

Global Change Information System (GCIS)

Semantic Web Update

U.S. Global Change Research Program

Justin Goldstein/USGCRP-UCAR
ESIP Summer Meeting - July 14, 2015

GCIS Semantic Web Update Outline

- Semantic Web Underpinning
- Ontology 1.2
- Ontology Updates Underway
- Lexicon Updates

Semantic Web Underpinning

- RESTful API at data.globalchange.gov
- URLs correspond to ontology URIs
- Primary storage: RDBMS (PostgreSQL)
- Representation is serialized (for JSON) or used in templates (for Turtle)
- Turtle representation is exported into a triple store (Virtuoso) which provides a SPARQL endpoint.

Ontology 1.2

- Completed in late 2013
- Has been used since release of NCA3 in May 2014
- Small differences between the ontology and implementation
- NCA focused
- First-cut for some definitions and relationships
- Link: <http://data.globalchange.gov/gcis.owl>

Ontology Improvements Underway

- Better relationships (e.g. more accurate domain and characterization)
- New characterization of activities
- Better characterization of observation entities (platforms, instruments, etc.)
- Updated definitions - better compatibility with other ontologies (primarily Dublin Core and Prov)
- Extensions to handle CDI “theme” concept (e.g. health, transportation)
- Expected to release this fall
- DRAFT Update Link (work in progress) :

<https://github.com/USGCRP/gcis-ontology/blob/master/gcis.ttl>

Example of Updated Ontology

```
@prefix dcterms: <http://pur1.org/dc/terms/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix dbpedia_owl: <http://dbpedia.org/ontology/> .
@prefix gcis: <http://data.globalchange.gov/activity/dcterms:identifier "02c53cf7-nca
@prefix meth: <http://sweet.jpl.nasa.gov/ontology/2008/09/01/meth.owl#> .
@prefix rdf: <http://www.w3.org/1999/rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://data.globalchange.gov/activity/dcterms:identifier "02c53cf7-nca
dcterms:description "Values of a
dcterms:extent [ rdf:value "6 hours" ] ;
dcterms:outputDataFile [ dbpedia_owl:filename "tx.ave.proc" ] ;
dcterms:computingEnvironment gcis:computingEnvironmentsUsed "Mac OS X (darwin x86_64 m64)"^^xsd:string;

prov:qualifiedAssociation [
  a prov:Association ;
  prov:agent [
    a prov:SoftwareAgent, gcis:Software ;
    rdfs:label "IDL (version 8.0); Microsoft Excel for Mac 2011 Version 11.5.2"^^xsd:string;
  ] ;
  prov:hadPlan [
    a prov:Plan, meth:Methodology ;
    rdf:value "Seasonal average values of temperature and precipitation from
on a scatter plot for each year, with data points for select years subsequent
that year."^^xsd:string;
  ] ;
] ;

a prov:Activity .

<http://data.globalchange.gov/activity/dcterms:identifier "02c53cf7-nca
dcterms:description "Values of a
dcterms:extent [ rdf:value "6 hours" ] ;
dcterms:outputDataFile [ dbpedia_owl:filename "tx.ave.proc" ] ;
dcterms:computingEnvironment gcis:computingEnvironmentsUsed "Mac OS X (darwin x86_64 m64)"^^xsd:string;

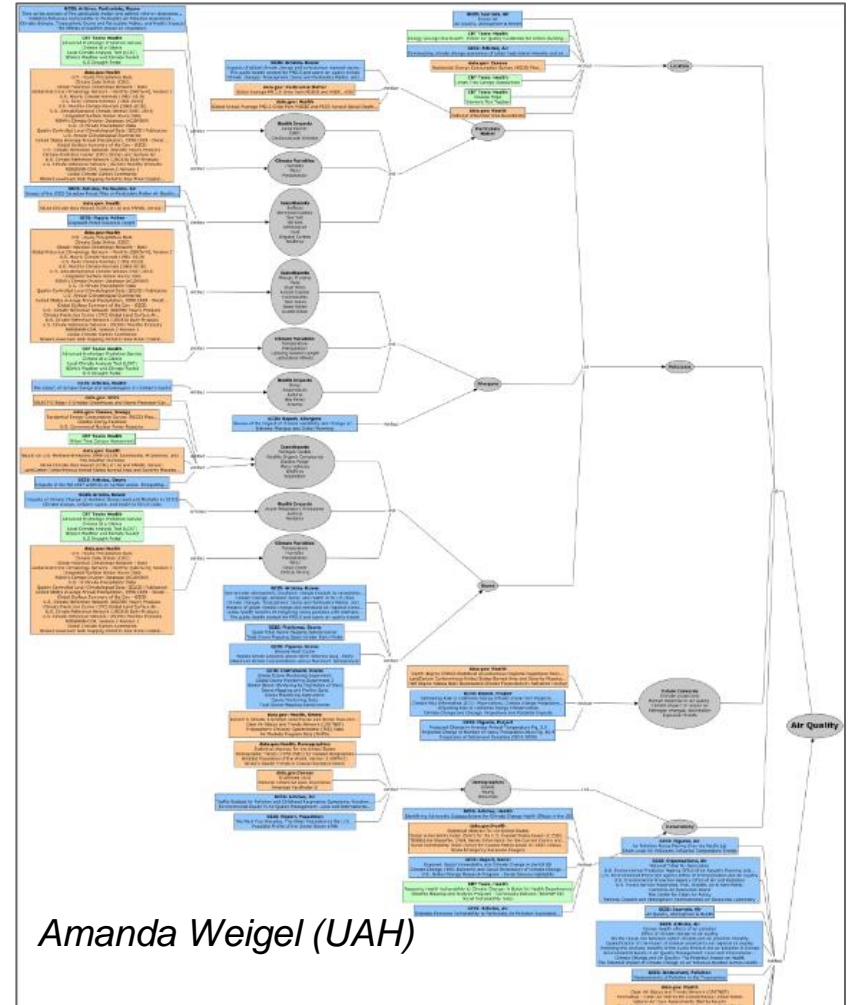
prov:qualifiedAssociation [
  a prov:Association ;
  prov:agent [
    a prov:SoftwareAgent, gcis:Software ;
    rdfs:label "IDL (version 8.0); Microsoft Excel for Mac 2011 Version 11.5.2"^^xsd:string;
  ] ;
  prov:hadPlan [
    a prov:Plan, meth:Methodology ;
    rdf:value "Seasonal average values of temperature and precipitation from
on a scatter plot for each year, with data points for select years subsequent
that year."^^xsd:string;
  ] ;
] ;

a prov:Activity .
```

- Extended use of qualified associated
- Created a new property for computing environment

Example of CDI Theme: Health Lexicon (preliminary)

- Using draft Climate and Health Assessment and NCA3 Health Chapter to develop and test vocabulary
- Developing use cases to inform the design



From LDOW 2015 (Florence, IT)

Normalizing Resource Identifiers using Lexicons in the Global Change Information System

Linking Earth Science Identifiers, Concepts, and Communities

Brian Duggan¹⁴
bduggan@usgcrp.gov

Curt Tilmes²
curt.tilmes@nasa.gov

Steven Aulenbach¹⁴
saulenbach@usgcrp.gov

Robert E. Wolfe¹²
rewolfe@usgcrp.gov

Justin C. Goldstein¹⁴
jgoldstein@usgcrp.gov

Gerald Manion³
geraldjohn.m.manion@jpl.nasa.gov

¹US Global Change Research Program
1717 Pennsylvania Ave NW
Washington, DC 20006

²NASA Goddard Space Flight Center
8800 Greenbelt Rd
Greenbelt, MD 20771

³NASA/Jet Propulsion Laboratory
4800 Oak Grove Dr
Pasadena, CA 91011

⁴University Corporation for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307

ABSTRACT

Earth Science informatics involves collaboration between multiple groups of people with diverse specializations and goals, often using variations in terminology to refer to common resources. The uniformity of the resource identifiers often does not cross organizational boundaries. Because of this, permanent, widely used, unambiguous identifiers for resources are elusive. We examine real world cases of changing and inconsistent identifiers which inherently work against persistence and uniformity. We also present a solution which mediates factors in these situations; namely the creation of lexicons: mappings of sets of terms to URIs which are curated within the Global Change Information System (GCIS).

We discuss aspects of the GCIS which facilitate the use of lexicons: an information model which disambiguates resources, a RESTful API which provides metadata through content-negotiation, and a strategy for long term curation of URIs, including mechanisms for handling changes to URIs and variations in terms used by different communities while providing persistent URIs and preserving relationships between resources.

We provide working definitions of *terms*, *contexts*, and *lex-*

icons of them to identifiers used colloquially within various earth science communities of practice.

Keywords

Linked Data, URI, Co-reference

1. INTRODUCTION

1.1 Background

The U.S. Global Change Research Program (USGCRP) was established in 1989 by Presidential Initiative and mandated by the U.S. Congress in the Global Change Research Act (GCRA) of 1990 to "assist the Nation and the world to understand, assess, predict, and respond to human-induced and natural processes of global change."^[1] The USGCRP has recently sponsored the creation of the Global Change Information System (GCIS) to better coordinate and integrate the use of federal information products on changes in the global environment and the implications of those changes for society.

In May, 2014, the USGCRP released the Third National Climate Assessment (NCA3). This 800 page document, authored by 300 people, each of which are affiliated with mul-

In
forthcoming
paper

Terms, Contexts, Lexicons

Lexicon	Context	Term	GCID (*)
podaac	Source	JASON-1	/platform/jason-1
ceos	MissionId	286	/platform/jason-1
gcmd	prefLabel	JASON-1	/platform/jason-1
echo	ShortName	JASON-1	/platform/jason-1
podaac	Sensor	POSEIDON-2	/instrument/poseidon-2
ceos	InstrumentId	182	/instrument/poseidon-2

(*) under <http://data.globalchange.gov>

See also: <http://data.globalchange.gov/lexicon>

15 / 28

http://events.linkeddata.org/ldow2015/papers/ldow2015_paper_12.pdf

Challenges / Questions

- How far can “lexicons” be used? Where do distinctions of “mission,” “platform,” and “instrument” break down? Multiple platforms e.g. GRACE?
- Reliance on definitions from outside sources (e.g. CEOS, PO.DAAC...)
- How does a health dataset differ from other physical Earth science datasets? What are the preservation, provenance, and documentation implications?
- How does one integrate with controlled vocabularies and taxonomies? What are the advantages/limitations to using lexicons for this practice?
- Extent of utilization of ISO 190115
- Using GCIS to query (SPARQL) other sources to derive relationships.

Forthcoming work - Semantic perspective

- Further updates to GCIS Ontology to reflect ongoing USGCRP efforts
- Structured provenance tracking for future efforts
- Further integration with lexicons, especially more work with dbpedia
- Forthcoming integration with PROV-ES