Douglas Fils (COL) Adam Shepherd (WHOI)

"I just want a box…" has been a request for some time…

Structured data on the web is in widespread use in the commercial sectors of the web… it powers job search, movie and restaurant listings and more..

Tooling and experience for the web are widespread

Scalable & resilient

...and there is schema.org/Dataset

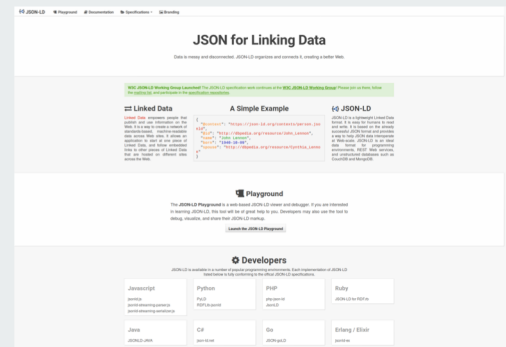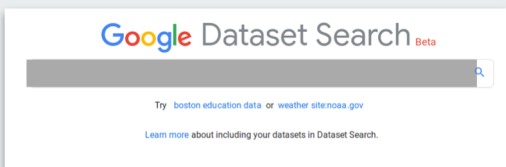Google and some providers have started down this path

#UseTheWeb

Using common web architecture patterns accelerates adoption and improves development & sustainability

Turns out… same for semantics

Google Dataset Search Beta

Try boston education data or weather site:noaa.gov

Learn more about including your datasets in Dataset Search.

schema.org

JSON for Linking Data

http://json-ld.org

Project 418

Describing
Using shared and connected vocabularies

Publishing
Using web architecture patterns
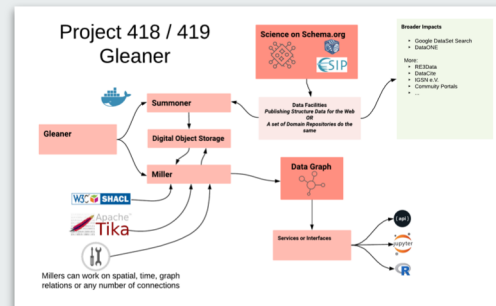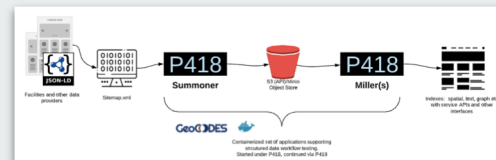
Indexing
Using tools on the web architecture patterns

Serving
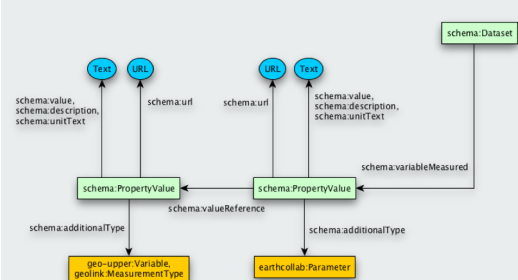To evaluate index use

Iterating multiple times through all the stages of the data life cycle helps evolve the recommendations in vocabularies and tools

earthcube.org

- Focus on schema.org/Dataset

- Connect to external vocabularies to address domain specific needs

- Address FAIR further with variableMeasured and measurementTechnique

- Improve spatial representation (OGC)

- Improve temporal representation (W3C OWL Time)

Vocabulary

"Data In Context"

Schema.org is a foundation to extend and provides a common voice on which to connect our our domain dialects

science-on-schema.org
geoschema.org
geoshapes.org
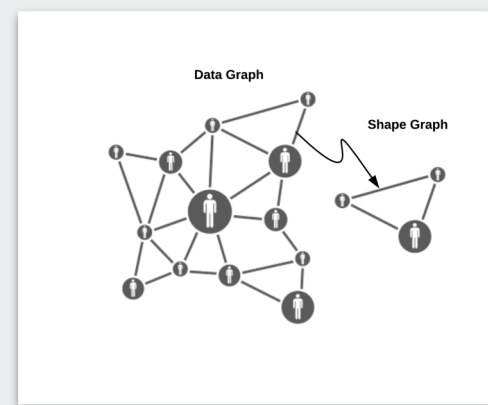
Shape graphs provide validation and constraint checking

P419 focusing on temporal and spatial aspects

Validation via W3C SHACL
https://www.w3.org/TR/shacl/

W3C DCAT (with schema.org alignment)
https://www.w3.org/TR/vocab-dcat-2/

Project 418 @ GitHub

Gleaner
Harvest, process, index

Tangram
Validation as a service

Fence
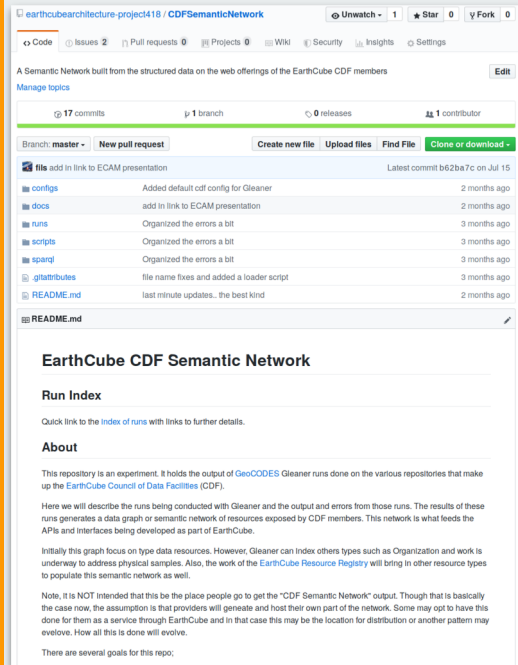Assessments and options

Open Source

Tools, Vocabulary and Data

Open Source is shared infrastructure (which is also a sustainability model)

gleaner.io

"Open" Covers Data Too!

CDF Semantic Network
Gleaner outputs is a semantic network. For the P418 this is the *EarthCube Council of Data Facilities (CDF) Semantic Network* published on GitHub via Large File Support.

Principles over Project

EarthCube:
Council of Data Facilities

ESIP:
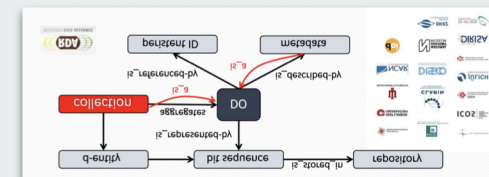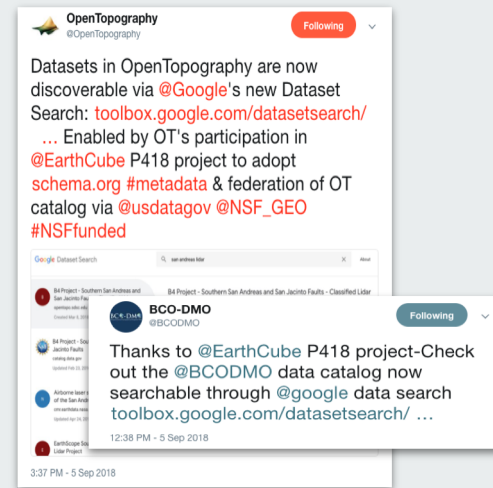*Semantic Committee and Schema.org Cluster*

RDA:
Schema.org Working Group

See more at:
- Polar Data Forum,
- RDA Plenary 14,
- AGU Fall Meeting,
- ESIP Winter Meeting
- EGU

Community

Community provides sustainability, helps to accelerate growth and evolves patterns

OpenTopography @OpenTopography · Following

Datasets in OpenTopography are now discoverable via @Google's new Dataset Search: toolbox.google.com/datasetsearch/ ... Enabled by OT's participation in @EarthCube P418 project to adopt schema.org #metadata & federation of OT catalog via @usdatagov @NSF_GEO #NSFfunded

3:37 PM · 5 Sep 2018

BCO-DMO @BCODMO · Following

Thanks to @EarthCube P418 project-Check out the @BCODMO data catalog now searchable through @google data search toolbox.google.com/datasetsearch/ ...

12:38 PM · 5 Sep 2018

XSEDE
Extreme Science and Engineering
Discovery Environment