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
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

- 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

Amanda Weigel (UAH)
Normalizing Resource Identifiers using Lexicons in the Global Change Information System

Linking Earth Science Identifiers, Concepts, and Communities

Brian Duggan1,2 bduuggan@usgs.gov
Curt Tilmes3 curt.tilmes@nasa.gov
Steven Aulenbach1,2 saulenbach@usgs.gov
Robert E. Wolle1,2 rewolle@usgs.gov
Justin C. Goldstein1,2 jgoldstein@usgs.gov
Gerald Manion1 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

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 resources 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 lexicons 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 (USCRP) 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 USCRP 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 USCRP 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

<table>
<thead>
<tr>
<th>Lexicon</th>
<th>Context</th>
<th>Term</th>
<th>GCID (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>podac</td>
<td>Source</td>
<td>JASON-1</td>
<td>/platform/jason-1</td>
</tr>
<tr>
<td>ceos</td>
<td>MissionId</td>
<td>286</td>
<td>/platform/jason-1</td>
</tr>
<tr>
<td>gcnd</td>
<td>prefLabel</td>
<td>JASON-1</td>
<td>/platform/jason-1</td>
</tr>
<tr>
<td>echo</td>
<td>ShortName</td>
<td>JASON-1</td>
<td>/platform/jason-1</td>
</tr>
<tr>
<td>podac</td>
<td>Sensor</td>
<td>POSEIDON-2</td>
<td>/instrument/poseidon-2</td>
</tr>
<tr>
<td>ceos</td>
<td>InstrumentId</td>
<td>182</td>
<td>/instrument/poseidon-2</td>
</tr>
</tbody>
</table>

(*) under http://data.globalchange.gov
See also: http://data.globalchange.gov/lexicon

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