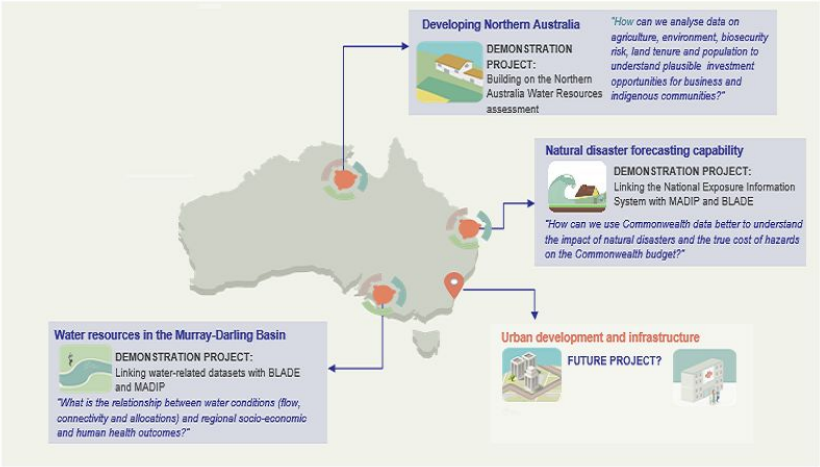
Home | Benefit | Project Partners | Project Progress | Datasets | Linksets | Definitional Items | Tools | DGGS | Linked Project | Contact | Application

Purpose

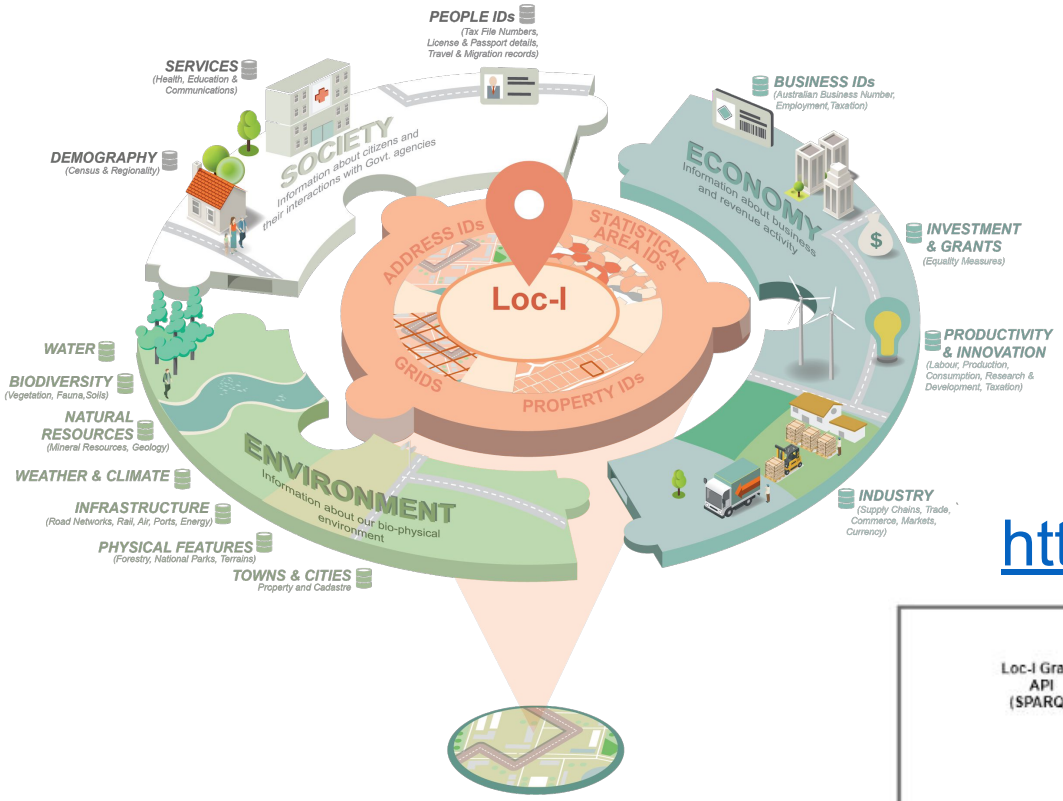
Location Index (Loc-I) is a framework that provides a consistent way to seamlessly integrate data on people, business, and the environment. Location Index aims to extend the characteristics of the foundation spatial data of taking geospatial data (multiple geographies) which is essential to support public 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

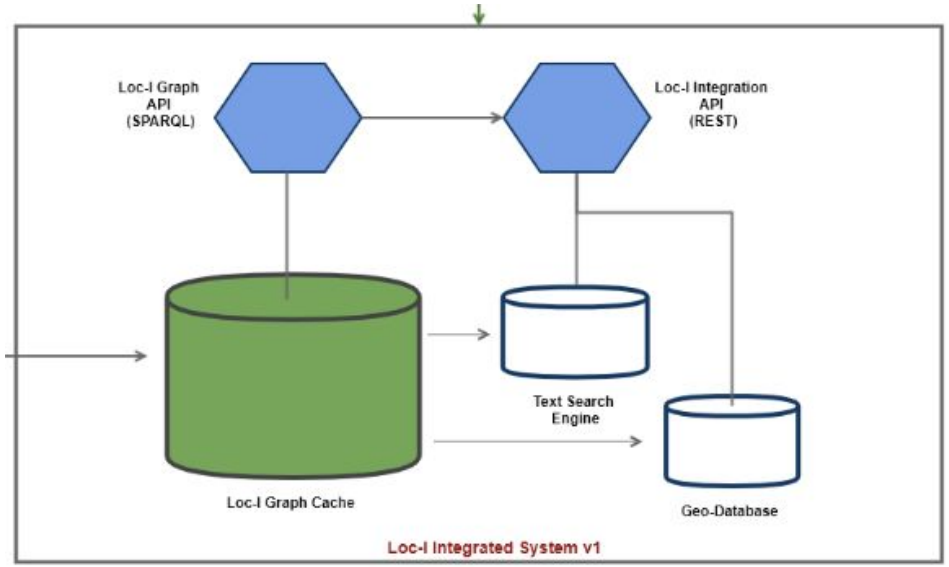
Demonstration Project: Natural disaster forecasting capability. Linking the National Exposure Information System with the Multi-Agency Data Integration Project (MADIP) and Business Longitudinal Analysis Data Environment (BLADE)



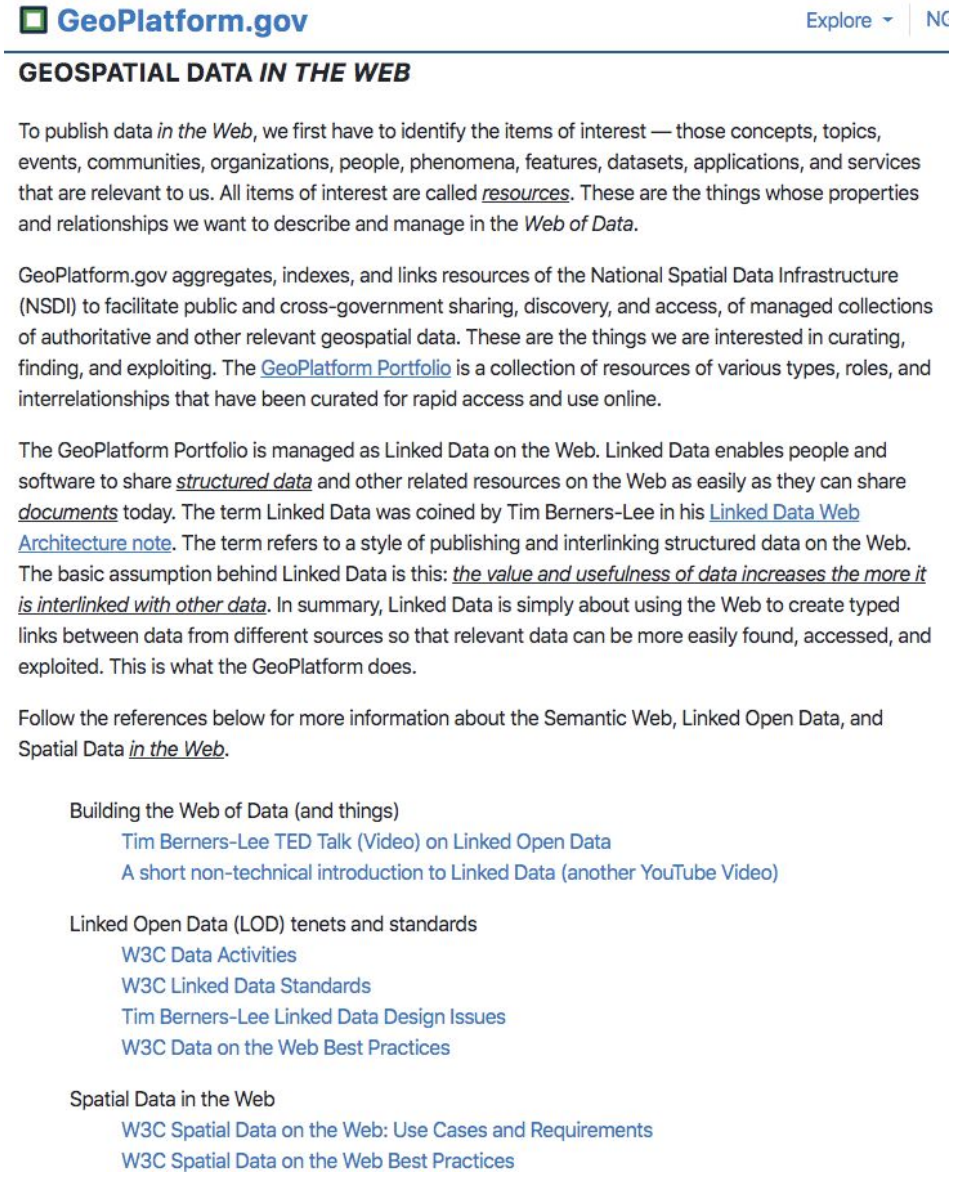
Loc-I Hybrid Approach



https://tinyurl.com/sbq3tjq



GeoPlatform.gov



GeoPlatform.gov Explore NC

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”

Paco Nathan
Known as a "player/coach", with core expertise in data science, natural language processing, machine learning, cloud computing; 35+ years tech industry experience, ranging from Bell Labs to early-stage start-ups. Co-chair Rev and JupyterCon. Advisor for NYU Coleridge Initiative, IBM Data Science Community, Amplify Partners, Anyscale, Recognai, Primer, Data Spartan. Former role: Director, Community Evangelism @ Databricks and Apache Spark. Cited in 2015 as one of the Top 30 People in Big Data and Analytics by Innovation Enterprise.

- Signal or secure email (our clientele)
- contact form (general public)
- Twitter public timeline (avoid DMs)

http://my.pronoun.is/he

0EEC 171D 3A38 7943 9E2E F23D 157E FBCA 16E9 2CF6

orcid.org/0000-0003-3167-1539

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

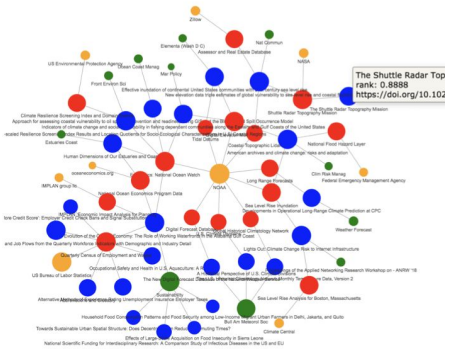
Administrative Data Research Facility

Coleridge Initiative
Julia Lane, et al. NYU Wagner

- FedRAMP-compliant **ADRF framework** 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**

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, Recognai, DLA, Primer AI, GESIS, AllenAI, etc.



https://tinyurl.com/vsq6q9g

Integration of Knowledge Graphs into Computational Toolchains

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

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)

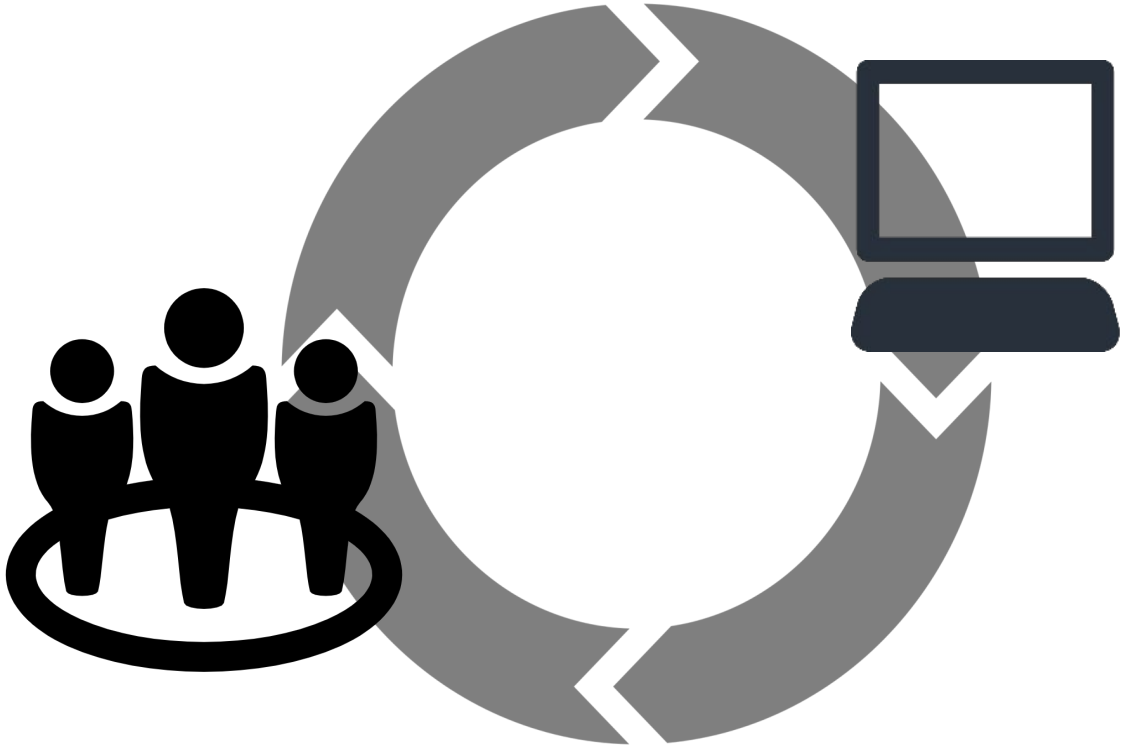
Emerging category: watch the "AI Natives"

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?



TECHNOLOGY + **POLICY** + **CULTURE**

ROADMAP

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

DavidPrevatt

Anne Rosinski

Erutgrul Taciroglu

Anne Wein

Yu Xiao

Missing? A-M Last Name

B ROOM

Michele Barbato

Rodrigo Costa

Wael Elhaddad

MartaGonzalez

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 3

- Trusted and vetted source
- Not secure data transfer
- Mostly interoperable
- Some "down time"
- Limited metadata

ORL 2

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

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

W3C Prov-O Ontology Model

The properties rdf:type and rdfs:label are used to express prov:type and prov:label, respectively.

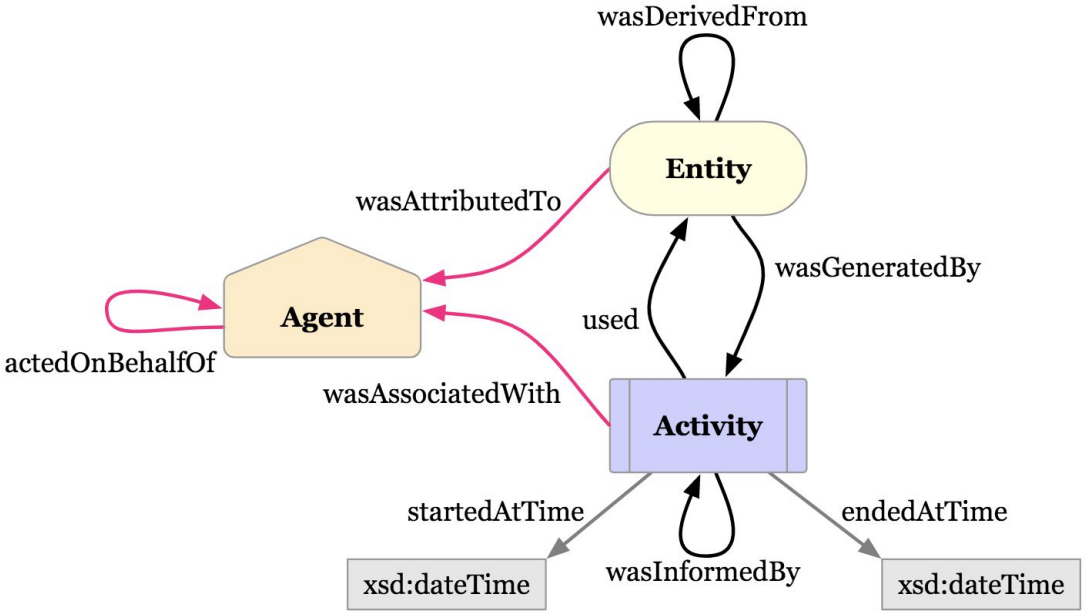


Figure 1. The three Starting Point classes and the properties that relate them.
The diagrams in this document depict Entities as yellow ovals,
Activities as blue rectangles, and Agents as orange pentagons.
The responsibility properties are shown in pink.

https://www.w3.org/TR/prov-o/

USGS Hybrid Experiment

GNIS Linked Data

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.