Project 418 / 419 Gleaner

Structured Data on the Web for GeoSciences

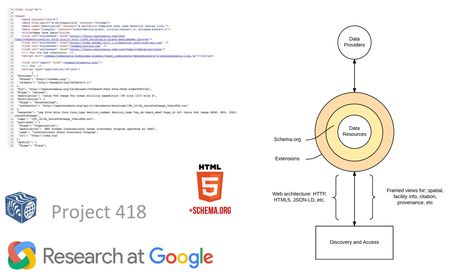
Douglas Fils
Data Manager
Ocean Leadership
dfils@oceanleadership.org

ORCID ID

Adam Shepherd
Technical Director
BCO-DMO
Woods Hole Oceanographic Institution
ashepherd@whoi.edu

Eric Lingerfelt
Technical Officer
EarthCube Science Support Office (ESSO)
eric2@ucar.edu

schema.org/Dataset



Google Research has built an index and interface to scientific data following schema.org/Dataset

EarthCube Project 418/9 has a working prototype using the schema.org/Dataset

EarthCube Project 419 is working on time and service descriptions to support more facility data

Semantic Web Methodology

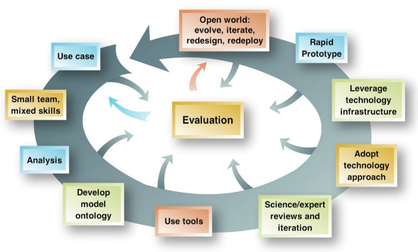
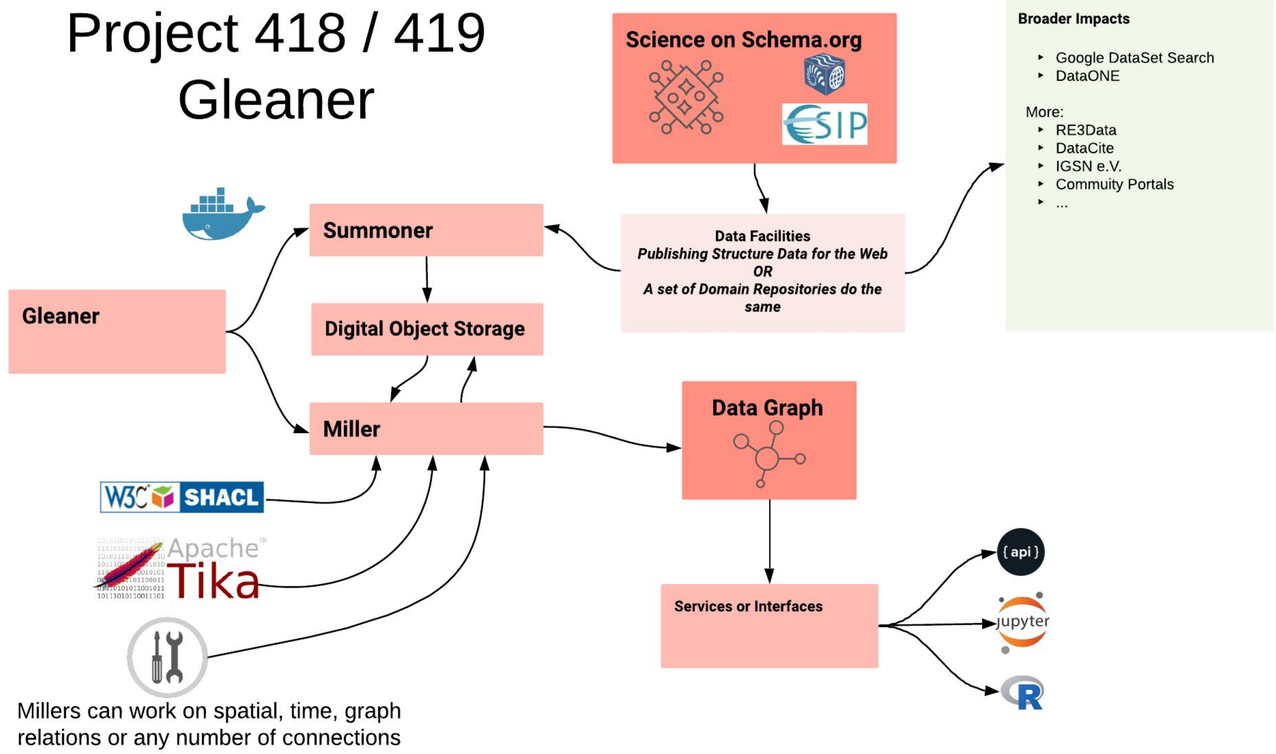


Image credit: Peter Fox
https://tw.rpi.edu/web/doc/TWC_SemanticWebMethodology

Facilities

BALTO
CHORDS
CDF RWG
MagIC
BCO-DMO
Biological & Chemical Oceanography Data Management Office
HYDROSHARE
UNAVCO
IEDA
INTERDISCIPLINARY EARTH DATA ALLIANCE
OpenTopography
High-Resolution Topography Data and Tools
LinkedEarth
IRIS
CSDCO
Continental Scientific Drilling Coordination Office
IODES Resolution Science Operator
Neotoma
R2R
Open Core

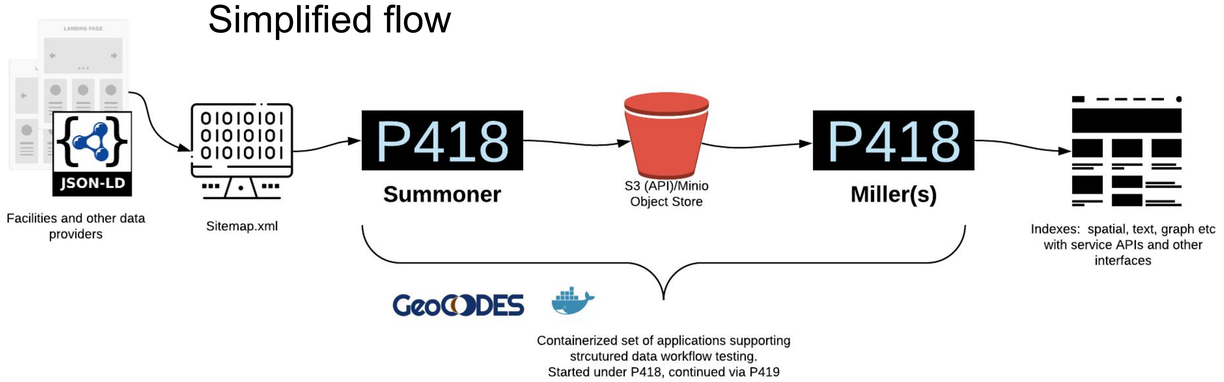
Project 418 / 419 Gleaner



Millers can work on spatial, time, graph relations or any number of connections

Gleaner is a tool for harvesting and processing structured data on the web. It can process data graphs all the way to a searchable index and support building various indexes and routing data through data processing workflows.

Simplified flow



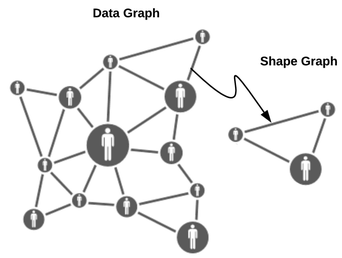
Containerized set of applications supporting strcutured data workflow testing.
Started under P418, continued via P419

Validation

Community Defined Shapes

Shapes can test
- number of occurrences
- term structure
- terms in lists
- location in graph
- much more!

We can use to to ensure support for FAIR elements, citation, sample relations, controlled voc use, etc!

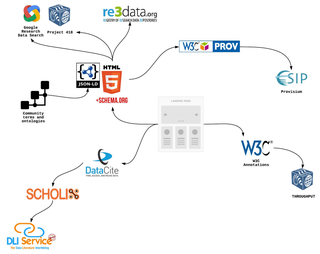


W3C SHACL

Connection Possibilities

Leverage web architecture to expose holdings in as many venues and patterns as possible.

The pattern allows 3rd parties to add value to the exposed data and structured data while enabling citation and provenance.



Acknowledgments

Project 418 was funded by EarthCube Science Support Office (ESSO) though NSF award number: 1623751

Further Information

All code is at Github:
https://github.com/earthcubearchitecture-project418
and https://geodex.org