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 Corporation 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`

United States Global Change Research Program

Outline

1. Introduction and Functionality
    - About
    - The Third National Climate Assessment
    - Identifiers
    - Sources of Information
    - RESTful API
    - Semantic API
2. Information Model
    - Relational
    - Semantic
3. System Architecture
    - Diagram
    - Lexicons
    - Relational Model Changes
    - Semantic Model Changes
4. Conclusion, Ongoing Work, Future Plans

Outline

1. Introduction and Functionality
   - About
   - The Third National Climate Assessment
   - Identifiers
   - Sources of Information
   - RESTful API
   - Semantic API
2. Information Model
   - Relational
   - Semantic
3. System Architecture
   - Diagram
   - Lexicons
   - Relational Model Changes
   - Semantic Model Changes
4. Conclusion, Ongoing Work, Future Plans

About

http://data.globalchange.gov

The Third National Climate Assessment

On May 6, 2014, the US Global Change Research Program released the Third National Climate Assessment (NCA3).

- PDF(s) (`http://data.globalchange.gov/report/nca3`)
- Website (`http://nca2014.globalchange.gov`)

The Third National Climate Assessment

- 829 pages
- Over 300 authors, numerous editors, analysts, graphics producers, scientists, data scientists, software developers, and web teams
- 161 findings, 284 figures, 3,395 bibliographic references (journal articles, books, reports

The Third National Climate Assessment

Report Construction

- Various sources spreadsheets, google docs, email, endnote, scientific software, CMSs.
- Common points of reference and vocabulary
- Uniform Resource Identifiers and URLs
- Fine grained tracking of all changes.
- Highly scalable information retrieval

Identifiers

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

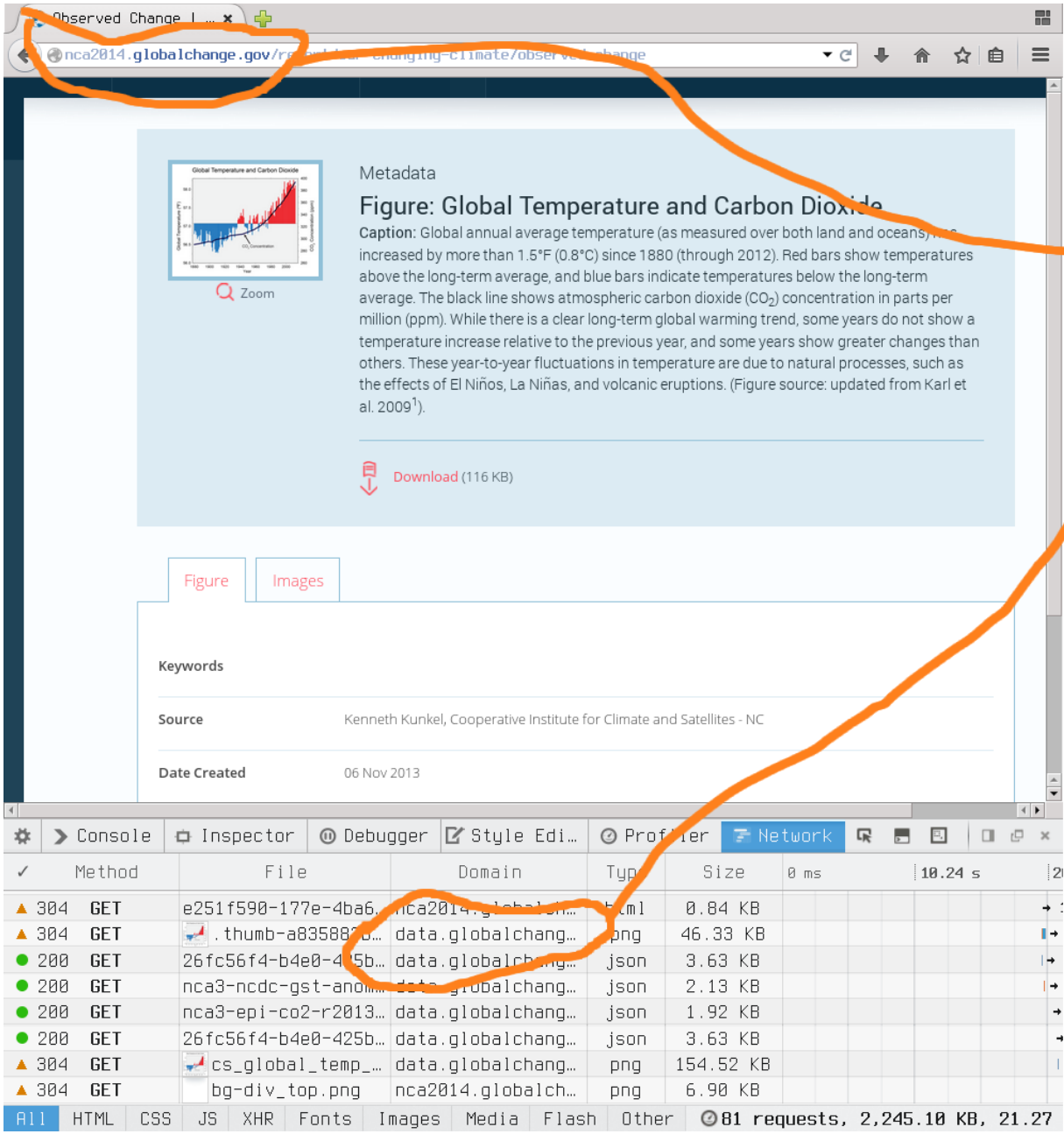
Identifiers

- /article/10.1080/15287390801997625
- /report/usfs-pnw-gtr-855
- /reference/007a7014-723e-4ceb-a395-5c986b1bf884
- /image/26fc56f4-b4e0-425b-adc8-14c6d961d558
- /organization/nasa
- /person/0000-0001-6667-7047
- /dataset/nca3-cddv2-r1
- /platform/aqua
- /instrument/modis
- /model/ccsm3

Sources of Information

- Committee on Earth Observation Satellites (CEOS) (platforms, instruments)
- NASA Physical Oceanography DAAC (datasets)
- Lawrence Livermore National Lab, PCMDI (model runs)
- crossref.org (journal articles)
- `http://github.com/usgcrp/gcis-sync`

RESTful API


nca2014.globalchange.gov
data.globalchange.gov
Metadata
Figure: Global Temperature and Carbon Dioxide
Caption: Global annual average temperature (as measured over both land and oceans) has 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 (CO_2) 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. 2009[1]).
Zoom
Download (116 KB)
Figure
Images
Keywords
Source
Kenneth Kunkel, Cooperative Institute for Climate and Satellites - NC
Date Created
06 Nov 2013
Console
Inspector
Debugger
Style Edi…
Network
Method
File
Domain
Size
304 GET e251f590-177e-4ba6… 0.84 KB
304 GET .thumb-a83588… data.globalchang… png 46.33 KB
200 GET 26fc56f4-b4e0-4… data.globalch… json 3.63 KB
200 GET nca3-ncdc-gst-ano… data.globalchang… json 2.13 KB
200 GET nca3-epi-co2-r2013… data.globalchang… json 1.92 KB
200 GET 26fc56f4-b4e0-425b… data.globalchang… json 3.63 KB
304 GET cs_global_temp_… data.globalchang… png 154.52 KB
304 GET bg-div_top.png nca2014.globalch… png 6.90 KB
All HTML CSS JS XHR Fonts Images Media Flash Other
81 requests, 2,245.10 KB, 21.27

Semantic API

platform

↑

instrument

↑

dataset

↑

dataset

↑

journal article

↑

figure

Semantic API

```
PREFIX gcis: <http://data.globalchange.gov/gcis.owl#> .
PREFIX prov: <http://www.w3.org/ns/prov#> .

SELECT  ?figure ?article ?dataset ?dataset ?instrument ?platform
FROM <http://data.globalchange.gov>
WHERE {
   ?figure prov:isDerivedFrom ?article .
   ?article prov:isDerivedFrom ?dataset .
   ?dataset prov:isDerivedFrom ?another_dataset .
   ?another_dataset gcis:hasInstrument ?instrument .
   ?instrument gcis:hasPlatform ?platform .

   ?figure gcis:isFigureOf ?chapter .
   ?chapter gcis:ChapterNumber "2" .
   ?report gcis:hasChapter ?chapter .
   ?report dcterms:title "The Third National Climate Assessment"
}
```

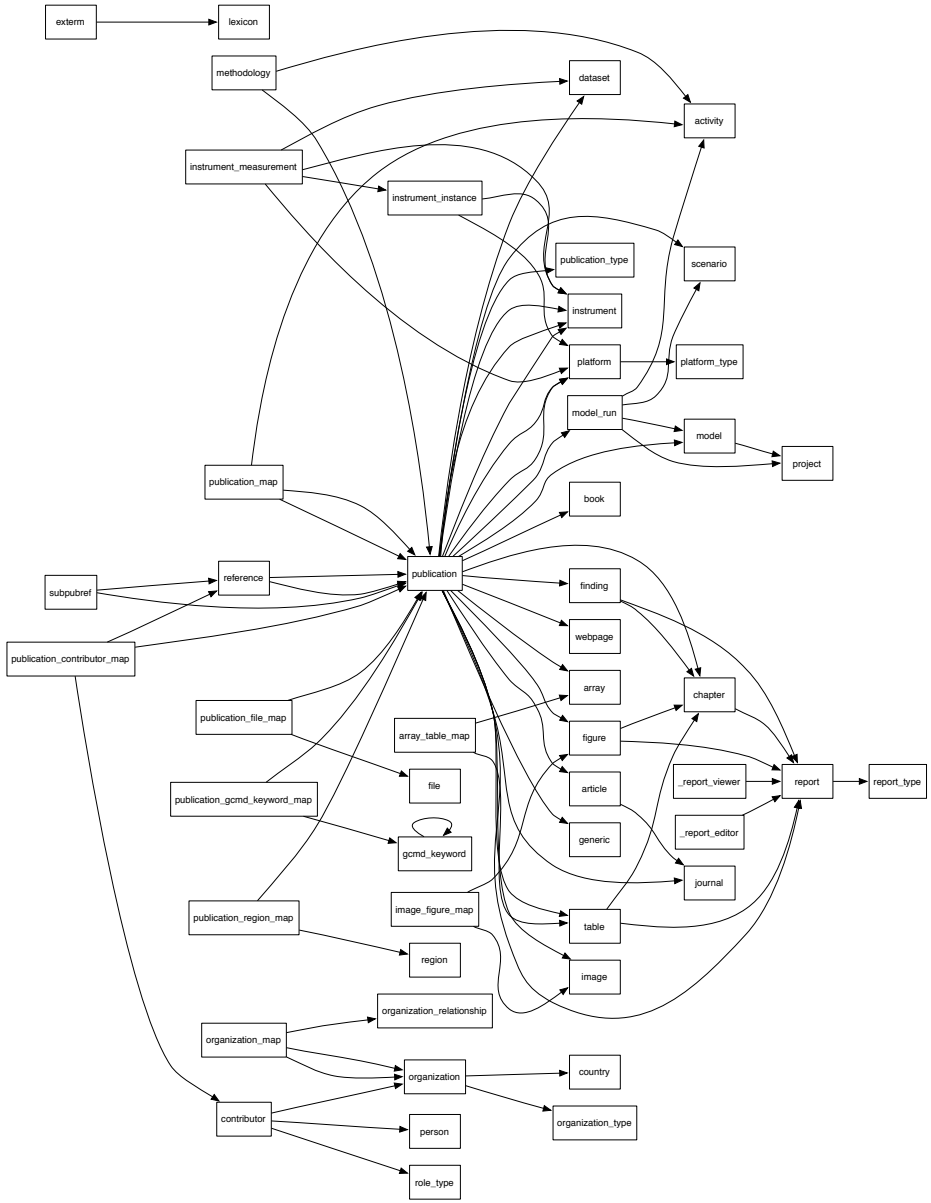
Outline

1. Introduction and Functionality
   - About
   - The Third National Climate Assessment
   - Identifiers
   - Sources of Information
   - RESTful API
   - Semantic API
2. Information Model
   - Relational
   - Semantic
3. System Architecture
   - Diagram
   - Lexicons
   - Relational Model Changes
   - Semantic Model Changes
4. Conclusion, Ongoing Work, Future Plans

Relational

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

Relational


exterm
lexicon
methodology
dataset
activity
instrument_measurement
instrument_instance
publication_type
scenario
instrument
platform
platform_type
model_run
model
project
publication_map
book
publication
reference
subpubref
finding
webpage
publication_contributor_map
array
chapter
publication_file_map
array_table_map
figure
file
publication_gcmd_keyword_map
article
_report_viewer
report
report_type
generic
_report_editor
gcmd_keyword
journal
image_figure_map
publication_region_map
table
region
image
organization_relationship
organization_map
organization
country
contributor
organization_type
person
role_type

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

Semantic

http://bit.ly/gcis-dbpedia

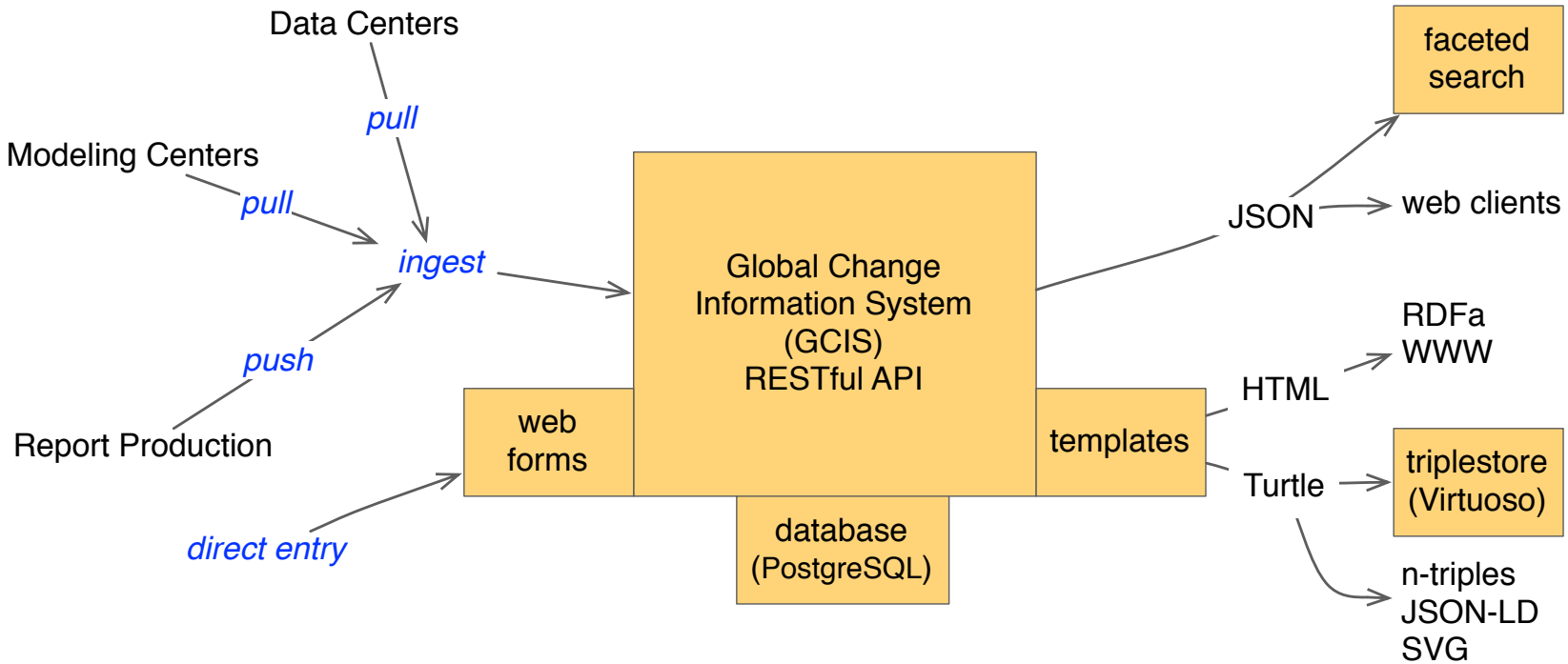
```
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX gcis: <http://data.globalchange.gov/gcis.owl#>
PREFIX cito: <http://purl.org/spar/cito/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dbprop: <http://dbpedia.org/property/>
PREFIX dbpo: <http://dbpedia.org/ontology/>

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 that have had an article cited by the NCA3 report.

Outline

1. Introduction and Functionality
   - About
   - The Third National Climate Assessment
   - Identifiers
   - Sources of Information
   - RESTful API
   - Semantic API
2. Information Model
   - Relational
   - Semantic
3. System Architecture
   - Diagram
   - Lexicons
   - Relational Model Changes
   - Semantic Model 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
templates
database
(PostgreSQL)
JSON
faceted
search
web clients
HTML
RDFa
WWW
Turtle
triplestore
(Virtuoso)
n-triples
JSON-LD
SVG

Lexicons

Associations between GCIS identifiers and external terms.

- GCIS identifier : /podaac/aqua
- PODAAC identifier : AQUA
- CEOS identifier : 206

Relational Model 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.

Semantic Model 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.

Semantic Model 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
   - About
   - The Third National Climate Assessment
   - Identifiers
   - Sources of Information
   - RESTful API
   - Semantic API
2. Information Model
   - Relational
   - Semantic
3. System Architecture
   - Diagram
   - Lexicons
   - Relational Model Changes
   - Semantic Model 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 Manipon, Angel Li, April Sides, Sarah Champion, Bradley Akamine, Amanda McQueen, Peter Fox, Marshall Ma and the Tetherless World Institute at RPI