Adaptable Information Models in the Global Change Information System

Brian Duggan[12], Andrew Buddenberg[3],
Steve Aulenbach[12], Robert Wolfe[14], Justin Goldstein[12]

[1]US Global Change Research Program
[2]University Coporation for Atmospheric Research
[3]National Oceanic and Atmospheric Administration
[4]National 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
   - Queries
2. Information Model
   - Relational
   - Semantic
   - Example
3. System Architecture
   - Diagram
   - 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
   - Queries
2. Information Model
   - Relational
   - Semantic
   - Example
3. System Architecture
   - Diagram
   - 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 a RESTful API for retrieving global change information. The GCIS also provides a triple store. URLs in the triple store are resolvable using the API. URIs in the triple store are described by the GCIS ontology.

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, 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

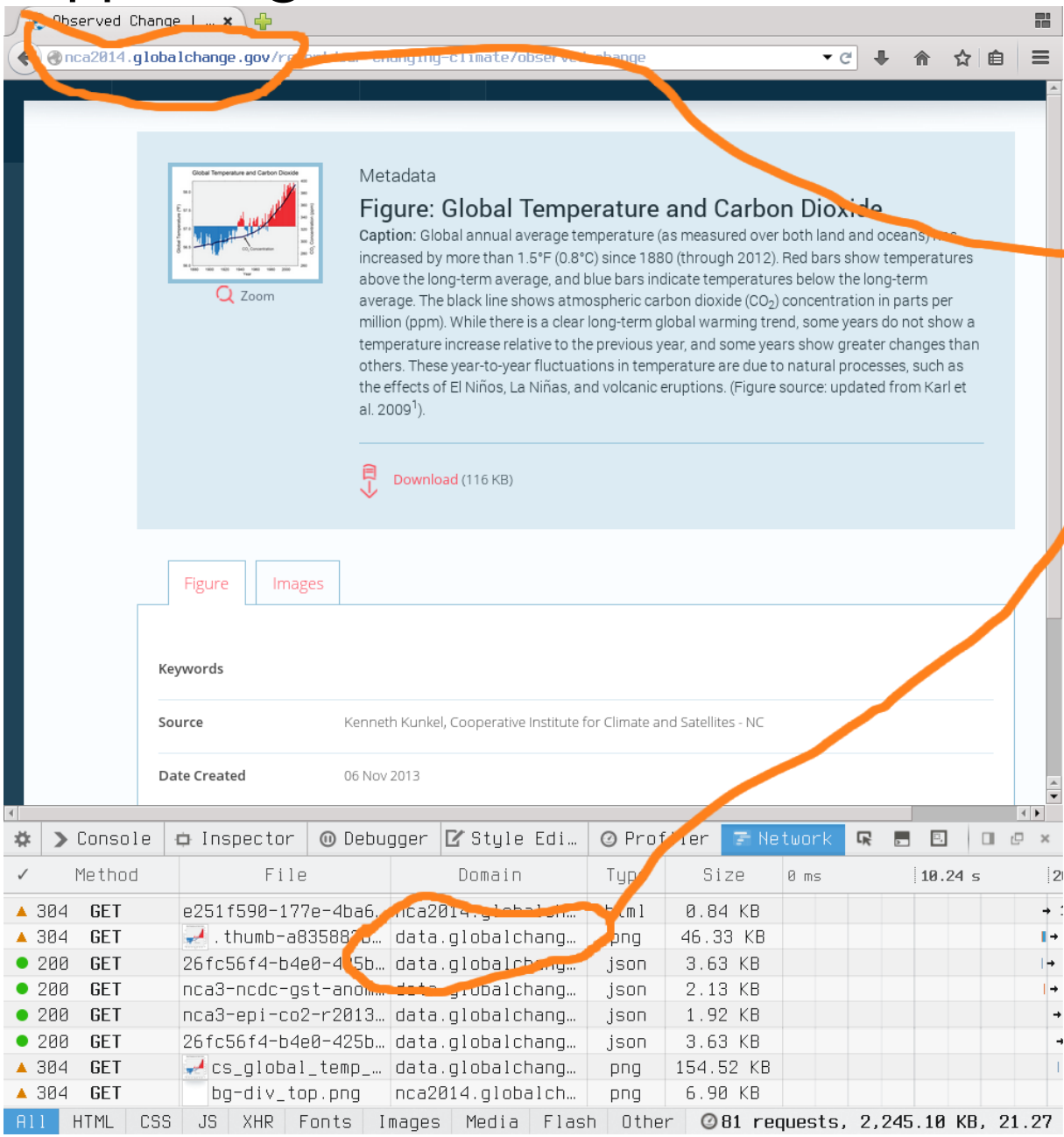
http://data.globalchange.gov/report/nca3
http://data.globalchange.gov/report/nca3.html
http://data.globalchange.gov/report/nca3.json
http://data.globalchange.gov/report/nca3.ttl

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 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



Provenance

The identifiers within GCIS can be used to trace the provenance of figures, findings, and other 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.

Provenance


Platform : jason-1
Instrument : poseidon-2
Advanced Microwave Radiometer
dataset : nasa-podaac-merged_tp_j1_ostm_ost_cycles_v2
dataset : nasa-podaac-integrated-multi-mission-ocean-altimeter-data-for-climate-research
article : 10.1080/01490419.2010.491031
Estimating Mean Sea Level Change from the TOPEX and Jason Altimeter Missions
Figure : past-and-projected-changes-in-global-sea-level-rise
Past and Projected Changes in Global Sea Level Rise
Figure 2.26

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
   - Queries
2. Information Model
   - Relational
   - Semantic
   - Example
3. System Architecture
   - Diagram
   - Schema Changes
   - Ontology Changes
4. Conclusion, Ongoing Work, Future Plans

Relational

Canonical representation : PostgreSQL database.

- 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.

PostgreSQL hstores allow key-value storage.

Closed world assumption.

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

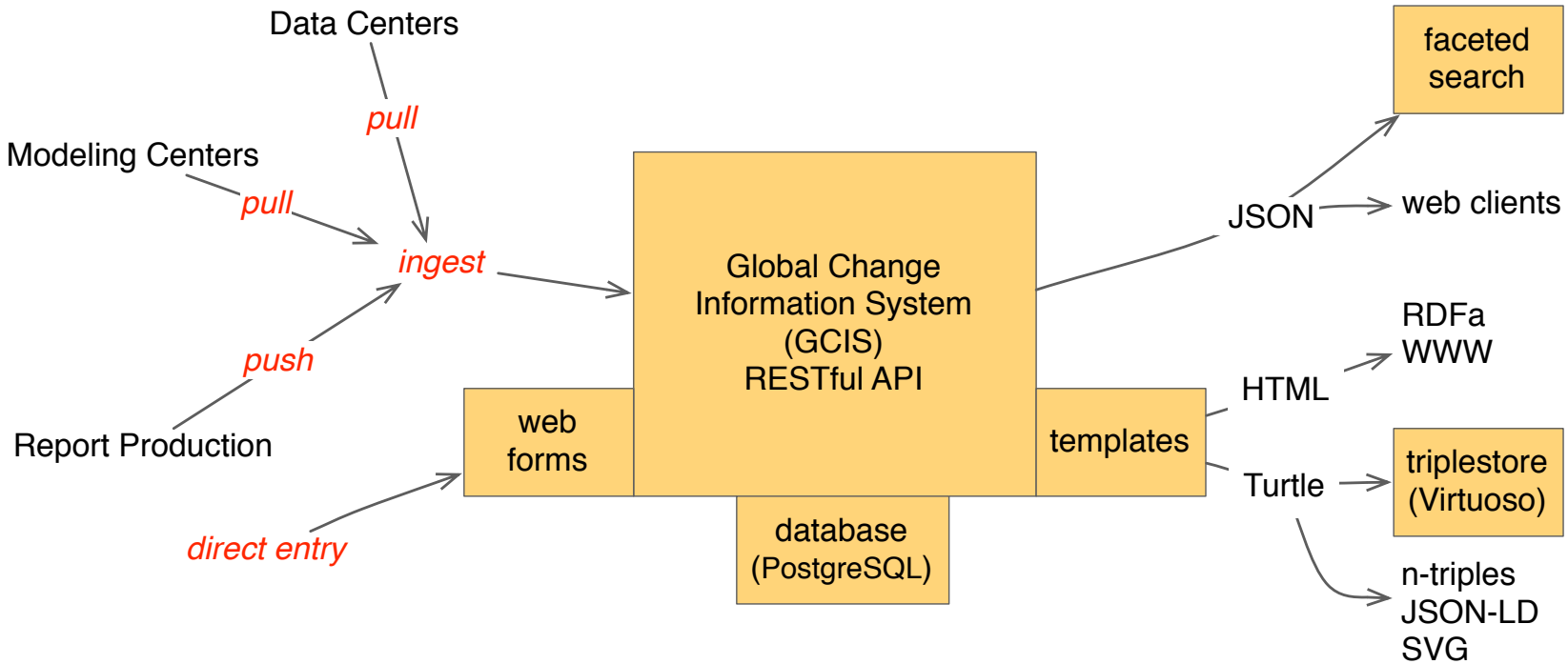
```
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX gcis: <http://data.globalchange.gov/gcis.owl#>
PREFIX cito: <http://purl.org/spar/cito/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dbprop: <http://dbpedia.org/property/>
PREFIX dbpo: <http://dbpedia.org/ontology/>

SELECT  DISTINCT ?dbpjournal ?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
    ?dbpjournal dbprop:frequency "Monthly" @en .
    ?dbpjournal dbpo:issn ?issnd .
   }
   FILTER(?issnd = ?issn)
}
```

Find monthly journals which 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
   - Queries
2. Information Model
   - Relational
   - Semantic
   - Example
3. System Architecture
   - Diagram
   - Schema Changes
   - Ontology Changes
4. Conclusion, Ongoing Work, Future Plans

Diagram


Data Centers
pull
Modeling Centers
pull
ingest
push
Report Production
direct entry
Global Change
Information System
(GCIS)
RESTful API
web
forms
database
(PostgreSQL)
templates
JSON
faceted
search
web clients
HTML
RDFa
WWW
Turtle
triplestore
(Virtuoso)
n-triples
JSON-LD
SVG

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 they test 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 :

```
<%= 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
   - Queries
2. Information Model
   - Relational
   - Semantic
   - Example
3. System Architecture
   - Diagram
   - Schema Changes
   - Ontology Changes
4. Conclusion, Ongoing Work, Future Plans

Current work involves extending the data model to include models, in situ observations, datasets from more DAACs, and identifying lexicons and APIs for GCIS resources.

Thank you

http://github.com/usgcrp/gcis
http://data.globalchange.gov
http://www.globalchange.gov