CI/CS WORKSHOP

THE COMMUNITY TOGETHER

ResearchSOC | CI CoE PILOT

What is FAIR for Research Data?

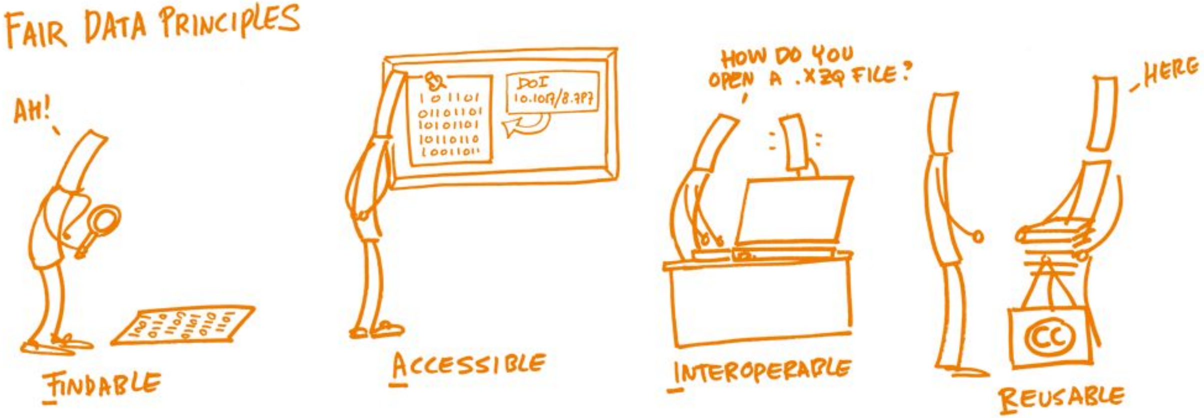
Charles F Vardeman II, Research Assistant Professor, University of Notre Dame

CI/CS WORKSHOP ResearchSOC CI CoE PILOT

Why is it so hard to be FAIR?

…and can we make it less difficult?

What do we mean by FAIR?



https://doi.org/10.5281/zenodo.1212496 CC0 1.0

www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

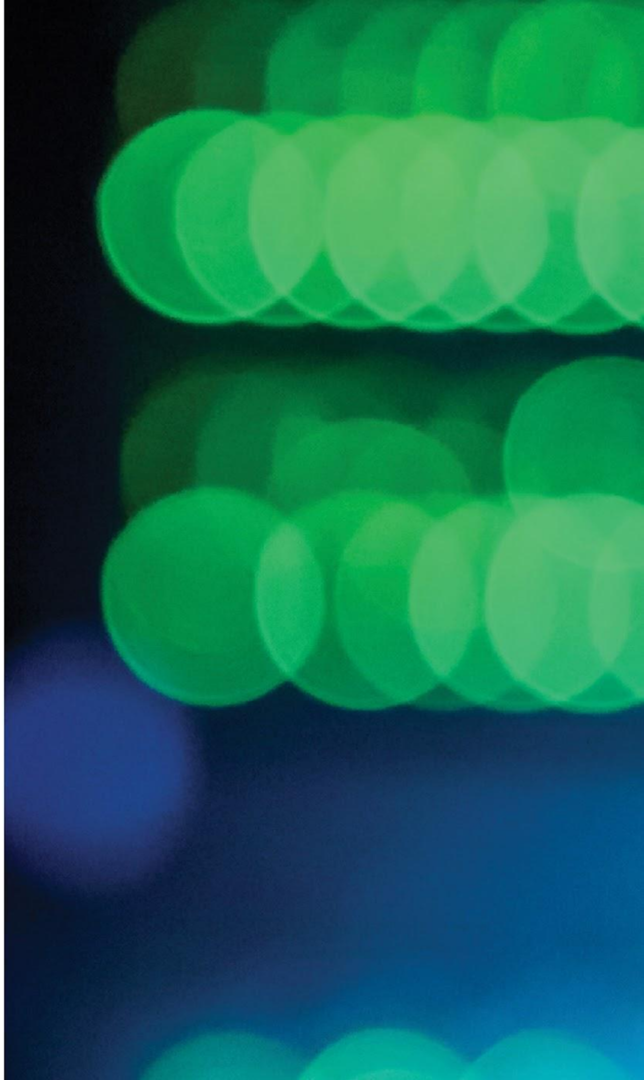
Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

http://dx.doi.org/10.1038/sdata.2016.18



Go FAIR Community

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

https://www.go-fair.org/fair-principles/

Go FAIR Community

Accessible

Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

https://www.go-fair.org/fair-principles/

Go FAIR Community

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

https://www.go-fair.org/fair-principles/

Go FAIR Community

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

- **R1.1. (Meta)data are released with a clear and accessible data usage license**
- **R1.2. (Meta)data are associated with detailed provenance**
- **R1.3. (Meta)data meet domain-relevant community standards**

https://www.go-fair.org/fair-principles/

FAIR Community -- “Not One Size Fits All”

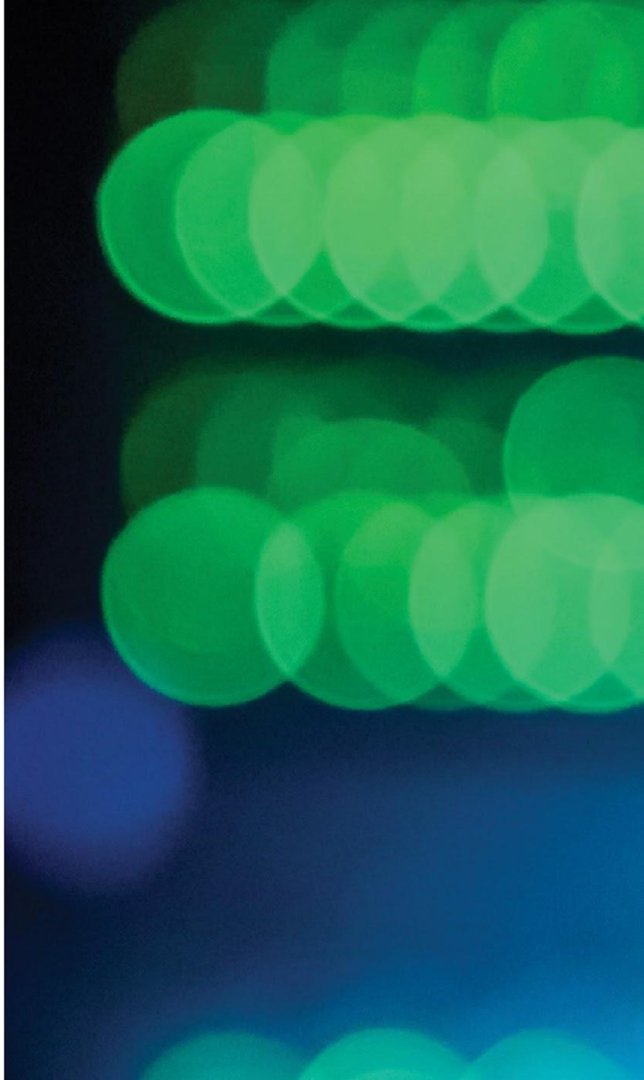
What is an Implementation Network ?

A GO FAIR Implementation Network (IN) is a consortium committed to defining and creating specific materials and tools as elements of the **Internet of FAIR Data and Services (IFDS)**.

INs are the core drivers of GO FAIR as open, inclusive, community-led and self-governed consortia working across disciplines and countries. Individuals, institutions and organisations from all over the world engage within different GO FAIR Implementation Networks.

The INs commit to implementing elements of the Internet of FAIR Data and Services within the three pillars: **GO Build** (Technology), **GO Change** (Culture) and **GO Train** (Training).

https://www.go-fair.org/implementation-networks/



Example Community Network

Chemistry

Home › Implementation Networks › Current Implementation Networks › Chemistry

Active GO FAIR Implementation Network

The Chemistry community needs to create a FAIR culture which is supported by standards and infrastructure development promoting machine readability of chemical data and other digital resources. Hence, the Chemistry Implementation Network (ChIN) commits to the following guiding principles:

- Effecting a change in culture around FAIR data stewardship and data sharing practice
- Findable chemistry data
- Reusable code and data – validation, compilation/aggregation, incorporation into future work, data mining
- The use of standards at source and throughout the information lifecycle
- Availability and accessibility of tools and infrastructure
- The use of persistent identifiers and machine readability at the core
- Management and oversight of standards
- Use of general data standards outside of chemistry where appropriate and FAIR in their implementation
- Enable and promote use of chemical data standards in other disciplines that work with chemical data

The ChIN operates in tandem with the **Chemistry Research Data Interest Group** (CRDIG) of the Research Data Alliance.

“The primary limitation of humans, however, is that we are unable to operate at the scope, scale, and speed necessitated by the scale of contemporary scientific data and complexity of e-Science. It is for this reason that humans increasingly rely on computational agents to undertake discovery and integration tasks on their behalf. “

What FAIR is Not

Home › Resources › FAQ › What FAIR is not...

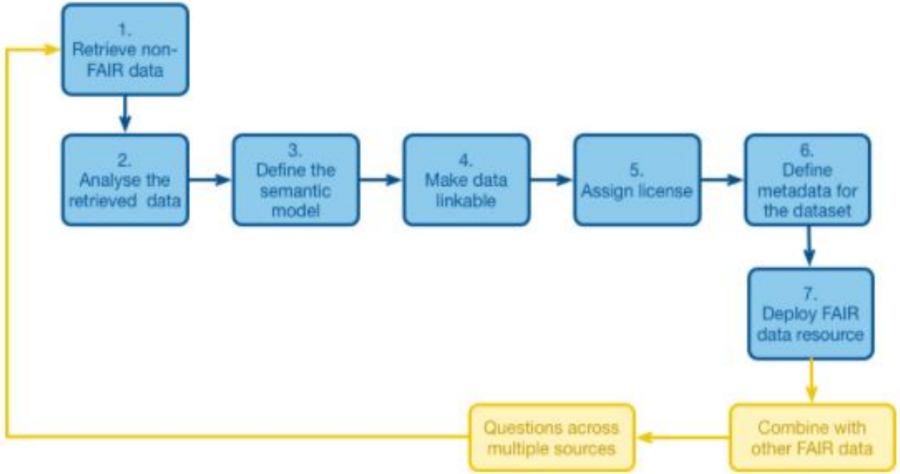
FAIR is **not a standard**, although the acronym is frequently used in that context. The GO FAIR view is that standards are needed for the **Internet of FAIR Data and Services** and that ideally, standards, API's and protocols are developed 'following FAIR guiding principles'.

FAIR is **not equivalent to open** (and open is not equivalent to 'free'): There are many reasons why data may be non-open and only available under certain conditions to certain users, including machines. As long as the accessibility conditions are properly described, non-open data can be entirely FAIR. Reciprocally, fully open and unrestricted data may score very low in FAIR metrics as they may for instance be non-actionable for machines.

FAIR principles do not, in themselves, cover the crucial aspects of **intrinsic data quality or ethics**. However, FAIR guiding principles request that optimal care is taken to enable users to determine the 'usefulness' (for their purpose) of the data and other research objects they find, which includes rich, machine readable provenance. Obviously, user defined metadata and comments on existing research objects will be increasingly useful to judge the reusability of the research objects.

FAIRification Process

The **FAIR Data Principles** apply to metadata, data, and supporting infrastructure (e.g., search engines). Most of the requirements for findability and accessibility can be achieved at the metadata level. Interoperability and reuse require more efforts at the data level. The scheme below depicts the FAIRification process adopted by GO FAIR, focusing on data, but also indicating the required work for metadata:

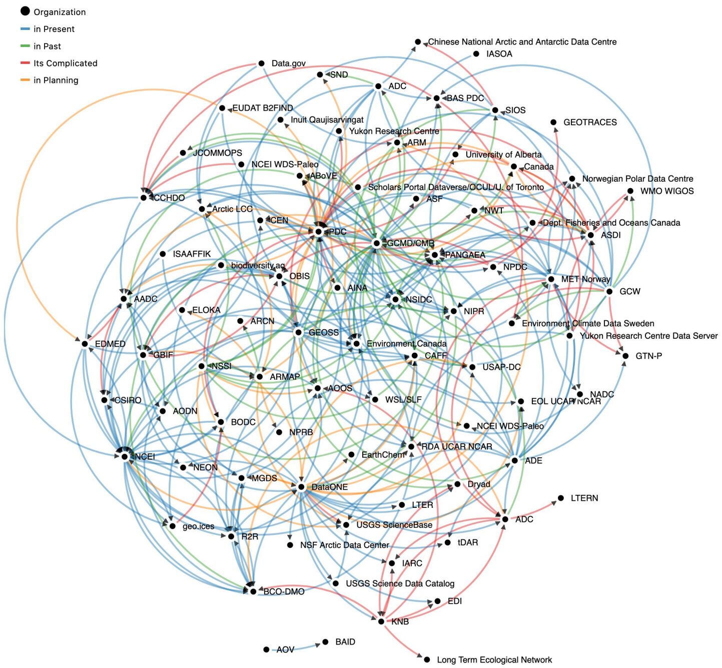


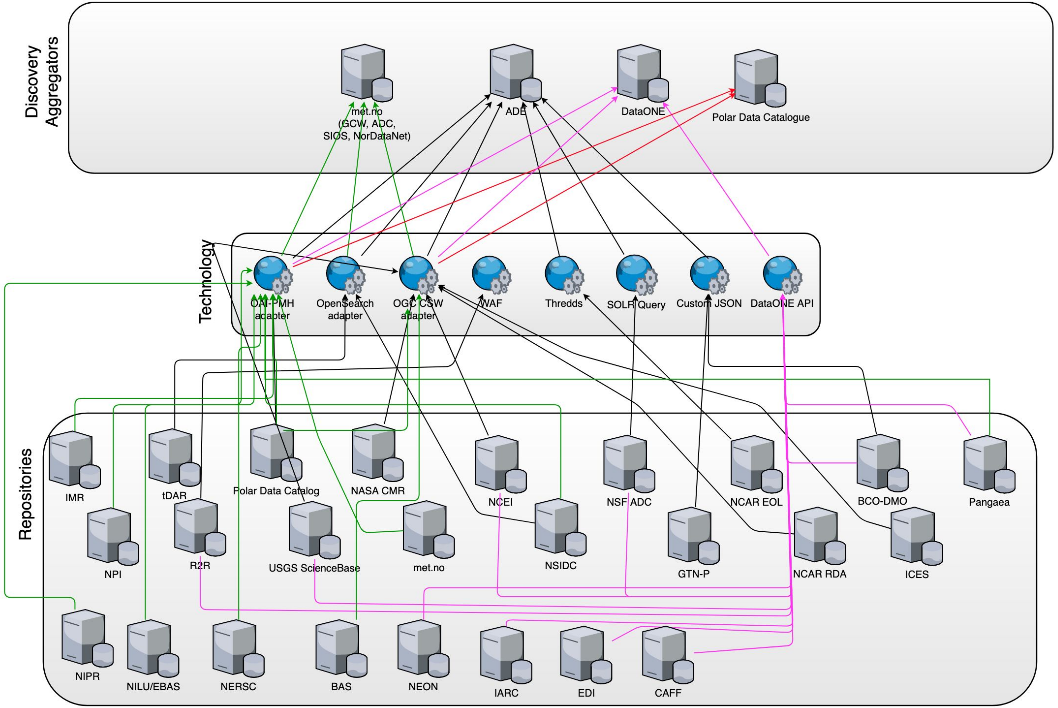
https://www.go-fair.org/fair-principles/fairification-process/

So now, what’s the problem?

Durr, R. Structured data on the web: putting best practice to work, ESIP Summer Meeting 2020 [2020 ESIP Summer Meeting: Structured data on the web: putting best...]

- Standard metadata isn’t standard!
- There are many repositories providing “standard” metadata in incompatible ways (81 identified with polar data)
- There are many aggregators harvesting information which often has already been aggregated by some other group!
- There are many technologies in use





A partial polar landscape map from 2018

Problem 1: Driving force towards <u>Centralization</u>

Linked Data As a Solution?

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

https://www.w3.org/DesignIssues/LinkedData.html

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.

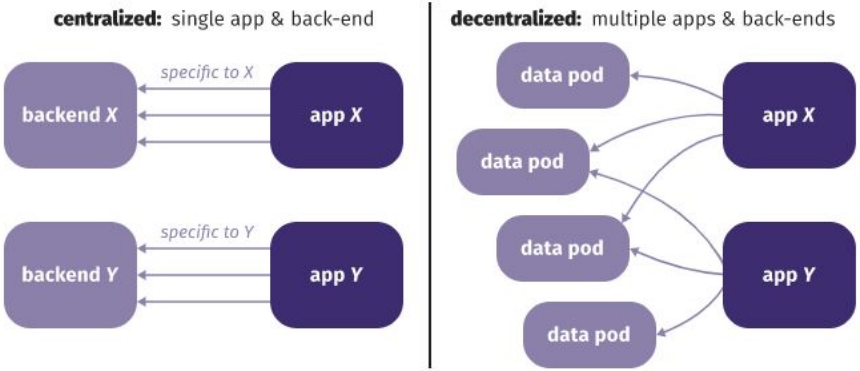
28 December 2018

https://ruben.verborgh.org/blog/2018/12/28/designing-a-linked-data-developer-experience/

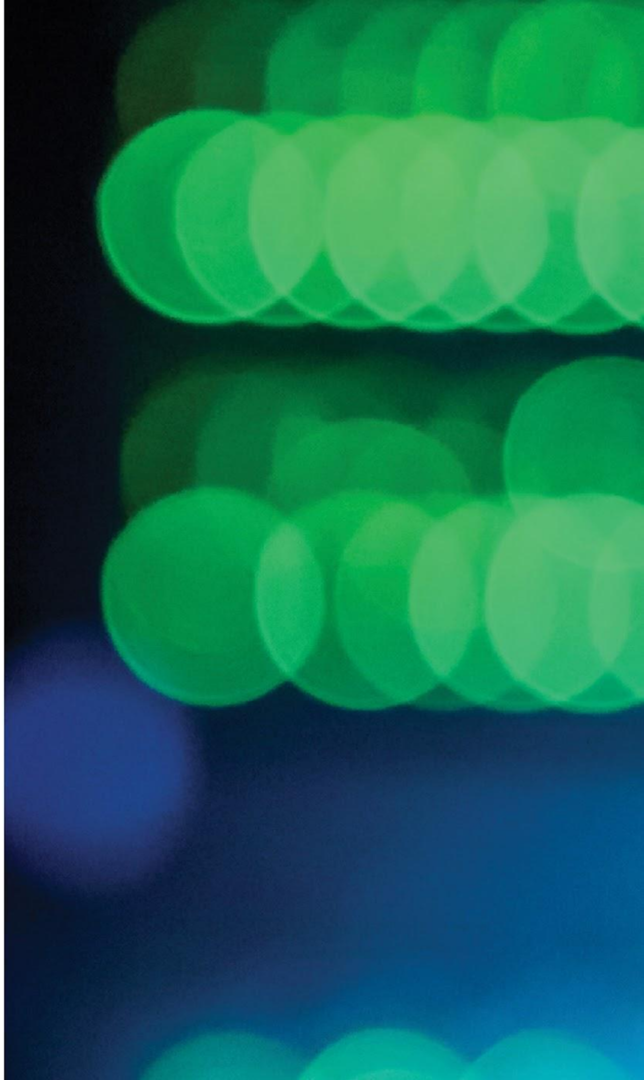
Why Linked Data?

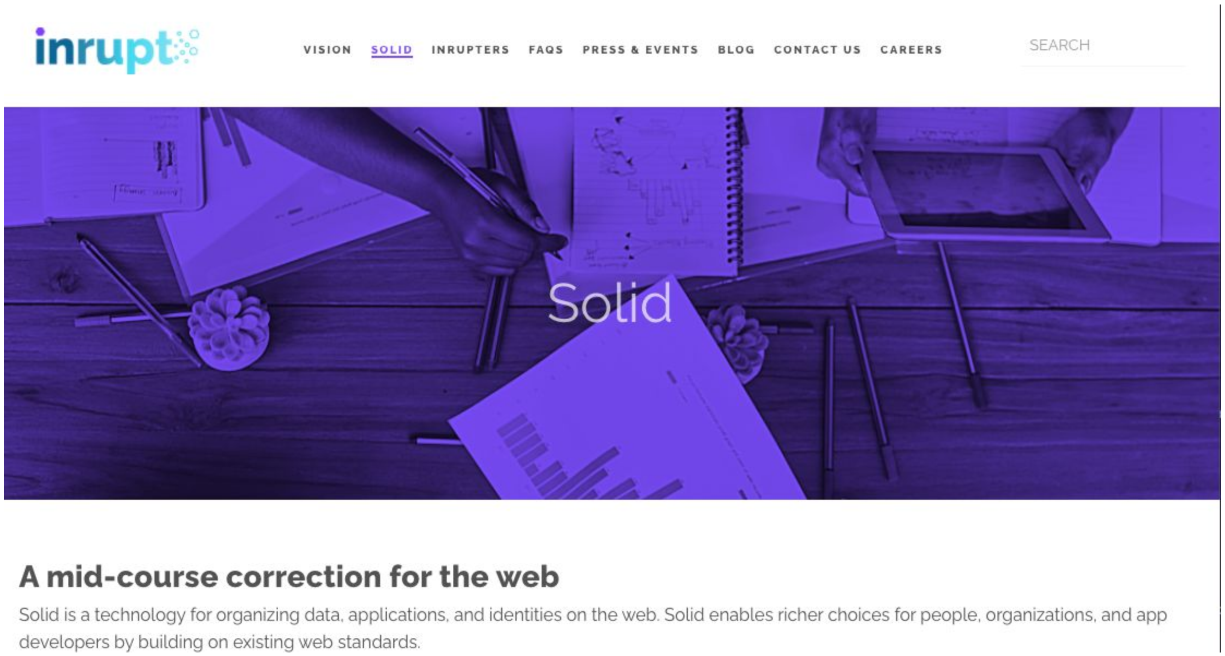
Decentralized Web apps have multiple back-ends

A crucial first question is whether decentralized Web apps need Linked Data at all. Why not just do like every other Web API, where the server sends custom JSON that the client can easily decipher? The idea behind decentralization in Solid is that **apps do *not* have their own data store**. Data is instead stored in a place of the user's choice. Apps thus need to be more flexible in order to become compatible with different back-ends. **Multiple back-ends** might be used at the same time by **multiple apps**. For instance, social media apps show data of multiple profiles, and every profile in a decentralized network can be stored in a different place.



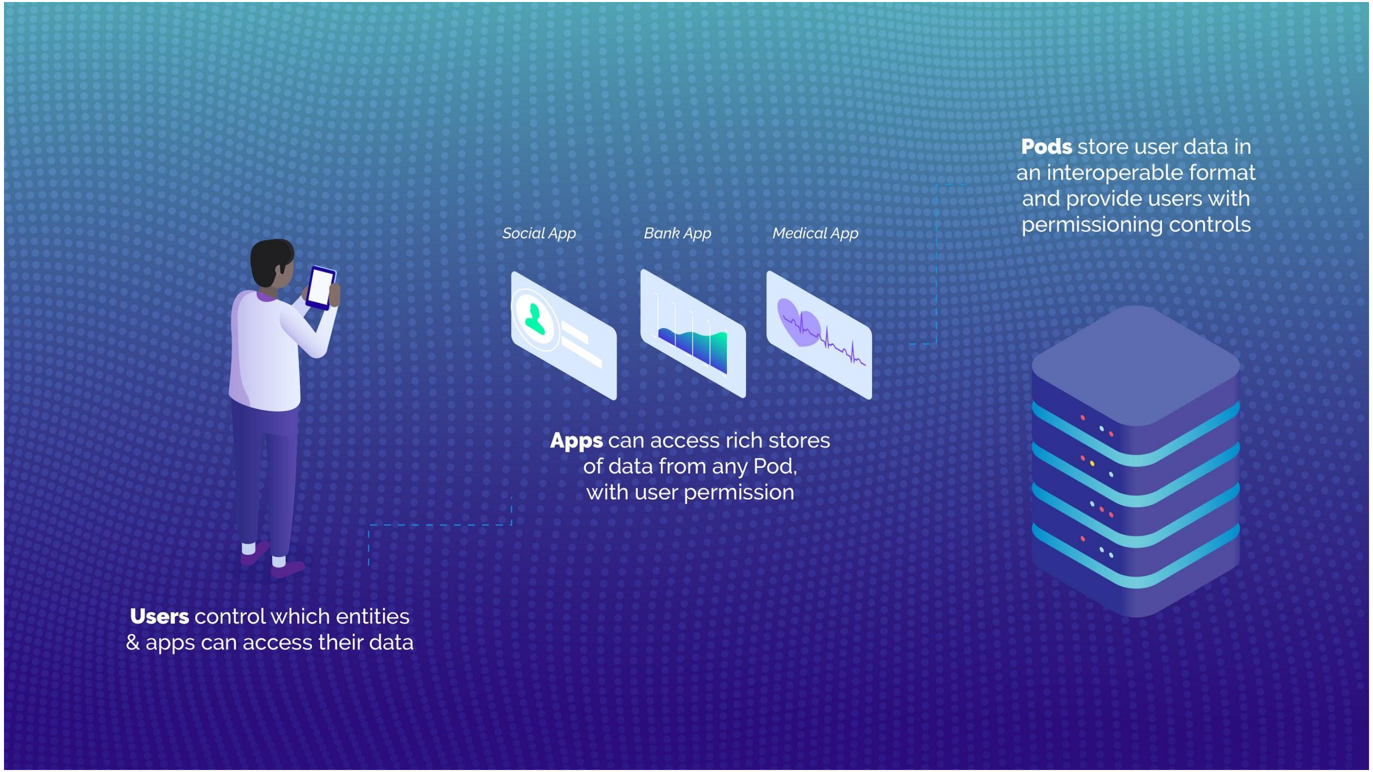
https://ruben.verborgh.org/blog/2018/12/28/designing-a-linked-data-developer-experience/





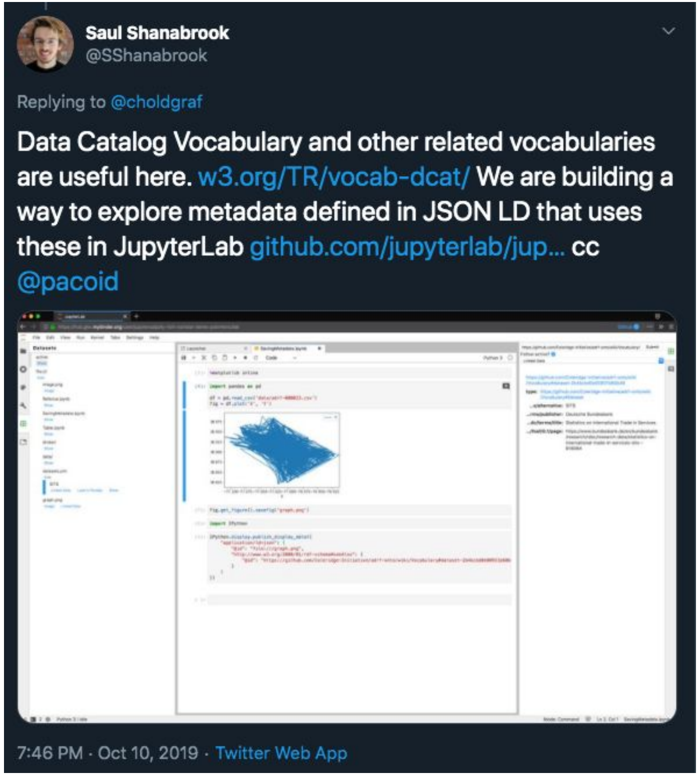
https://inrupt.com/solid

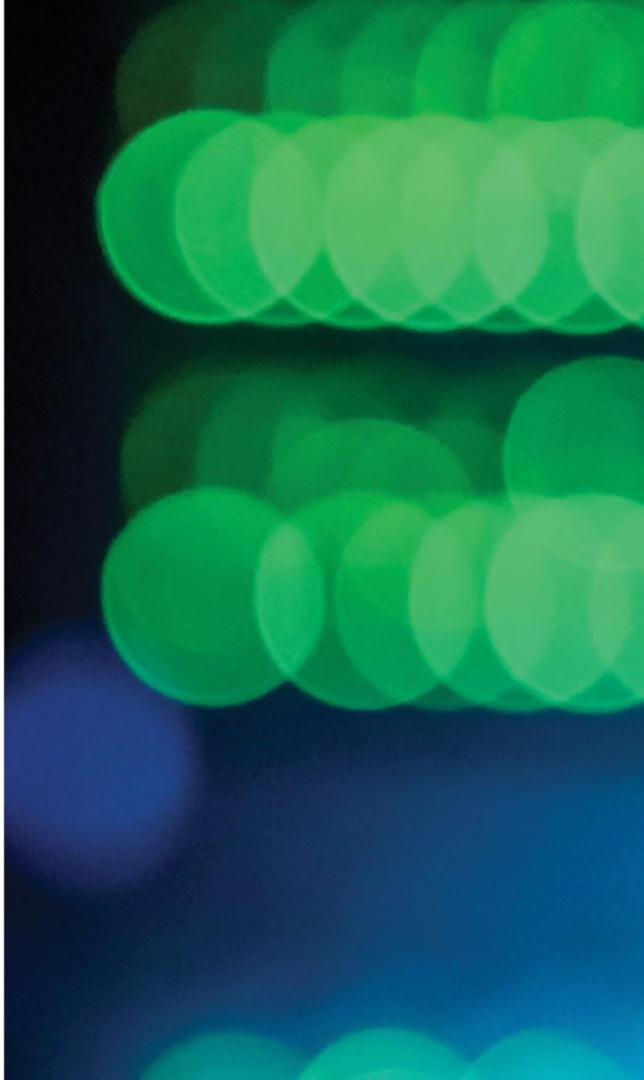
Inrupt Solid Idea



https://inrupt.com/solid

jupyter





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

Paco Nathan's Excellent talk on Deep Context and AI:

https://esip.figshare.com/articles/presentation/Rich_Context_providing_support_for_cross-agency_data_stewardship_and_measuring_dataset_impact_on_public_policy/11573343

JOSS
The Journal of Open Source Software

Ecological Metadata as Linked Data

Carl Boettiger[1]

1 University of California, Berkeley

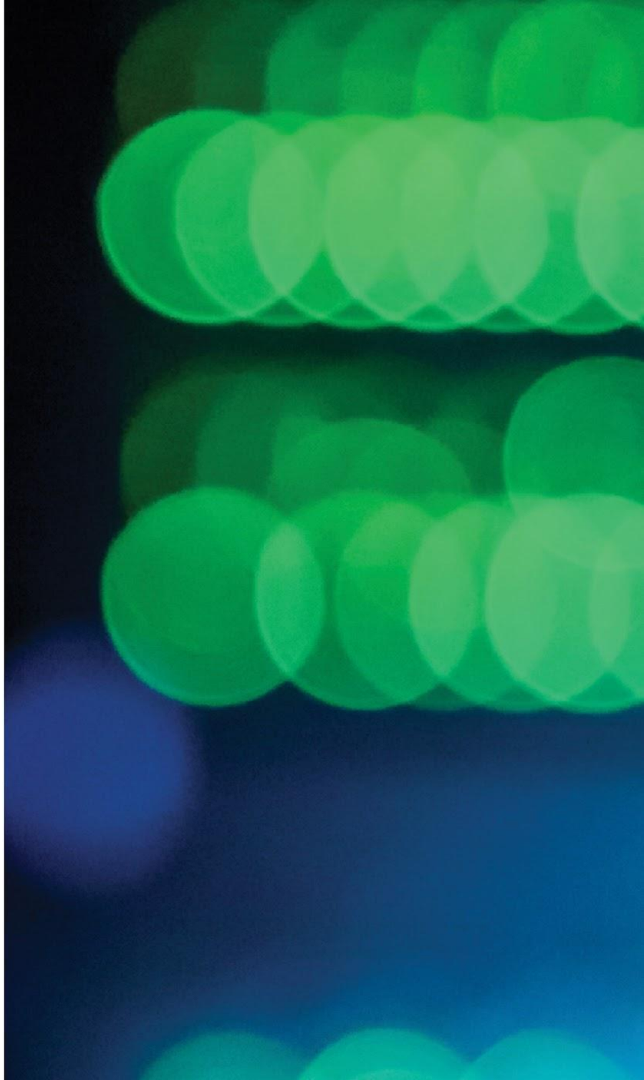
DOI: 10.21105/joss.01276

rOpenSci

This package is part of the rOpenSci project

To learn more, please visit http://ropensci.org

https://github.com/ropensci/emld



AIMS

SEARCH CONNECT SIGN UP LOGIN

GET INVOLVED | ACCESS TO INFORMATION | AGROVOC | SEMANTICS | ACTIVITIES | THEMES | ABOUT

FAIR data: What and Why? Easier said than implemented?

(Image credit: Farm Data Train project that aims to connect agricultural data to make them more usable. It is a joint initiative of GODAN, CABI, Wageningen UR and DTL).

In 2016, a Nature article "The FAIR Guiding Principles for scientific data management and stewardship" launched the FAIR concept.

FAIR stands for Findable, Accessible, Interoperable, Re-usable principles (The FAIR principles as published by FORCE11). The FAIR Data principles act as an international guideline for high quality data stewardship. Throughout the FAIR Principles, we use the phrase '(meta)data' in cases where the Principle should be applied to both metadata and data.

AREAS OF INTEREST

advocacy

research information and data available and accessible

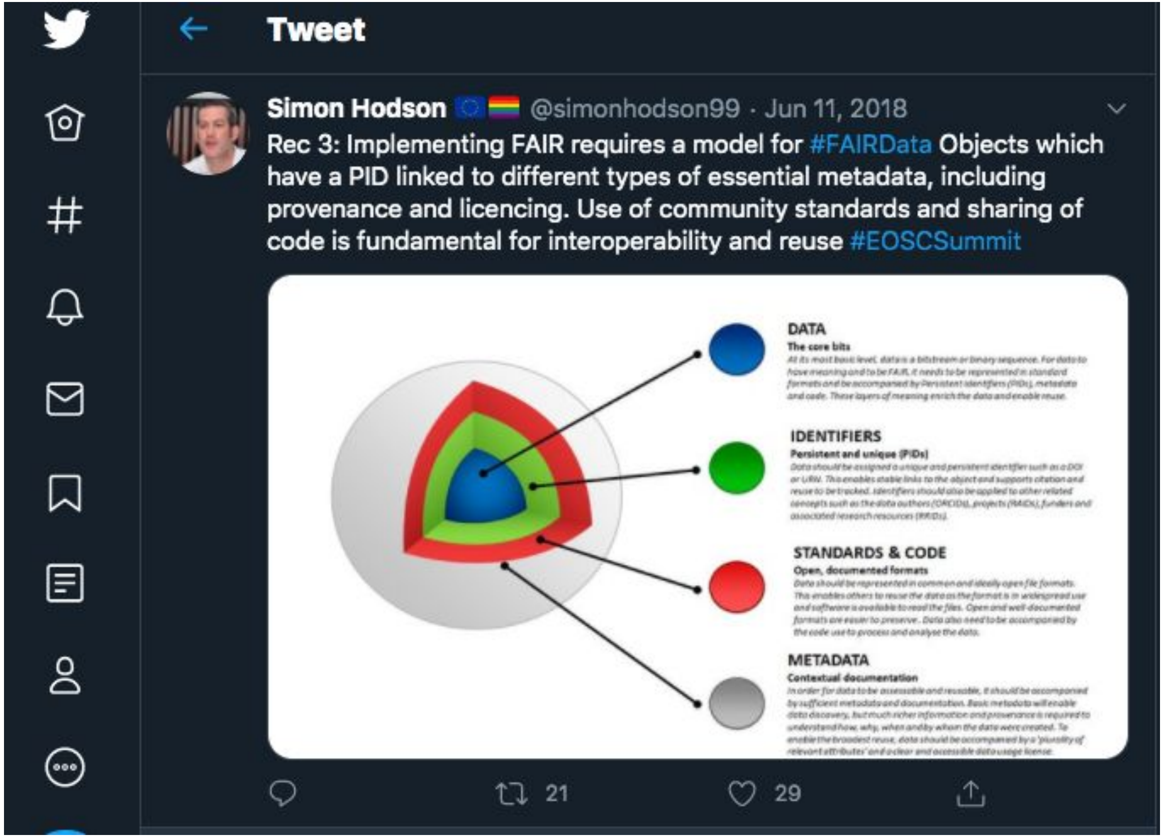
data

fostering knowledge exchange

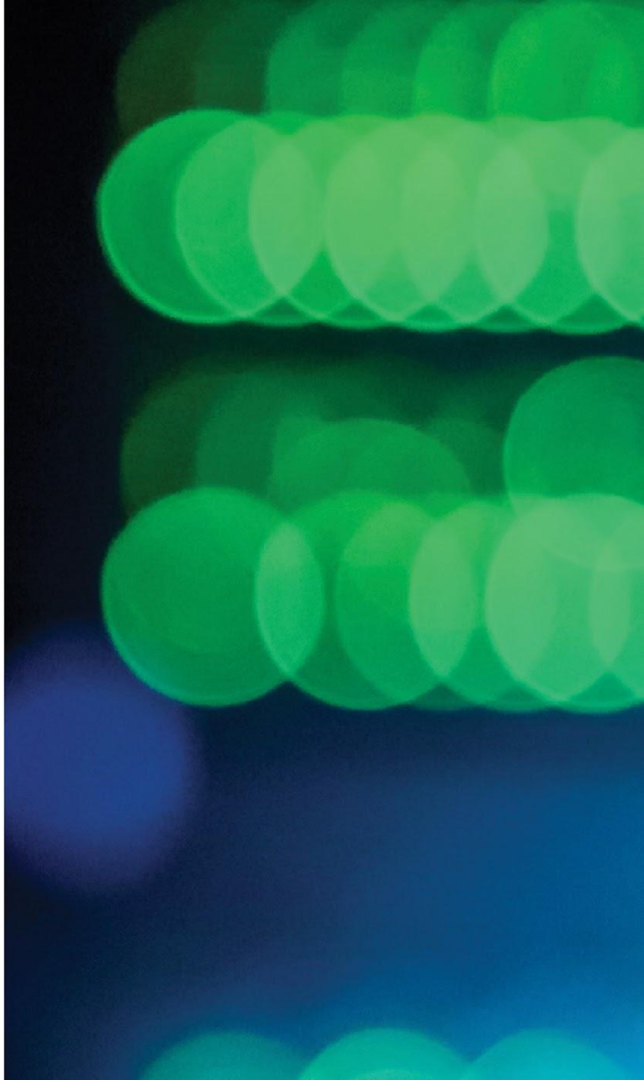
SHARE

5

http://aims.fao.org/activity/blog/fair-data-what-and-why-easier-said-implemented



https://twitter.com/simonhodson99/status/1006106264215195648



Problem 2: Persistent Identifiers Are Hard to Implement (as Web Resources)

Linked Data As a Solution?

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

https://www.w3.org/DesignIssues/LinkedData.html

URI Resources also have semantics themselves as to <u>HOW</u> they identify entities

PLOS BIOLOGY

BROWSE PUBLISH ABOUT SEARCH

advanced search

PERSPECTIVE

Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data

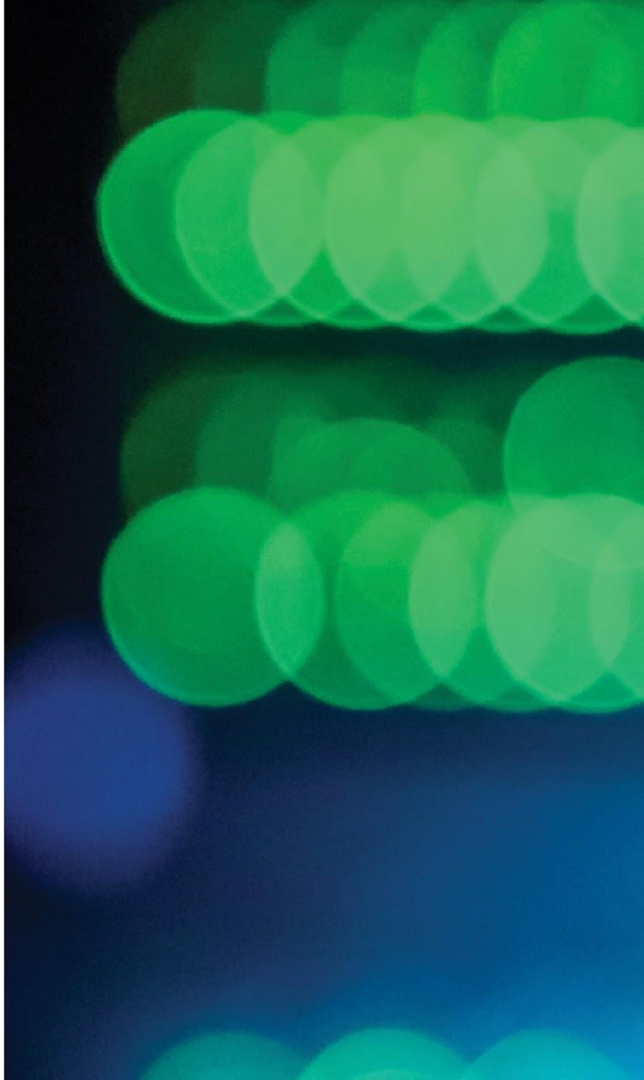
Julie A. McMurry, Nick Juty, Niklas Blomberg, Tony Burdett, Tom Conlin, Nathalie Conte, Mélanie Courtot, John Deck, Michel Dumontier, Donal K. Fellows, Alejandra Gonzalez-Beltran, Philipp Gormanns, Jeffrey Grethe, [...], Helen Parkinson
[view all]

Published: June 29, 2017 • https://doi.org/10.1371/journal.pbio.2001414 • >> See the preprint

Abstract

In many disciplines, data are highly decentralized across thousands of online databases (repositories, registries, and knowledgebases). Wringing value from such databases depends on the discipline of data science and on the humble bricks and mortar that make integration possible; identifiers are a core component of this integration infrastructure. Drawing on our experience and on work by other groups, we outline 10 lessons we have learned about the identifier qualities and best practices that facilitate large-scale data integration. Specifically, we propose actions that identifier practitioners (database providers) should take in the design, provision and reuse of identifiers. We also outline the important considerations for those referencing identifiers in various circumstances, including by authors and data generators. While the importance and relevance of each lesson will vary by context, there is a need for increased awareness about how to avoid and manage common identifier problems, especially those related to persistence and web-accessibility/resolvability. We focus strongly on web-based identifiers in the life sciences; however, the principles are broadly relevant to other disciplines.

https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2001414



Legend

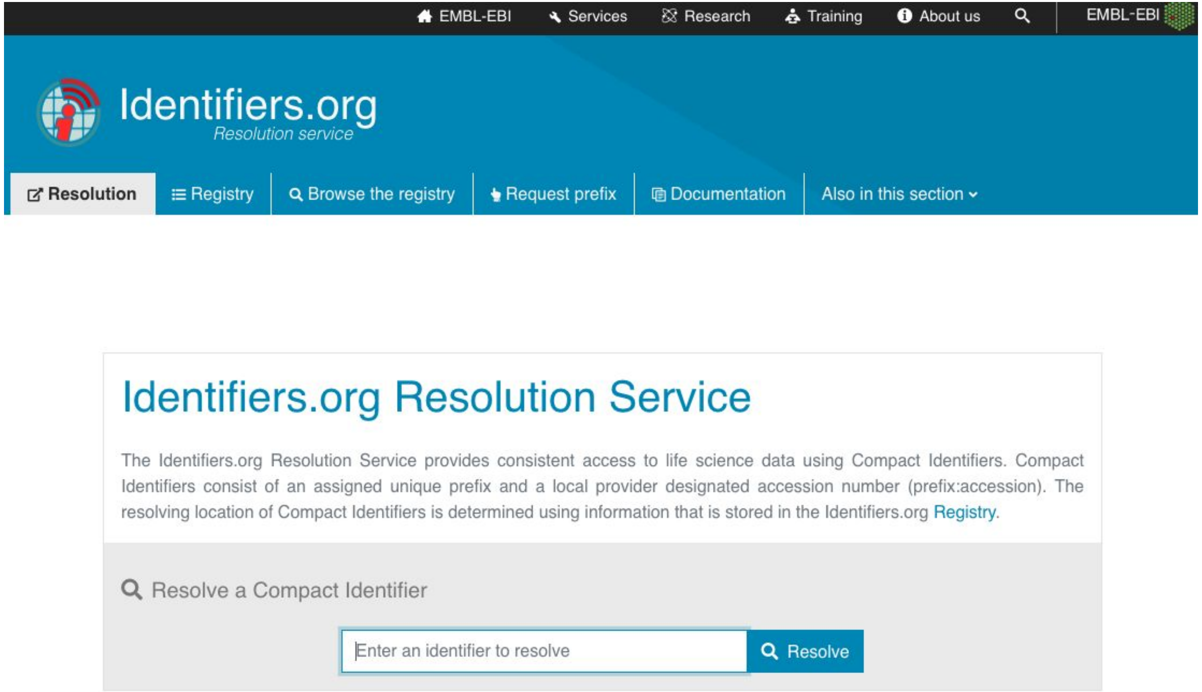
◒ Indirectly relevant for

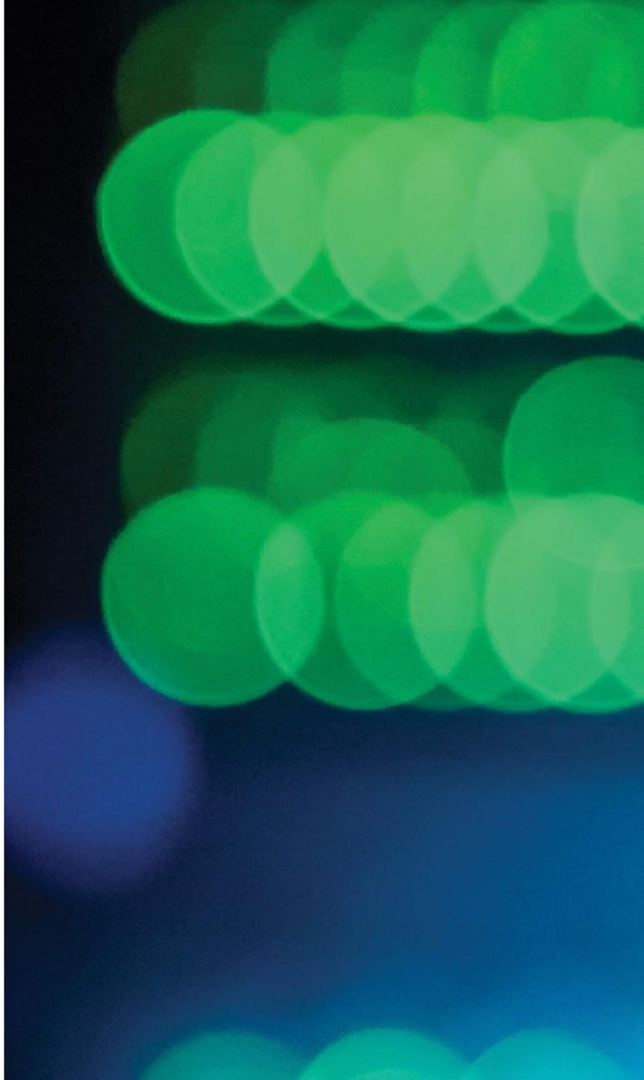
● Directly relevant for

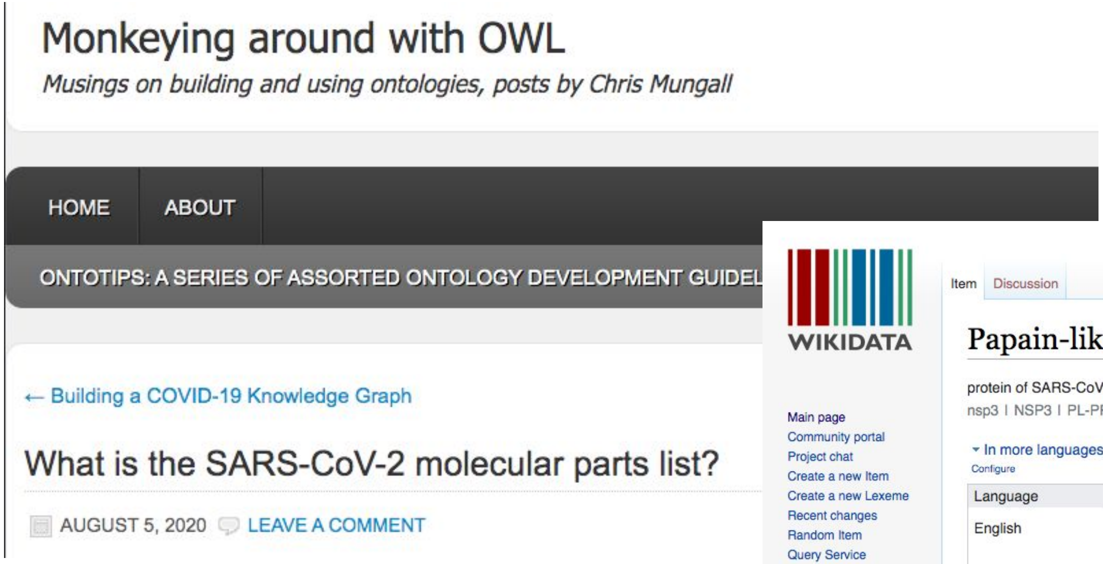
Lesson	Types of actors: Designers & creators	Types of actors: Providers & mainteners	Types of actors: Reusers & referencers
Lesson 1. Credit any derived content using its original identifier	●	◒	●
Lesson 2. Help local identifiers travel well: document prefix and patterns	●	●	●
Lesson 3. Opt for simple, durable web resolution	●	◒	◒
Lesson 4. Avoid embedding meaning, or relying on it for uniqueness	●	●	◒
Lesson 5. Design new identifiers for diverse uses by others	●	●	◒
Lesson 6. Implement a version-management policy	●	●	◒
Lesson 7. Do not reassign or delete identifiers	●	●	◒
Lesson 8. Make URIs clear and findable	●	●	◒
Lesson 9. Document the identifiers you issue and use	●	●	●
Lesson 10. Reference and display responsibly	●	●	●

https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2001414

Identifiers.org







Monkeying around with OWL

Musings on building and using ontologies, posts by Chris Mungall

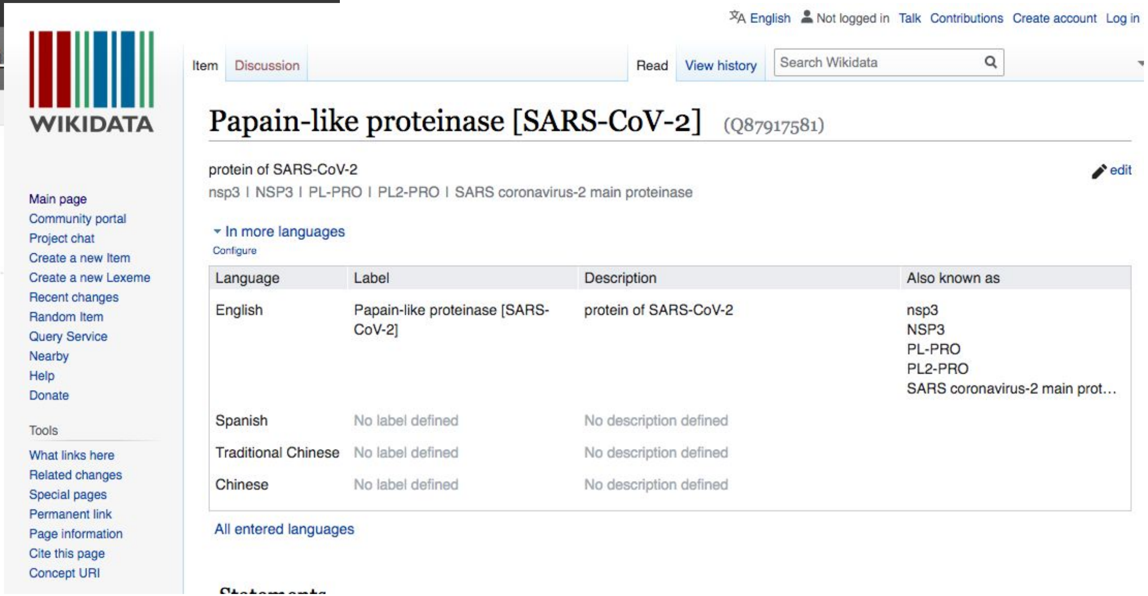
HOME ABOUT

ONTOTIPS: A SERIES OF ASSORTED ONTOLOGY DEVELOPMENT GUIDEL

← Building a COVID-19 Knowledge Graph

What is the SARS-CoV-2 molecular parts list?

AUGUST 5, 2020 LEAVE A COMMENT



English Not logged in Talk Contributions Create account Log in

WIKIDATA

Item Discussion Read View history Search Wikidata

Main page
Community portal
Project chat
Create a new Item
Create a new Lexeme
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Tools

What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Concept URI

Papain-like proteinase [SARS-CoV-2] (Q87917581)

protein of SARS-CoV-2

nsp3 | NSP3 | PL-PRO | PL2-PRO | SARS coronavirus-2 main proteinase

edit

In more languages
Configure

Language	Label	Description	Also known as
English	Papain-like proteinase [SARS-CoV-2]	protein of SARS-CoV-2	nsp3 NSP3 PL-PRO PL2-PRO SARS coronavirus-2 main prot…
Spanish	No label defined	No description defined	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

All entered languages

https://douroucouli.wordpress.com/2020/08/05/what-is-the-sars-cov-2-molecular-parts-list/
https://www.nytimes.com/interactive/2020/04/03/science/coronavirus-genome-bad-news-wrapped-in-protein.html
https://www.wikidata.org/wiki/Q87917581

W3IDs persistently redirect to sustainable resources like GitHub

Permanent Identifiers for the Web

Secure, permanent URLs for your Web application that will stand the test of time.

The purpose of this website is to provide a secure, permanent URL re-direction service for Web applications. This service is run by the W3C Permanent Identifier Community Group.

Web applications that deal with Linked Data often need to specify and use URLs that are very stable. They utilize services such as this one to ensure that applications using their URLs will always be re-directed to a working website. This website operates like a switchboard, connecting requests for information with the true location of the information on the Web. The switchboard can be reconfigured to point to a new location if the old location stops working.

There are a growing group of organizations that have pledged responsibility to ensure the operation of this website. These organizations are: Digital Bazaar, 3 Round Stones, OpenLink Software, Applied Testing and Technology, Openspring, and Bosatsu Consulting. They are responsible for all administrative tasks associated with operating the service. The social contract between these organizations gives each of them full access to all information required to maintain and operate the website. The agreement is setup such that a number of these companies could fail, lose interest, or become unavailable for long periods of time without negatively affecting the operation of the site.

This website operates in HTTPS-only mode to ensure end-to-end security. This means that it may be used for Linked Data applications that require high levels of security such as those found in the financial, medical, and public infrastructure sectors.

All identifiers associated with this website are intended to be around for as long as the Web is around. This means decades, if not centuries. If the final destination for popular identifiers used by this service fail in such a way as to be a major inconvenience or danger to the Web, the community will mirror the information for the popular identifier and setup a working redirect to restore service to the rest of the Web.

https://w3id.org/

Problem 3: Machine readable interfaces

The interface for the digital object is not the object

Use of Standardized APIs and Landing Pages

Second Environmental Linked Features Experiment:

Publication Date: YYYY-MM-DD

Approval Date: YYYY-MM-DD

Submission Date: YYYY-MM-DD

Reference number of this document: OGC 20-067

Reference URL for this document: http://www.opengis.net/doc/PER/SELFIE-ID

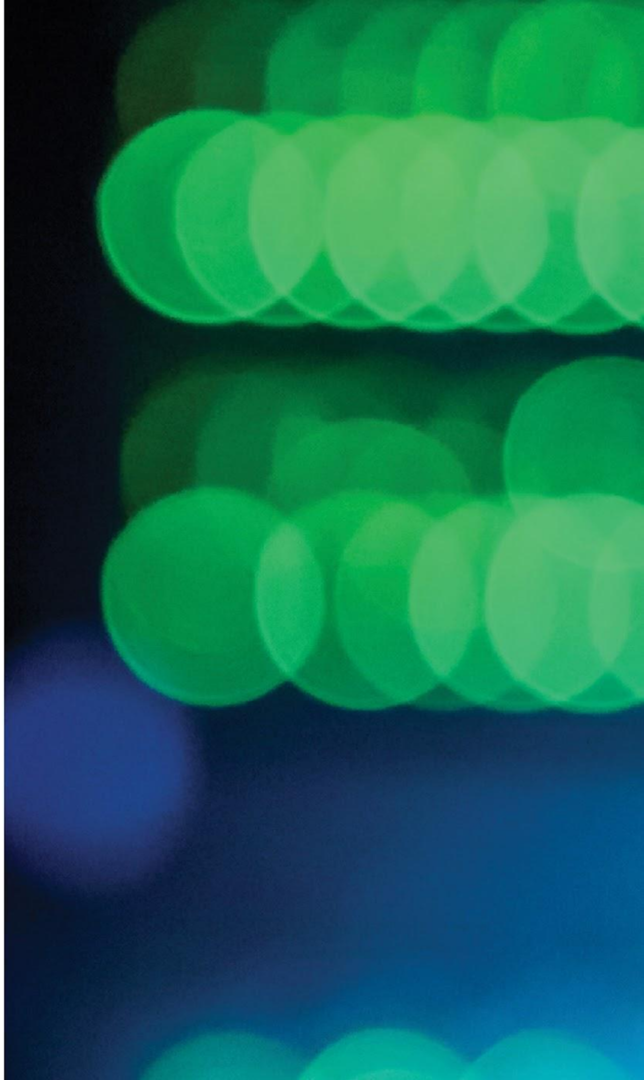
Category: OGC Public Engineering Report

Editor: David Blodgett

Title: Second Environmental Linked Features Experiment:

---

https://github.com/opengeospatial/SELFIE



Opengeospatial Consortium SELFIE

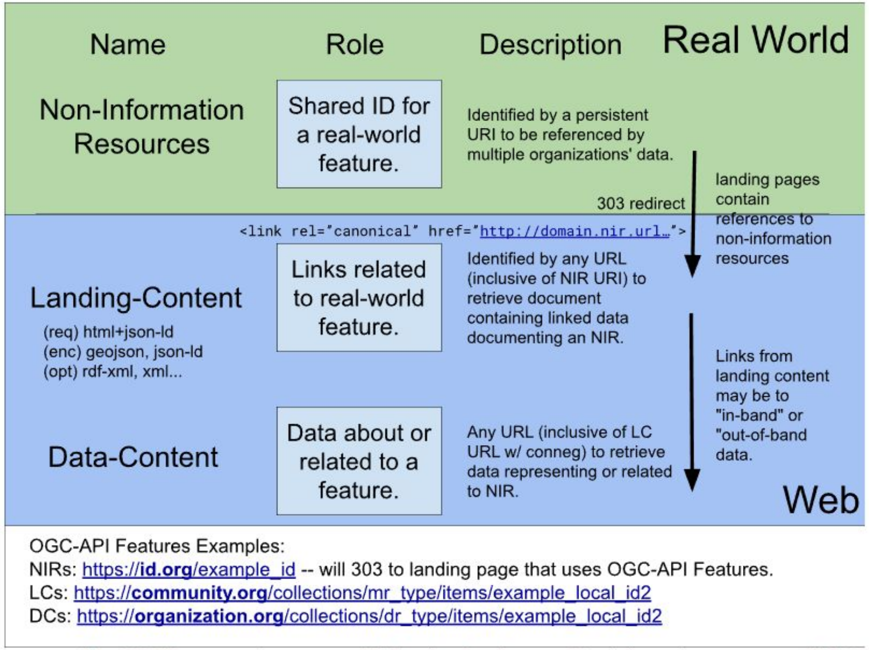
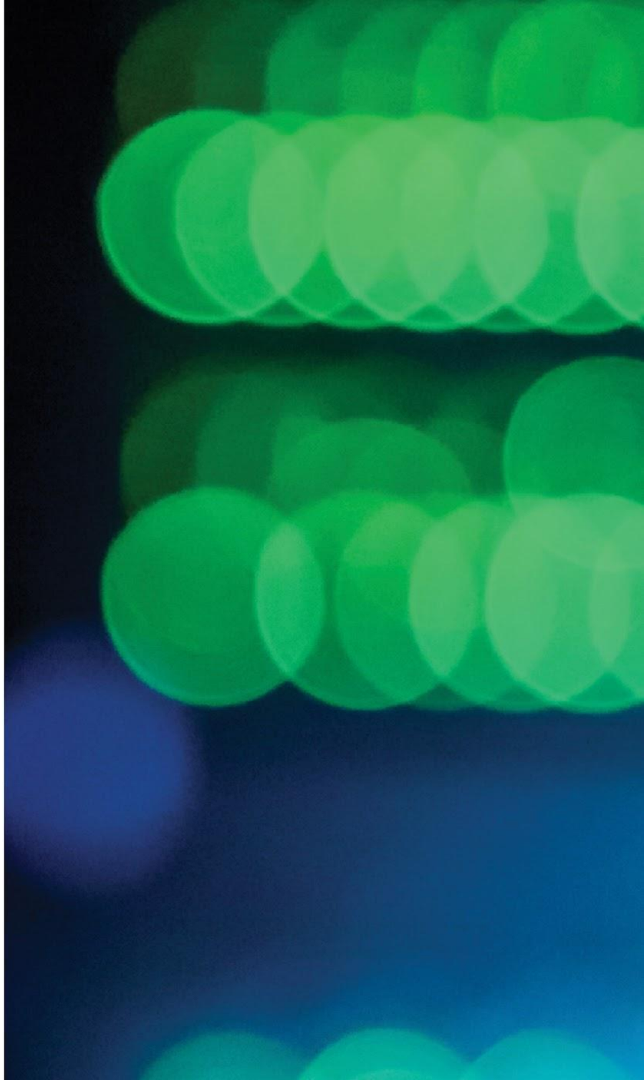


Figure 2. Summary of the SELFIE resource / content model showing that there are Non-information resources which 303 redirect to a resource intended to provide "landing content". The distinction between landing-content and data-content is use-case specific and methods for negotiating between the two is left for future work.



Content Negotiation by Profile

W3C Working Draft 26 November 2019

This version:
https://www.w3.org/TR/2019/WD-dx-prof-conneg-20191126/

Latest published version:
https://www.w3.org/TR/dx-prof-conneg/

Latest editor's draft:
https://w3c.github.io/dxwg/connegp/

Test suite:
https://github.com/w3c/prof-conneg-testing

Implementation report:
https://w3c.github.io/dxwg/connegp-implementation-report/

Previous version:
https://www.w3.org/TR/2019/WD-dx-prof-conneg-20190430/

Latest Recommendation:
https://www.w3.org/TR/2019/WD-dx-prof-conneg-20190430/

Editors:
Lars G. Svensson (Invited Expert and before at Deutsche Nationalbibliothek)
Rob Atkinson (Metalinkage, Open Geospatial Consortium)
Nicholas J. Car (SURROUND Australia Pty Ltd and before at CSIRO)

Participate:
GitHub w3c/dxwg
File a bug
Commit history
Pull requests

Contributors:
Ruben Verborgh

https://www.w3.org/TR/dx-prof-conneg/

Example: Australian Geographic Name Service

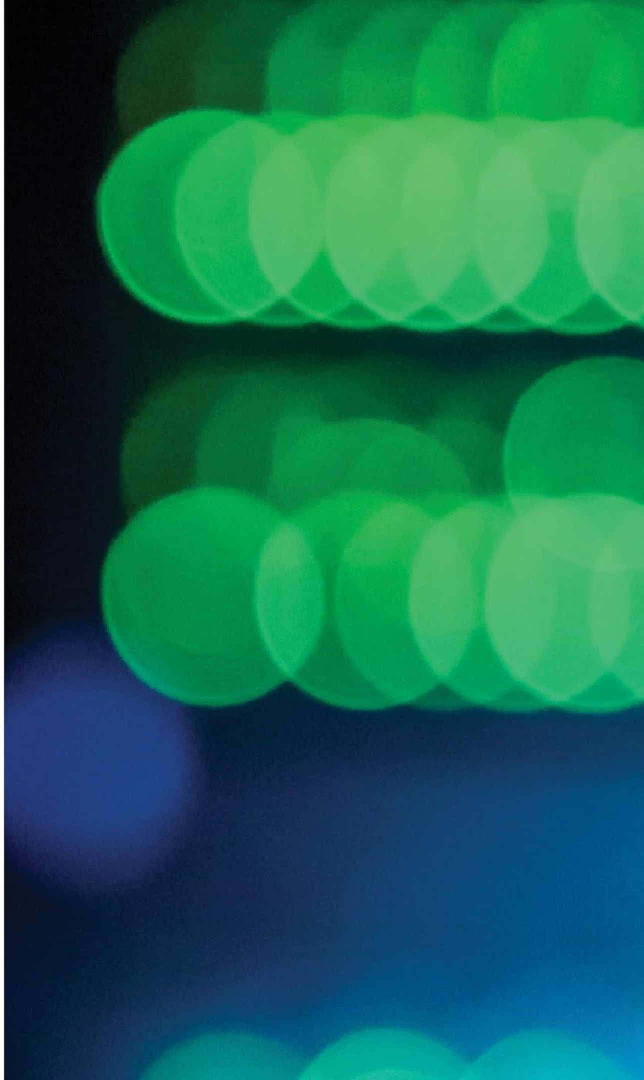
Address GAACT714845933

G-NAF View

Property	Value
Address Line	6 Packham Place, Charnwood, ACT 2615
First Street Number	6
Street Locality	Packham Place
Locality	Charnwood
State/Territory	ACT
Postcode	2615
Legal Parcel ID	BELC/CHAR/15/16/
Address Site PID	710446419
Level Geocoded Code	7
GNAF Confidence	Confidence level 2
Date Created	2004-04-29
Date Last Modified	2018-02-01
Geometry	Frontage Centre Setback → <http://www.opengis.net/def/crs/EPSG/0/4283> POINT(149.03865604 -35.20113263)
Mesh Blocks 2011	Parcel Level Match → 80006300000
Mesh Blocks 2016	Parcel Level Match → 80006300000

Other views

https://gnafld.net/address/GAACT714845933



https://gnafld.net/address/GAACT714845933?_view=alternates

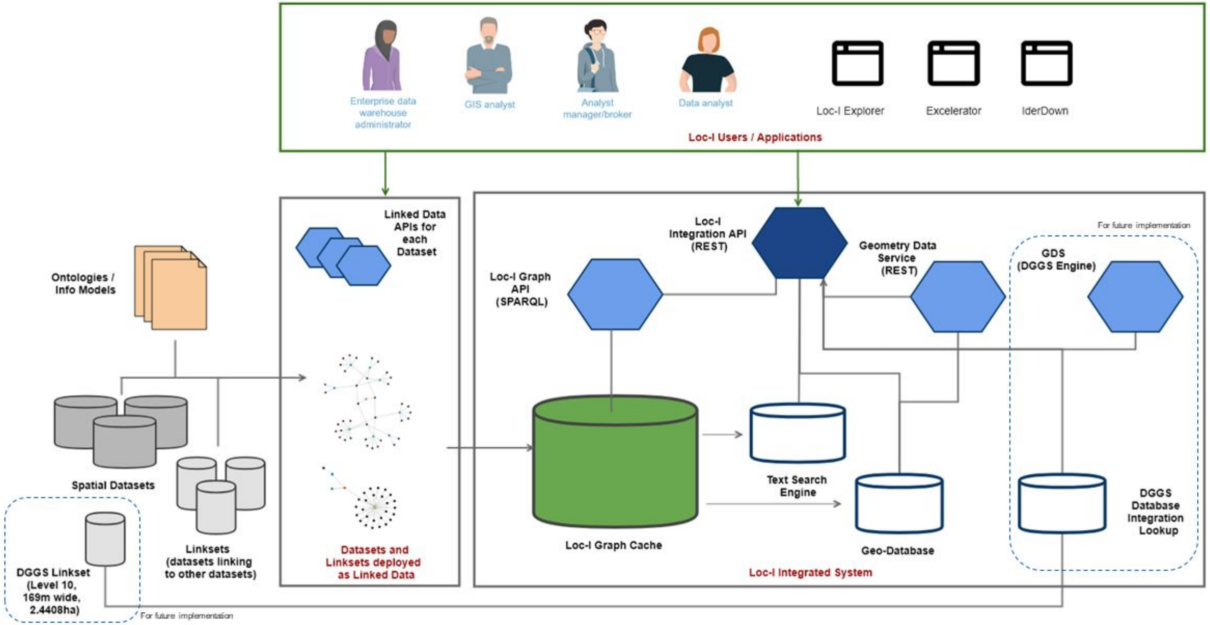
Address API

Alternates view of a http://linked.data.gov.au/def/gnaf#Address

Instance http://linked.data.gov.au/dataset/gnaf/address/GAACT714845933

View	Formats	View Namespace	View Desc.
schemaorg	**application/ld+json** *	http://schema.org	An initiative by Bing, Google and Yahoo! to create and support a common set of schemas for structured data markup on web pages. It is serialised in JSON-LD
gnaf *	**text/html** * text/turtle application/rdf+xml application/ld+json application/n-triples application/xml	http://reference.data.gov.au/def/ont/gnaf/	G-NAF web page view. A simple human-readable, web page-only view, based on the data model of PSMA's G-NAF as of August, 2017
dct	**text/html** * text/turtle application/rdf+xml application/ld+json application/n-triples application/xml	http://purl.org/dc/terms/	Dublin Core Terms from the Dublin Core Metadata Initiative
alternates	**text/html** * application/json text/turtle application/rdf+xml application/ld+json text/n3 application/n-triples	https://promsns.org/def/alt	The view that lists all other views
ISO19160	**text/turtle** * application/rdf+xml application/ld+json	http://reference.data.gov.au/def/ont/iso19160-1-address	The OWL ontology view of the ISO 19160-1:2015 Address standard from the OGC's TC211 UML to OWL mapping: https://github.com/ISO-

Hybrid approaches that store metadata in Knowledge Graphs



http://loci.cat/

Problem 4: Which Vocabularies?

Problem 4: Vocabularies are also need to be FAIR, Linked and Harmonized

Vocabularies Need to be FAIR

Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web

Daniel Garijo[1] [0000-0003-0454-7145] and
María Poveda-Villalón[2] [0000-0003-3587-0367]

[1] Information Sciences Institute, University of Southern California
dgarijo@isi.edu
[2] Ontology Engineering Group, Universidad Politécnica de Madrid
mpoveda@fi.upm.es

Abstract. With the adoption of Semantic Web technologies, an increasing number of vocabularies and ontologies have been developed in different domains, ranging from Biology to Agronomy or Geosciences. However, many of these ontologies are still difficult to find, access and understand by researchers due to a lack of documentation, URI resolving issues, versioning problems, etc. In this chapter we describe guidelines and best practices for creating accessible, understandable and reusable ontologies on the Web, using standard practices and pointing to existing tools and frameworks developed by the Semantic Web community. We illustrate our guidelines with concrete examples, in order to help researchers implement these practices in their future vocabularies.

Keywords: Ontology metadata · Ontology publication · Ontology access · FAIR principles · Linked Data principles.

https://arxiv.org/pdf/2003.13084.pdf

Are We Better Off With Just One Ontology on the Web?

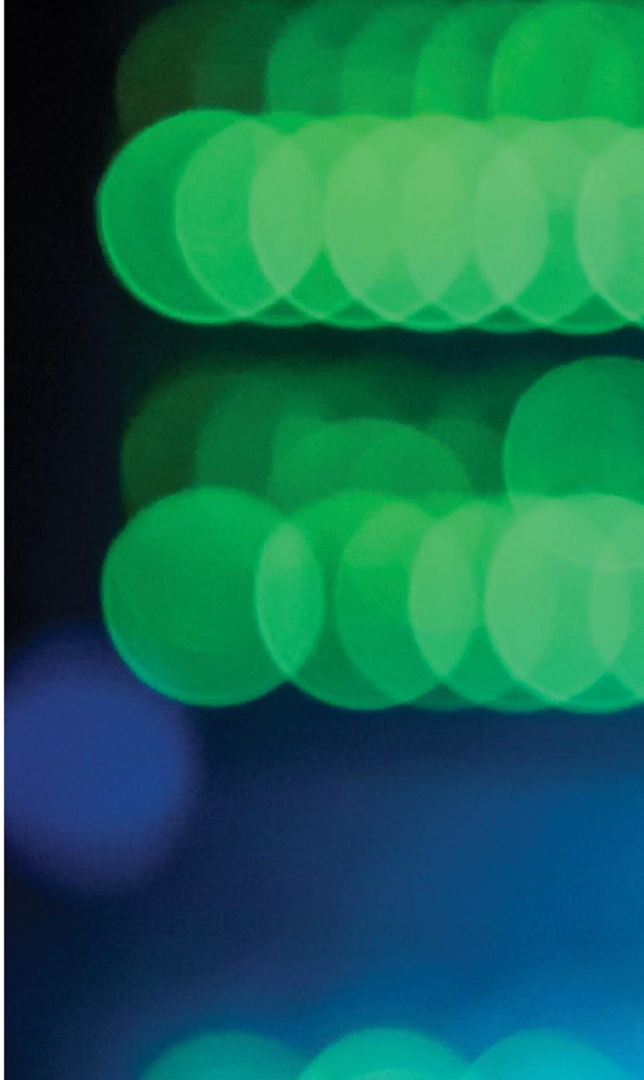
Armin Haller[a,*] Axel Polleres[b]
[a] *Research School of Management & Research School of Computer Science, Australian National University, Canberra, Australia*
E-mail: armin.haller@anu.edu.au
[b] *Institute for Information Business, Vienna University of Economics and Business, Vienna, Austria*
E-mail: axel.polleres@wu.ac.at

Editors: Pascal Hitzler, Wright State University, USA; Krzysztof Janowicz, University of California, Santa Barbara, USA

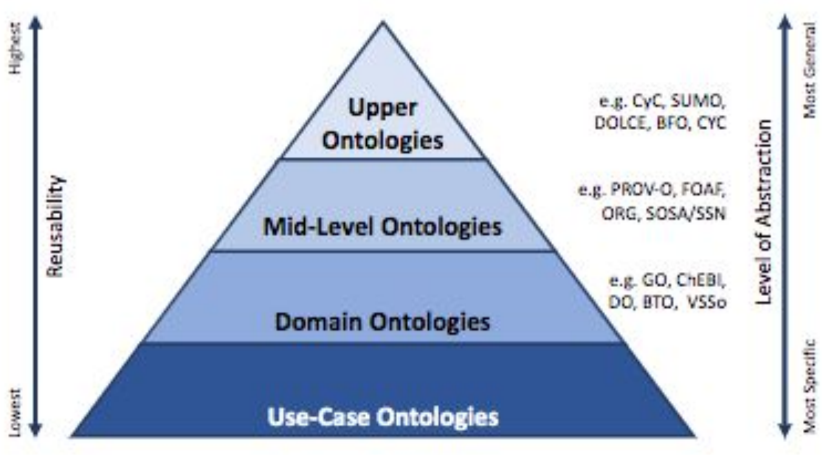
Abstract. Ontologies have been used on the Web to enable semantic interoperability between parties that publish information independently of each other. They have also played an important role in the emergence of Linked Data. However, many ontologies on the Web do not see much use beyond their initial deployment and purpose in one dataset and therefore should rather be called what they are – (local) schemas, which per se do not provide any interoperable semantics. Only few ontologies are truly used as a shared conceptualization between different parties, mostly in controlled environments such as the BioPortal. In this paper, we discuss open challenges relating to true re-use of ontologies on the Web and raise the question: "are we better off with just one ontology on the Web?"

Keywords: Ontology, Knowledge Representation

http://www.semantic-web-journal.net/content/are-we-better-just-one-ontology-web-0

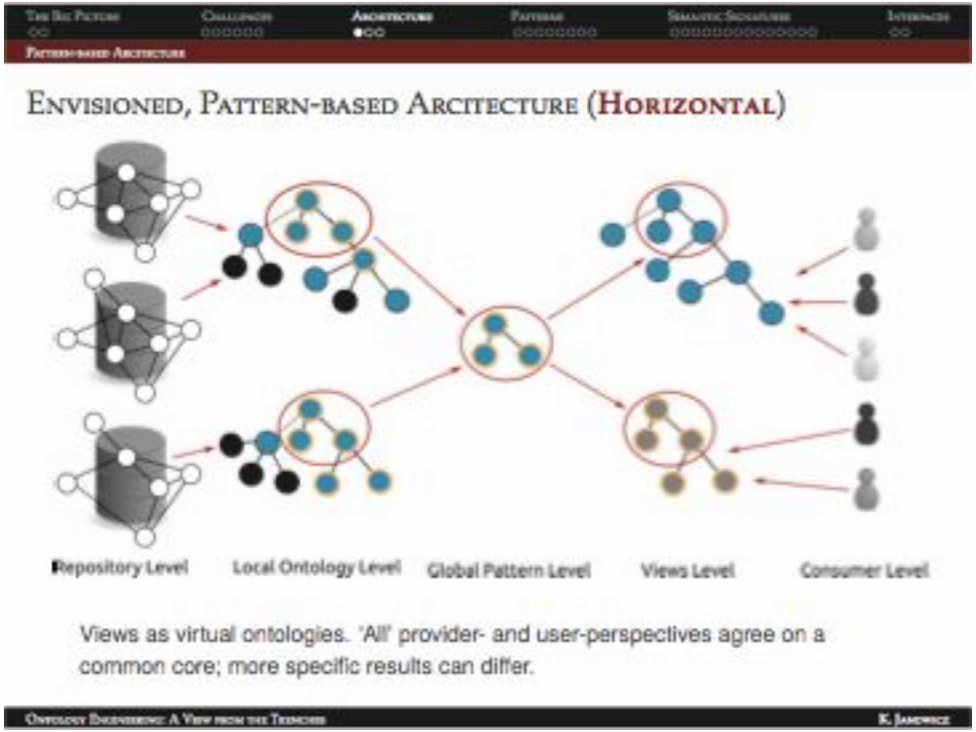


One Ontology to Rule them All?



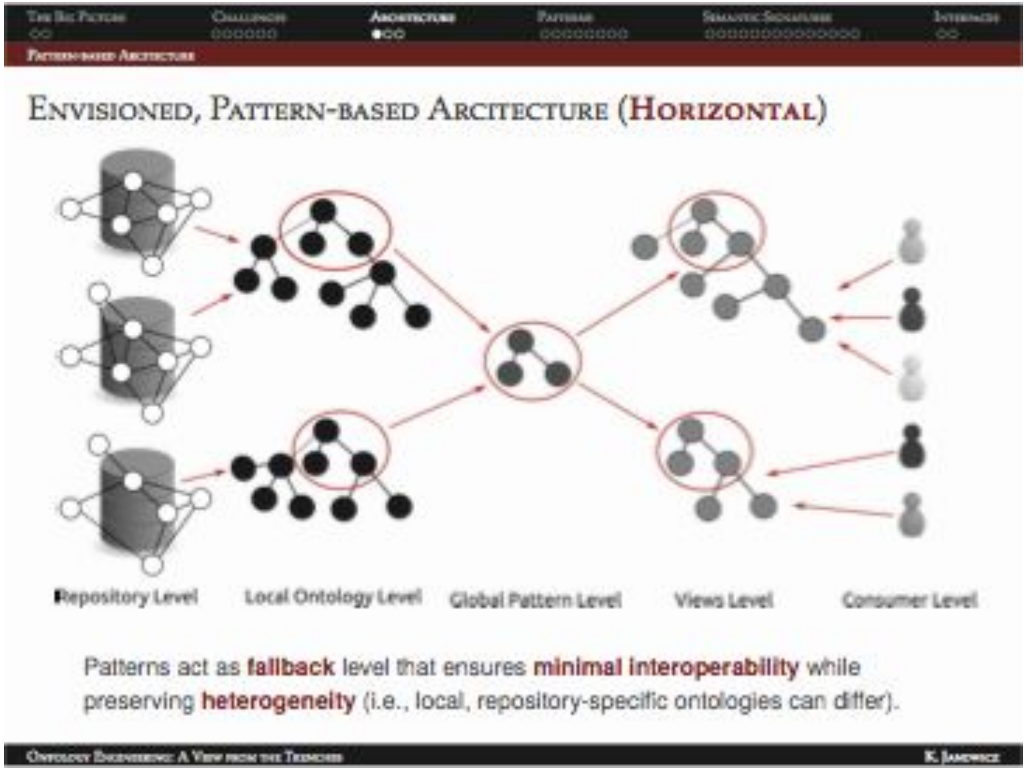
http://www.semantic-web-journal.net/content/are-we-better-just-one-ontology-web-0

Ontology Design Pattern (ODP) Approach - Janowicz WOP 2015



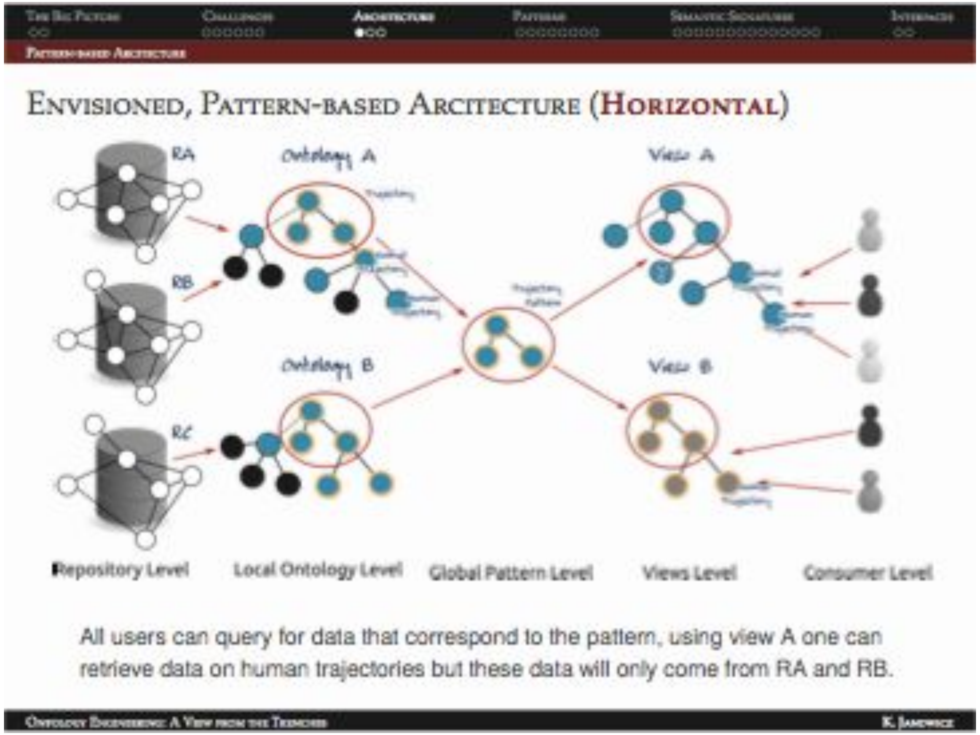
https://www.slideshare.net/kjanowicz/ontology-engineering-a-view-from-the-trenches-wop-2015-keynote

Ontology Design Pattern (ODP) Approach



https://www.slideshare.net/kjanowicz/ontology-engineering-a-view-from-the-trenches-wop-2015-keynote

Ontology Design Pattern (ODP) Semantic Trajectory



https://www.slideshare.net/kjanowicz/ontology-engineering-a-view-from-the-trenches-wop-2015-keynote

A Tutorial on Modular Ontology Modeling with Ontology Design Patterns: The Cooking Recipes Ontology

Pascal Hitzler, Wright State University, USA, pascal@pascal-hitzler.de
Adila Krisnadhi, Universitas Indonesia, krisnadhi@gmail.com

August 2018

We provide a detailed example for modular ontology modeling based on ontology design patterns. It is similar to the Chess Ontology tutorial in [6], which we suggest to read first. We will be less verbose in this tutorial; we provide it because additional examples should be helpful for those interested in adopting the modular ontology modeling methodology – see [6] and the book [2] in which it is contained.

We assume that the reader is familiar with the Web Ontology Language OWL [5, 4].

Before we dive into the actual modeling, let us present the general workflow which we recommend for ontology modeling, and which is the same as in [6]. The steps of this workflow are laid out in Figure 1. We will refer to these steps, and explain them in more detail, as we advance through the tutorial.

https://arxiv.org/abs/1808.08433

Examples of Mid-Level Ontologies for FAIR

Web Standard Middle Ontologies -- Provenance

W3C®

PROV-O: The PROV Ontology

W3C Recommendation 30 April 2013

This version:
http://www.w3.org/TR/2013/REC-prov-o-20130430/
Latest published version:
http://www.w3.org/TR/prov-o/
Implementation report:
http://www.w3.org/TR/2013/NOTE-prov-implementations-20130430/
Previous version:
http://www.w3.org/TR/2013/PR-prov-o-20130312/
Editors:
Timothy Lebo, Rensselaer Polytechnic Institute, USA
Satya Sahoo, Case Western Reserve University, USA
Deborah McGuinness, Rensselaer Polytechnic Institute, USA
Contributors:
(In alphabetical order)
Khalid Belhajjame, University of Manchester, UK
James Cheney, University of Edinburgh, UK
David Corsar, University of Aberdeen, UK
Daniel Garijo, Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
Stian Soiland-Reyes, University of Manchester, UK
Stephan Zednik, Rensselaer Polytechnic Institute, USA
Jun Zhao, University of Oxford, UK

Please refer to the **errata** for this document, which may include some normative corrections.

The English version of this specification is the only normative version. Non-normative translations may also be available.

Web Standard Middle Ontologies -- Time

OGC®
Making location count.

W3C®

Time Ontology in OWL

W3C Candidate Recommendation 26 March 2020

This version:
https://www.w3.org/TR/2020/CR-owl-time-20200326/

Latest published version:
https://www.w3.org/TR/owl-time/

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

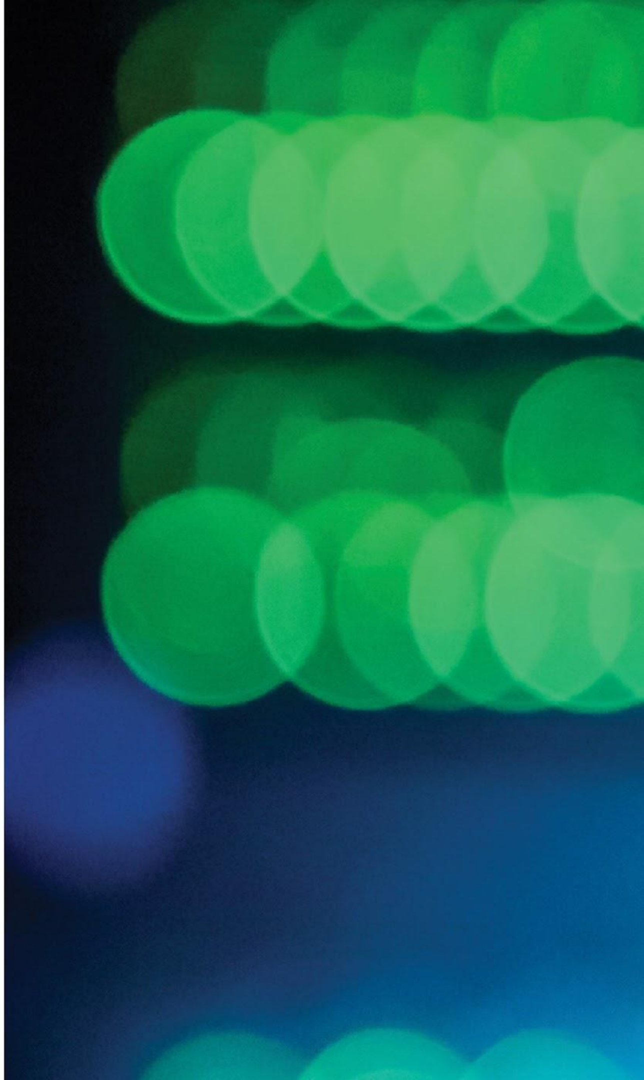
Implementation report:
https://www.w3.org/2015/spatial/wiki/OWL_Time_Ontology_adoption

Previous version:
https://www.w3.org/TR/2017/REC-owl-time-20171019/

Editors:
Simon Cox (CSIRO)
Chris Little (Met Office)

Contributors:
Jerry R. Hobbs
Feng Pan

OGC Document Number:
OGC 16-071r3



Web Standard Middle Ontologies -- GeoSPARQL

Open Geospatial Consortium

Submission Date: 2019-11-21

Approval Date: <yyyy-mm-dd>

Publication Date: <yyyy-mm-dd>

External identifier of this OGC® document: http://www.opengis.net/doc/{doc-type}/{standard}/{m.n}

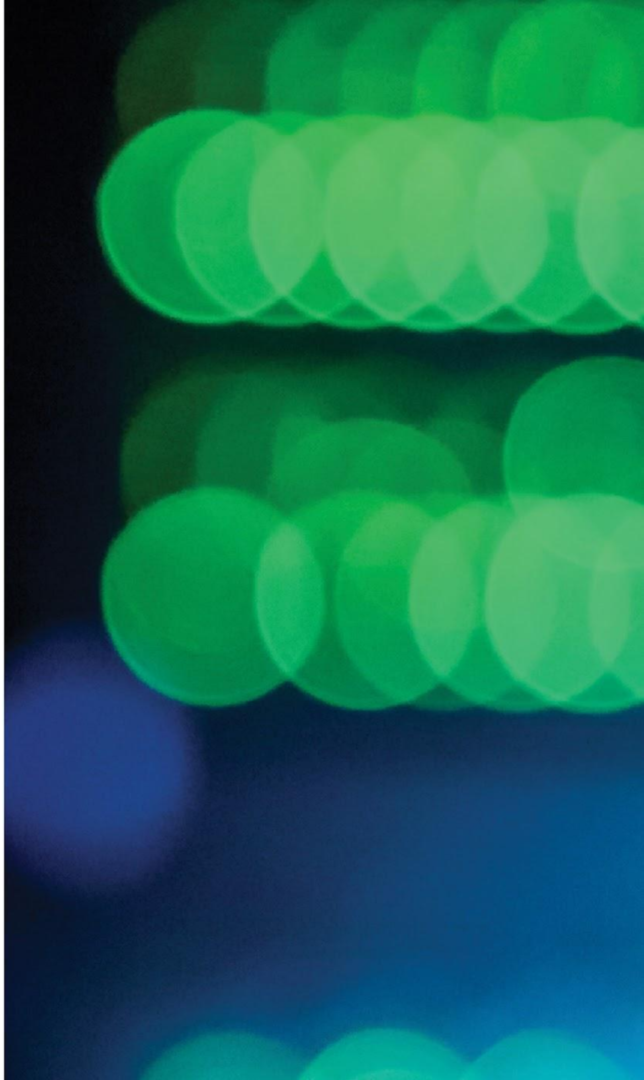
Internal reference number of this OGC® document: 19-078r1

Category: OGC® White Paper

Editor: Joseph Abhayaratna, Linda van den Brink, Nicholas Car, Rob Atkinson, Timo Homburg, Frans Knibbe, Kris McGlinn, Anna Wagner, Mathias Bonduel, Mads Holten Rasmussen, Florian Thiery

OGC Benefits of Representing Spatial Data Using Semantic and Graph Technologies

https://github.com/opengeospatial/geosemantics-dwg
https://github.com/opengeospatial/geosemantics-dwg/blob/master/white_paper/wp.pdf



Web Standard Middle Ontologies -- Observations

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/

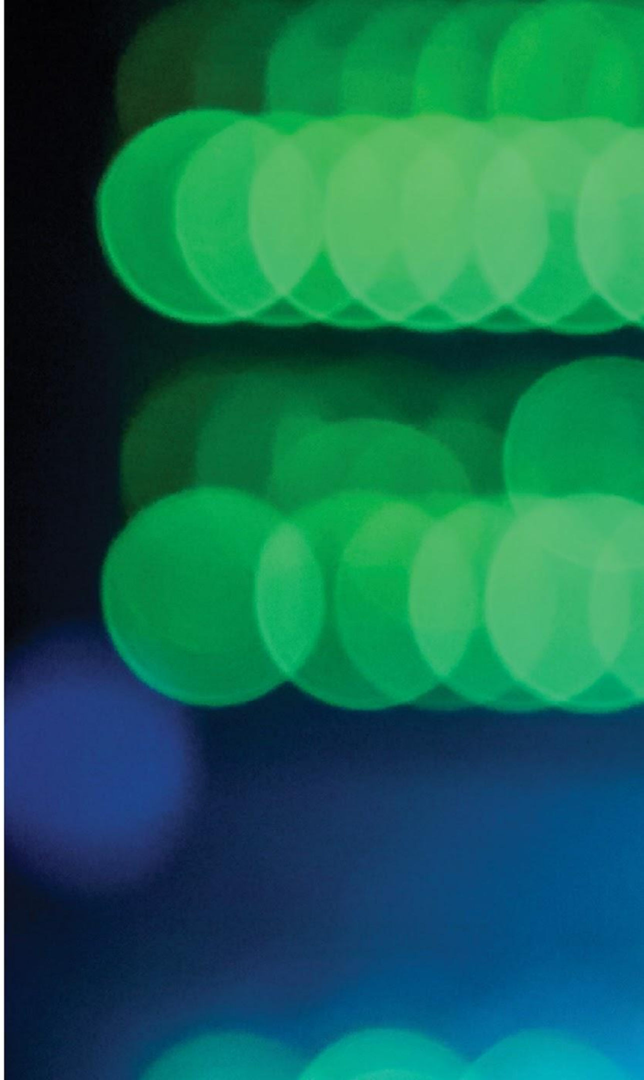
Latest editor's draft:
https://w3c.github.io/sdw/ssn/

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 de Saint-Étienne

Contributors (ordered alphabetically):
Rob Atkinson, Metalinkage
Raúl García-Castro, Universidad Politécnica de Madrid
Joshua Lieberman, Tumbling Walls
Claus Stadler, Universität Leipzig



Web Standard Middle Ontologies -- Datasets

Data Catalog Vocabulary (DCAT) - Version 2

W3C Recommendation 04 February 2020

This version:
https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/

Latest published version:
https://www.w3.org/TR/vocab-dcat-2/

Latest editor's draft:
https://w3c.github.io/dxwg/dcat/

Implementation report:
https://w3c.github.io/dxwg/dcat-implementation-report/

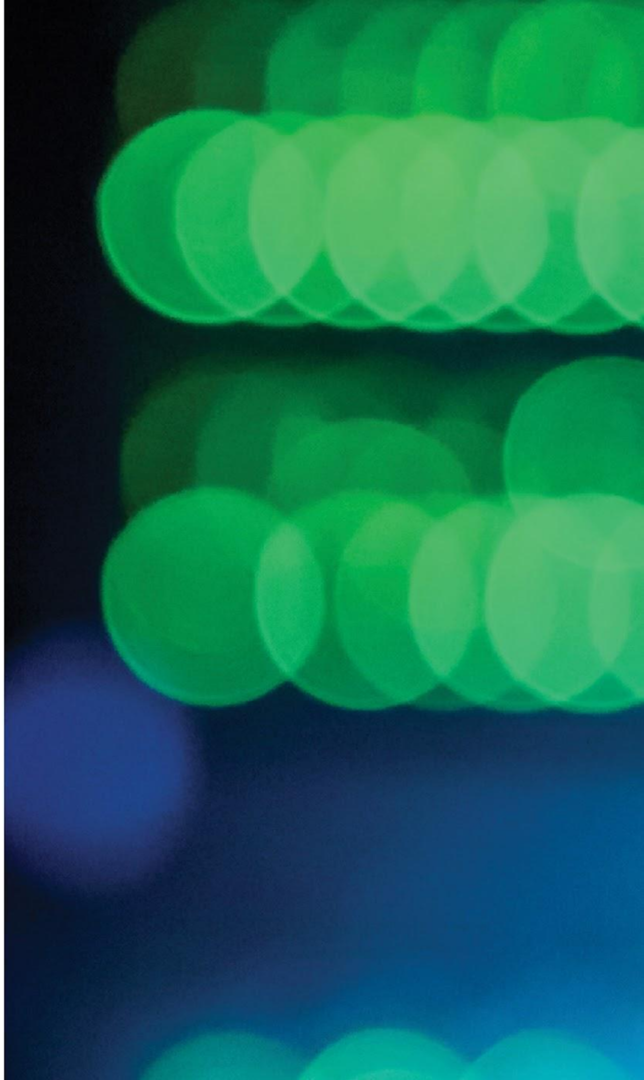
Previous version:
https://www.w3.org/TR/2019/PR-vocab-dcat-2-20191119/

Previous Recommendation:
https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/

Editors:
Riccardo Albertoni (CNR - Consiglio Nazionale delle Ricerche, Italy)
David Browning (Refinitiv)
Simon Cox (CSIRO)
Alejandra Gonzalez Beltran (Scientific Computing Department, Science and Technology Facilities Council, UK) (Previously at the University of Oxford)
Andrea Perego (European Commission, Joint Research Centre)
Peter Winstanley (Scottish Government)

Former editors:
Fadi Maali (DERI)
John Erickson (Tetherless World Constellation (RPI))

Participate:
GitHub w3c/dxwg
File a bug
Commit history
Pull requests



Schema.org and Wikidata as the start of ‘core’ ontologies

COMPUTATIONAL AND SYSTEMS BIOLOGY

Science Forum: Wikidata as a knowledge graph for the life sciences

Andra Waagmeester, Gregory Stupp, Sebastian Burgstaller-Muehlbacher, Benjamin M Good, Malachi Griffith, Obi L Griffith, Kristina Hanspers, Henning Hermjakob, Toby S Hudson see all »

Micelio, Belgium; Department of
States; Center for Integrative Bioi
of Vienna, Austria; McDonnell G
Data Science and Biotechnology, (
Kingdom; School of Chemistry, Th
of Medicine, University of Wa

Abstract

Wikidata is a community-maintained knowledge base that has been assembled from repositories in the fields of genomics, proteomics, genetic variants, pathways, chemical compounds, and diseases, and that adheres to the FAIR principles of findability, accessibility, interoperability and reusability. Here we describe the breadth and depth of the biomedical knowledge contained within Wikidata, and discuss the open-source tools we have built to add information to Wikidata and to synchronize it with source databases. We also demonstrate several use cases for Wikidata, including the crowdsourced curation of biomedical ontologies, phenotype-based diagnosis of disease, and drug repurposing.

Schema.org as a start for Findability

zenodo

Search | Upload | Communities

April 1, 2020

Software | Open Access

Schema.org Publishing Guidelines for the Geosciences v1.1

Adam Shepherd; Matt Jones; Dave Vieglais; Douglas Fils; Stephen Richard; Ruth Duerr; Lewis John McGibbney; Charles Vardeman II

FIXES

- Using `@type` instead of `schema:additionalType` (decision, Issue #74)

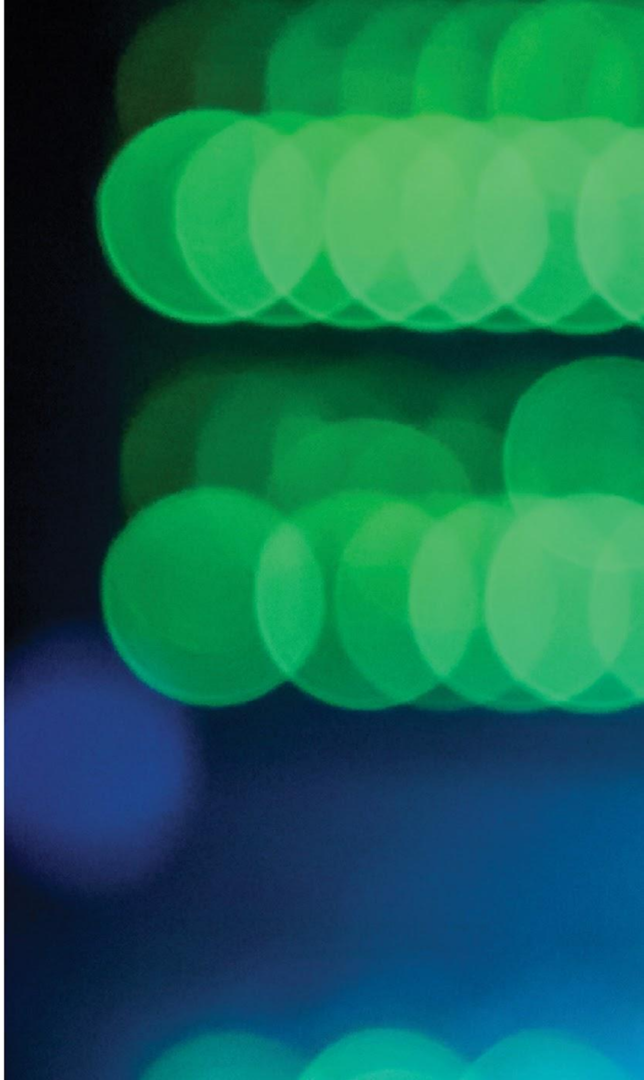
IMPROVEMENTS

- Use `schema:PropertyValue` for describing Persistent Identifiers (decision, Issue #13)
  - Guide: Dataset - Identifier
  - Guide: Data Repository- Identifier
- Use SPDX URLs for Dataset License (decision, Issue #47)
  - Guide: Dataset - License

NEW FEATURES

- Describing Dataset Metadata files (decision, Issue #4)
  - Guide: Dataset - Metadata Files
- Gitflow Release Workflow (decision, Issue #30)
  - CONTRIBUTING.md

https://zenodo.org/record/3736235#.Xy2QypNKhZ4



Harmonization and Linking of Domain Ontologies

CI/CS WORKSHOP ResearchSOC | CI CoE PILOT

Harmonization Efforts: ENVO and Semantic Web for Earth and Environment Terminology (SWEET)

Community Ontology Repository

v3.9.6 Help Contact us

Home Term Search SPARQL Search

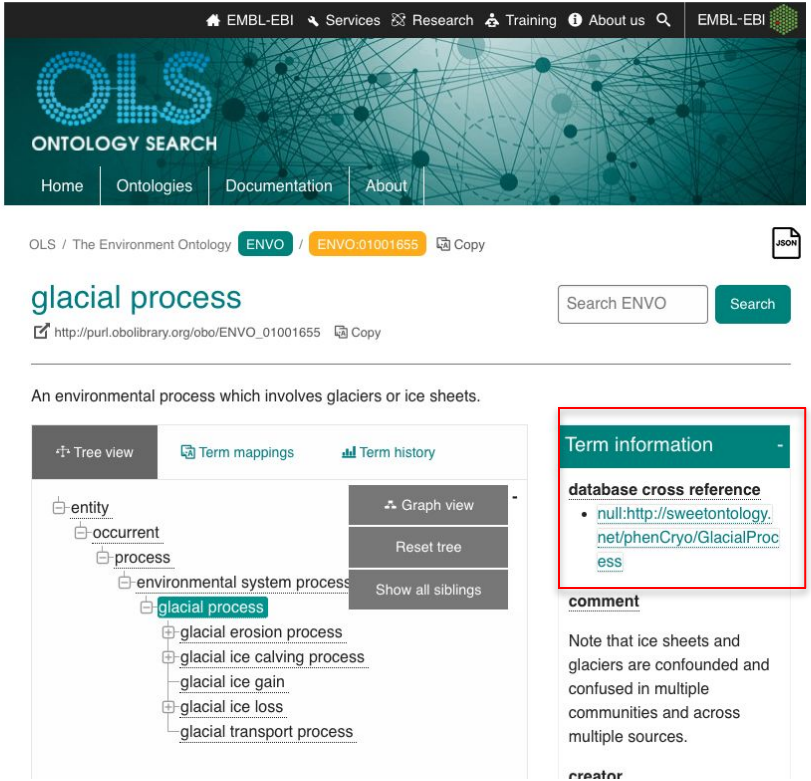
Sign in Create account

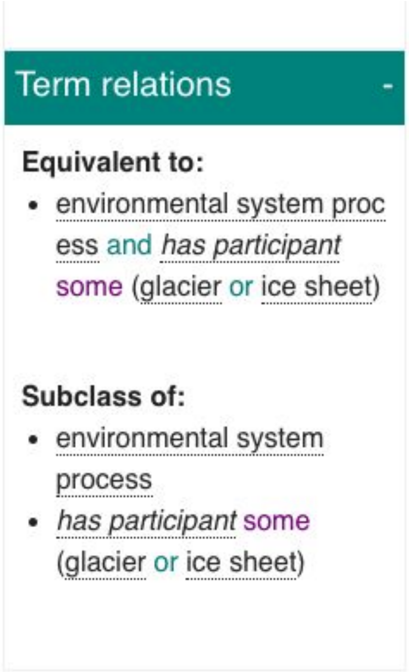
http://sweetontology.net/phenCryo/GlacialProcess View/download as

GlacialProcess

property	value
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#Class
http://www.w3.org/2000/01/rdf-schema#label	"glacial process"@en
http://www.w3.org/2000/01/rdf-schema#subClassOf	http://sweetontology.net/procPhysical/PhysicalProcess
http://www.w3.org/2004/02/skos/core#closeMatch	http://purl.obolibrary.org/obo/ENVO_01001655
http://www.w3.org/2004/02/skos/core#definition	_:bD3C9ACCEx191940

Harmonization Efforts: ENVO and SWEET





Alignments to Mid-Level Ontologies

ESIP Community Ontology Repository

v3.9.6 Help Contact us

Home Term Search SPARQL Search

Sign in Create account

http://sweetontology.net/alignment/ssn Status: **stable** View/download as Versions

SWEET-SSN/SOSA alignment graph

Version: 20191211T002942 - Author: ESIP Federation Owner: esip

Metadata details

General | Usage/License/Permissions | Original source | Other

rdfs **label:**	SWEET-SSN/SOSA alignment graph
rdfs **comment:**	A preliminary axiomatization of the alignment of SWEET with the W3C SSN/SOSA ontologies
owl **versionInfo:**	3.5.0

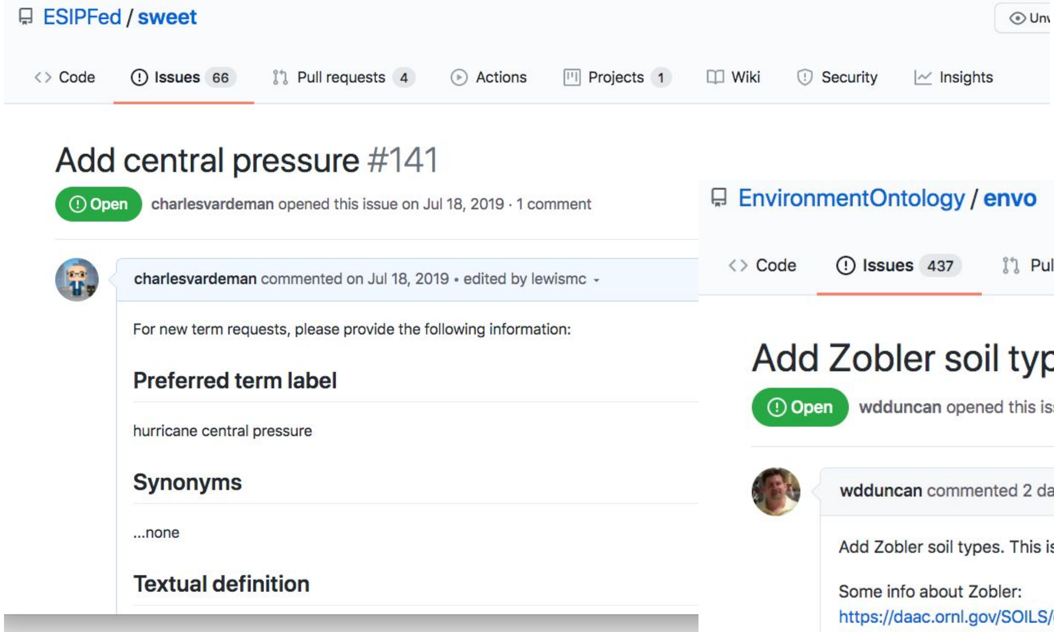
Data

Display contents using:

Triple table pyLODE

Subject	Predicate	Object
http://sweetontology.net/humanResearch/Observation	http://www.w3.org/2000/01/rdf-schema#subClassOf	http://www.w3.org/ns/sosa/Observation
http://sweetontology.net/humanResearch/Result	http://www.w3.org/2000/01/rdf-schema#subClassOf	http://www.w3.org/ns/sosa/Result
http://sweetontology.net/humanResearch/Variable	http://www.w3.org/2000/01/rdf-schema#subClassOf	http://www.w3.org/ns/ssn/Property
http://sweetontology.net/humanResearch/res_Observation-subClas...	http://www.w3.org/1999/02/22-rdf-syntax-ns#predicate	http://www.w3.org/2000/01/rdf-schema#subClassOf
http://sweetontology.net/humanResearch/res_Observation-subClas...	http://www.w3.org/1999/02/22-rdf-syntax-ns#subject	http://sweetontology.net/humanResearch/Observation
http://sweetontology.net/humanResearch/res_Observation-subClas...	http://www.w3.org/1999/02/22-rdf-syntax-ns#object	http://www.w3.org/ns/sosa/Observation

GitHub Driven Development



ESIPFed / sweet

Code | Issues 66 | Pull requests 4 | Actions | Projects 1 | Wiki | Security | Insights

Add central pressure #141

Open charlesvardeman opened this issue on Jul 18, 2019 · 1 comment

charlesvardeman commented on Jul 18, 2019 • edited by lewismc

For new term requests, please provide the following information:

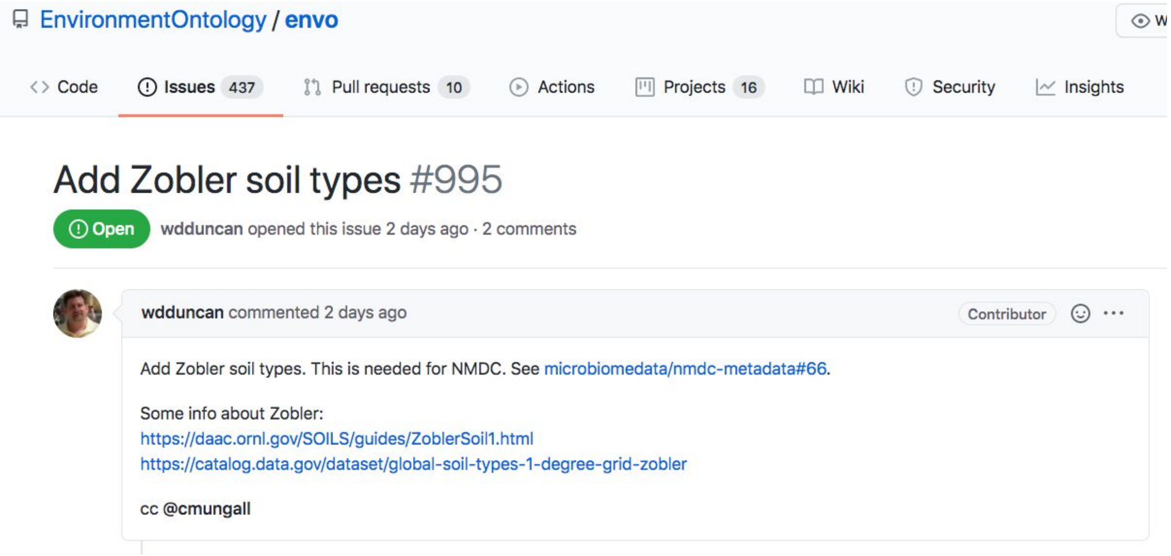
Preferred term label

hurricane central pressure

Synonyms

...none

Textual definition



EnvironmentOntology / envo

Code | Issues 437 | Pull requests 10 | Actions | Projects 16 | Wiki | Security | Insights

Add Zobler soil types #995

Open wdduncan opened this issue 2 days ago · 2 comments

wdduncan commented 2 days ago Contributor

Add Zobler soil types. This is needed for NMDC. See microbiomedata/nmdc-metadata#66.

Some info about Zobler:
https://daac.ornl.gov/SOILS/guides/ZoblerSoil1.html
https://catalog.data.gov/dataset/global-soil-types-1-degree-grid-zobler

cc @cmungall

Lessons from CI-CoE Pilot

How do we move toward FAIR Principles

- Work to connect Large Facilities to scientific communities through organizations like ESIP, RDA, CODATA, Go-FAIR to advance FAIR adoption because FAIR is a community process

- Leverage Community Software Building Blocks such as RDFlib, RDFJS, tools built by Zazuko GmbH like trifid, Inrupt Solid

- **Knowledge engineering** to create better ontologies through harmonization, adoption of FAIR practices, formalization of vocabularies in ontologies

Thank you!

https://cicoe-pilot.org/

https://wiki.esipfed.org/Schema.org_Cluster

https://wiki.esipfed.org/SemanticHarmonization

RDA - i-Adopt https://bit.ly/33D6IEL

cvardema@nd.edu