"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

Leveraging Structured Data on the Web to address FAIR data Principles

Douglas Fils (COL) Adam Shepherd (WHOI)

#UseTheWeb

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

Turns out... same for semantics



Try boston education data or weather site:noaa.gov

Learn more about including your datasets in Dataset Search

schema.org

JSDN-LD @ Paypoint @ Incomentation & Specifications - Skithanding

JSON for Linking Data

WSC JSON LS Working Group Laure The matters lat, and participate in the g	ubed The JODHLD specification work continues at the WDC JSDHLD Working , andihadian manadarias	Genuge' Please join us Parro, ballow
≓ Linked Data	A Simple Example	(4) JSON-LD
United Cala empowers people that publish and use information on the	Calescentry, Material/Jan-16.org/contexts/person.jac	JBON LD is a lightweight United Date Remail, 8 to easy for humans to read

Playground

 Store David Dav

	tated below to fully confi	sensing to the official JSON-LD specification	mertation of JSDN LD rst.
Javascript	Python	PHP	Ruby
porid)a porid stearing parar(s porid stearing setial(pr)s	PyLD RDPLIx-panid	php-jour-lef JournED	30410 to 101/p
Java	C4	64	Erlans / Elixir

http://json-ld.org



Describing

Using shared and connected vocabularies



Publishing

Using web architecture patterns



Indexing

Using tools on the web architecture patterns



Serving

To evaluate index use

Project 418

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

earthcubearchitecture-pro	/ork	og Unwatch - 1 ★ Star 0 ÿ Fork 0				
O Code ⊕ Issues 2 m	Pull requests 0 III Projects (🖲 🖂 Wiki 🕡 S	ecurity Ins	ights 🖧 S	Settings	
A Semantic Network built from t Manage topics	the structured data on the web off	erings of the EarthC	ube CDF memb	ers		Edit
@ 17 commits	1 branch	0 0	releases		1 contributor	
Branch: master - New put	ll request	Create new file	Upload files	Find File	Clone or downlo	ad -
🚳 fils add in link to ECAM pr	resentation			Latest co	mmit b62ba7c on Ji	ul 15
in configs	Added default cdf config for G	lleaner			2 months	ago
in docs	add In link to ECAM presenta	tion			2 months	ago
in runs	Organized the errors a bit				3 months	ago
iii scripts	Organized the errors a bit				3 months	ago
📷 sparql	Organized the errors a bit				3 months	ago
.gitattributes	file name fixes and added a l	pader script			3 months	ago
README.md	last minute updates the best	kind			2 months	ago
I README.md						1

EarthCube CDF Semantic Network

Run Index

Quick link to the index of runs with links to further details

About

This repository is an experiment. It holds the output of GeoCODES Gleaner runs done on the various repositories that make up the EarthCube Council of Data Facilities (CDF).

Here we will describe the runs being conducted with Gleaner and the output and errors from those runs. The results of these runs generates a data graph or semantic network of resources exposed by CDF members. This network is what leeds the APIs and interfaces being developed as part of EarthCube.

Initially this graph focus on type data resources. However, Gleaner can Index others types such as Organization and work is underway to address physical samples. Also, the work of the EarthCube Resource Registry will bring in other resource types to populate this semantic network as well.

Note, It is NOT intended that this be the place people go to get the "CDD" Semantic Network" output. Though that is basically the case now, the easumption is that provident will generate and host their own part of the network. Some may got to have this done for them as a service through EarthCube and in that case this may be the location for distribution or another pattern may evolverv. How all this is done will evolve.

There are several goals for this repo;

"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





Extreme Science and Engineering Discovery Environment