

Adaptable Information Models in the Global Change Information System

Brian Duggan¹², Andrew Buddenberg³,
Steve Aulenbach¹², Robert Wolfe¹⁴, Justin Goldstein¹²

¹US Global Change Research Program

²University Corporation for Atmospheric Research

³National Oceanic and Atmospheric Administration

⁴National Aeronautics and Space Administration

December 16, 2014

<http://data.globalchange.gov>

<http://github.com/USGCRP/gcis>



Outline

1. Introduction and Functionality

- Overview
- Producing the Third National Climate Assessment
- Supporting the NCA3 website
- Provenance
- Source of Information
- Queries

2. Information Model

- Relational
- Semantic
- Example

3. System Architecture

- Diagram
- Content Changes
- Schema Changes
- Ontology Changes

4. Conclusion, Ongoing Work, Future Plans

Outline

1. Introduction and Functionality

- Overview
- Producing the Third National Climate Assessment
- Supporting the NCA3 website
- Provenance
- Source of Information
- Queries

2. Information Model

- Relational
- Semantic
- Example

3. System Architecture

- Diagram
- Content Changes
- Schema Changes
- Ontology Changes

4. Conclusion, Ongoing Work, Future Plans

Overview

The US Global Change Research Program (USGCRP) has established 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.

Overview

The GCIS provides identifiers for resources (URIs), representations and relationships between those resources.

A RESTful API provides JSON and HTML representations.

A triple store provides a SPARQL endpoint.

URIs in the triple store are endpoints in the API.

An ontology defines the resources and the relationships between URIs.

Producing the Third National Climate Assessment

In May, 2014, the US Global Change Research Program released the 2014 National Climate Assessment.

Production of this 829 page report and its web site involved collaboration between over 300 [authors](#), numerous editors, analysts, graphics producers, scientists, data scientists, software developers, and web teams.

The content included 161 [findings](#), 284 [figures](#), 3,395 bibliographic [references](#) (journal [articles](#), [books](#), [reports](#)).

The GCIS facilitated the assembly of the report by providing common [identifiers](#) for resources and concepts, providing a common web interface for entering and viewing information, as well as an API for adding and removing information using a variety of formats.

Producing the Third National Climate Assessment



<http://data.globalchange.gov/report/nca3>
<http://data.globalchange.gov/report/nca3.html>
<http://data.globalchange.gov/report/nca3.json>
<http://data.globalchange.gov/report/nca3.ttl>

Supporting the NCA3 website

A website, <http://nca2014.globalchange.gov>, was released concurrently with the report. The site received over 200,000 visits in the first two days after the launch and continues to receive frequent main stream media attention.

GCIS serves as the backend; the website sends client-side requests to <http://data.globalchange.gov> and receives JSON responses which it uses to populate elements of some pages dynamically.

Supporting the NCA3 website

The screenshot shows a web browser displaying the NCA3 website. The address bar shows the URL `nca2014.globalchange.gov/r...`. The main content area features a figure titled "Figure: Global Temperature and Carbon Dioxide" with a caption explaining the data. Below the figure is a "Download (116 KB)" button. The browser's developer tools are open at the bottom, showing the Network tab with a list of requests. An orange line highlights the path from the browser's address bar, through the figure's metadata, to the network requests, indicating the data flow.

Observed Change I

nca2014.globalchange.gov/r...

Figure: Global Temperature and Carbon Dioxide

Caption: Global annual average temperature (as measured over both land and oceans) increased by more than 1.5°F (0.8°C) since 1880 (through 2012). Red bars show temperatures above the long-term average, and blue bars indicate temperatures below the long-term average. The black line shows atmospheric carbon dioxide (CO₂) concentration in parts per million (ppm). While there is a clear long-term global warming trend, some years do not show a temperature increase relative to the previous year, and some years show greater changes than others. These year-to-year fluctuations in temperature are due to natural processes, such as the effects of El Niños, La Niñas, and volcanic eruptions. (Figure source: updated from Karl et al. 2009¹).

Download (116 KB)

Figure Images

Keywords

Source Kenneth Kunkel, Cooperative Institute for Climate and Satellites - NC

Date Created 06 Nov 2013

Method	File	Domain	Type	Size	0 ms	18.24 s
384 GET	e251f59b-177e-4ba6...	nca2014.globalchange...	html	8.84 KB		
384 GET	.thumb-a83588...	data.globalchange...	png	46.33 KB		
200 GET	26fc56f4-b4e8-45b...	data.globalchange...	json	3.63 KB		
200 GET	nca3-ncdc-gst-and...	data.globalchange...	json	2.13 KB		
200 GET	nca3-epi-co2-r2013...	data.globalchange...	json	1.92 KB		
200 GET	26fc56f4-b4e8-425b...	data.globalchange...	json	3.63 KB		
384 GET	cs_global_temp...	data.globalchange...	png	154.52 KB		
384 GET	bg-div_top.png	nca2014.globalch...	png	6.90 KB		

81 requests, 2,245.10 KB, 21.27

nca2014.globalchange.gov

data.globalchange.gov

Provenance

GCIS identifiers can be used to trace the provenance of resources.

A [figure](#) may be derived from a [journal article](#), which is derived from a [dataset](#), which is derived from a NASA standard product, which is derived from an [instrument](#), which is on a [platform](#).

The [activity](#) of producing a figure uses a [model run](#) which involves a [scenario](#) applied to a [model](#) run by an [organization](#).

Provenance



platform

instrument

dataset

dataset

journal article

figure

Source of Information

- Committee on Earth Observing Satellites (CEOS)
(platforms, instruments)
- NASA Physical Oceanography DAAC (datasets)
- PCMDI (model runs)
- crossref.org (journal articles)
- <http://github.com/usgcrp/gcis-sync>

Queries

Structured information allows for querying.

- Find reports with figures derived from a dataset generated by an instrument on a specific platform.
- Show figures associated with data generated by instruments on platforms flown by NOAA.

The structure of queries is determined by the information model.

Outline

1. Introduction and Functionality

- Overview
- Producing the Third National Climate Assessment
- Supporting the NCA3 website
- Provenance
- Source of Information
- Queries

2. Information Model

- Relational
- Semantic
- Example

3. System Architecture

- Diagram
- Content Changes
- Schema Changes
- Ontology Changes

4. Conclusion, Ongoing Work, Future Plans

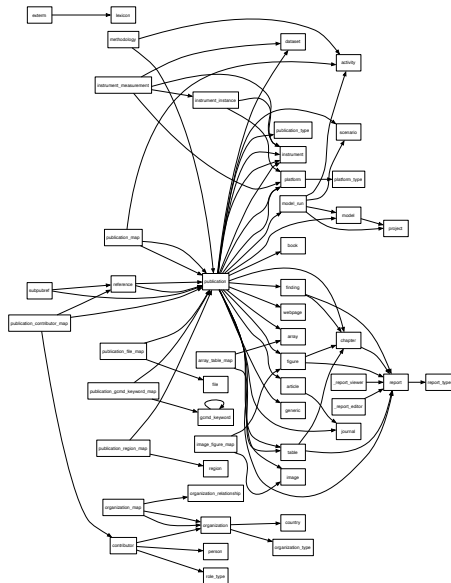
Relational

Canonical representation : PostgreSQL database.

<http://data.globalchange.gov/resources>

- One-many, many-one, many-many relationships
- Referential integrity
- String type checking
- Column constraints
- Cascading updates and deletes
- Well-known optimization techniques
- Wide-spread adoption
- Extendable (hstores in PostgreSQL)
- Closed world assumption

Relational



Semantic

- Relationships are first class objects
- Concepts are formally defined in an ontology
- Formal definitions help remove ambiguities
- Interoperability with other systems
- Open world assumption

Example

<http://bit.ly/gcis-dbpedia>

```
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX gcis: <http://data.globalchange.gov/gcis.owl#>
PREFIX cito: <http://purl.org/spar/cito/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dbprop: <http://dbpedia.org/property/>
PREFIX dbpo: <http://dbpedia.org/ontology/>

SELECT DISTINCT ?dbpjjournal ?gcisjournal ?issn
FROM <http://data.globalchange.gov>
WHERE {
    SERVICE <http://data.globalchange.gov/sparql> {
        ?gcisjournal a bibo:Journal .
        ?gcisjournal bibo:issn ?issn .
        ?gcisjournal dcterms:hasPart ?gcisarticle .
        ?gcisarticle a bibo:Article .
        ?gcisarticle dcterms:isPartOf ?gcisjournal .
        ?gcisarticle cito:isCitedBy <http://data.globalchange.gov/report/nca3> .
    }
    SERVICE <http://dbpedia.org/sparql> 1
        ?dbpjjournal dbprop:frequency "Monthly" @en .
        ?dbpjjournal dbpo:issn ?issnd .
    }
    FILTER(?issnd = ?issn)
}
```

Find monthly journals that have had an article cited by the NCA3 report.

Outline

1. Introduction and Functionality

- Overview
- Producing the Third National Climate Assessment
- Supporting the NCA3 website
- Provenance
- Source of Information
- Queries

2. Information Model

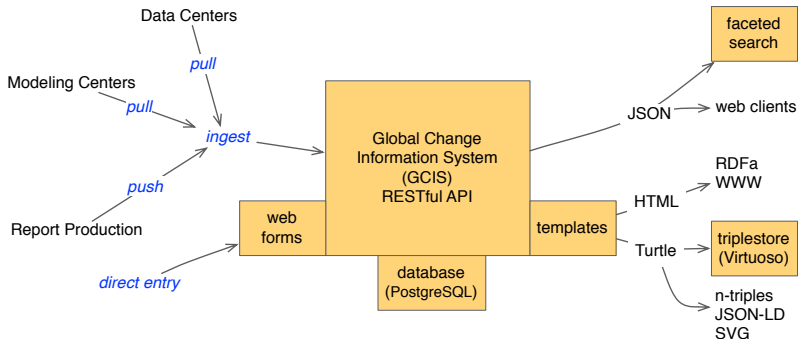
- Relational
- Semantic
- Example

3. System Architecture

- Diagram
- Content Changes
- Schema Changes
- Ontology Changes

4. Conclusion, Ongoing Work, Future Plans

Diagram



Content Changes

- Ingest happens through the API
- External identifiers are mapped to internal GCIS identifiers with lexicons
- Natural identifiers are used when possible (DOIs, ISBNs, ORCIDs)

Providing common identifiers allows external entities to map to each other.

- PODAAC : AQUA = /platform/aqua
- CEOS : 206 = /platform/aqua
- Therefore PODAAC : AQUA = CEOS 206

New sources of information may require changes to the information model.

Schema Changes

Changes to the schema propagate to the JSON API. JSON key names match the column names, and nested JSON objects correspond to relationships.

1. Write a test for new REST functionality.
2. Run the tests. Do the tests pass?
3. Yes? Done.
4. No? Write a schema patch.
5. Goto step 2.

The tests remain part of the test suite, which is run continuously.

Ontology Changes

Change to the triple are handled by turtle templates.

1. Write a test with a SPARQL query that should succeed.
2. Run the tests. Do they pass?
3. Yes? Done.
4. No? Modify the turtle templates.
5. Go to step 2.

The tests remain part of the test suite, which is run continuously.

Ontology Changes

Sample turtle template :

```
@prefix gcis: <http://data.globalchange.gov/gcis.owl#> .  
@prefix dcterms: <http://purl.org/dc/terms/> .  
  
<<%= article->uri %>> a gcis:Article;  
<<%= article->uri %>> dcterms:isPartOf  
    <<%= article->journal->uri %>>;
```


Outline

1. Introduction and Functionality

- Overview
- Producing the Third National Climate Assessment
- Supporting the NCA3 website
- Provenance
- Source of Information
- Queries

2. Information Model

- Relational
- Semantic
- Example

3. System Architecture

- Diagram
- Content Changes
- Schema Changes
- Ontology Changes

4. Conclusion, Ongoing Work, Future Plans

Current work involves extending the data model for models and model runs, *in situ* station data, spatial metadata (PostGIS), agency-wide publication systems, authors of journal articles, and connecting to disparate sources of information using lexicons, APIs and semantic queries.

Thank You

<http://data.globalchange.gov/about> (mailing list)

<http://www.globalchange.gov>

<http://github.com/usgcrp/gcis>

Also thanks : Curt Tilmes, Hook Hua, Brian Wilson, Gerald Manion, Angel Li, April Sides, Sarah Champion, Bradley Akamine, Amanda McQueen, Peter Fox, Marshall Ma and the Tetherless World Institute at RPI