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
    - Mission
    - Support NCA3 report production
    - Serve as the NCA3 website backend
    - Provide provenance of resources
    - Be a source of reliable information
    - Connect disparate sources of information
2. Information Model
    - Relational
    - Semantic
    - Example
3. System Architecture
    - Diagram
    - Schema Changes
    - Ontology Changes
    - Updating Content
4. Conclusion, Ongoing Work, Future Plans

Outline

1. Introduction and Functionality
    - Mission
    - Support NCA3 report production
    - Serve as the NCA3 website backend
    - Provide provenance of resources
    - Be a source of reliable information
    - Connect disparate sources of information
2. Information Model
    - Relational
    - Semantic
    - Example
3. System Architecture
    - Diagram
    - Schema Changes
    - Ontology Changes
    - Updating Content
4. Conclusion, Ongoing Work, Future Plans

Mission

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.

Mission

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.

Support NCA3 report production

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 data, as well as an API for accepting data in a variety of formats.

Serve as the NCA3 website backend

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 populated elements of some pages dynamically.

todo insert picture here

Provide provenance of resources

The GCIS ensures compliance with the Information Quality Act by providing traceable identifiers for sources of information.
Given a figure, find the datasets and instruments associated the data behind it.

todo sea level rise graphic

Be a source of reliable information

The inverse of provenance.

Sample questions :

Given a dataset, find reports with figures generated from the dataset.

Show figures associated with data generated by instruments funded by NASA.

Connect disparate sources of information

Identifiers for resources are curated by various organizations; find ways to link them together.

Outline

1. Introduction and Functionality
   - Mission
   - Support NCA3 report production
   - Serve as the NCA3 website backend
   - Provide provenance of resources
   - Be a source of reliable information
   - Connect disparate sources of information
2. Information Model
   - Relational
   - Semantic
   - Example
3. System Architecture
   - Diagram
   - Schema Changes
   - Ontology Changes
   - Updating Content
4. Conclusion, Ongoing Work, Future Plans

Relational

A relational model provides one/many to one/many relationships.
One report has many chapters.
One journal has many articles.
Many datasets support many articles.

Critical features : referential integrity, type checking, check constraints, cascading updates of primary keys (identifiers), high performance, well established scalable tools. Postgres also offers some useful features like hstores for looser information.

Closed world assumption for some contexts. A figure is related to a chapter if and only if it is in the chapter.

The canonical representation of GCIS data is in the Postgres database.

Semantic

The semantic representation has benefits such as
The ability to describe how two entities are related.
A figure **was influenced by** a report.
An article **was cited by** a report.
A dataset **was derived from** another dataset.
A GCIS ontology defines semantic relationships between entities.
The semantic representation allows GCIS data to be used with external data.
Open world assumption : a figure may be related to a dataset, but this is not in GCIS.

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)
}
```

Outline

1. Introduction and Functionality
   - Mission
   - Support NCA3 report production
   - Serve as the NCA3 website backend
   - Provide provenance of resources
   - Be a source of reliable information
   - Connect disparate sources of information
2. Information Model
   - Relational
   - Semantic
   - Example
3. System Architecture
   - Diagram
   - Schema Changes
   - Ontology Changes
   - Updating Content
4. Conclusion, Ongoing Work, Future Plans

Diagram

todo

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 %>;
```

Updating Content

todo lexicons here

Outline

1. Introduction and Functionality
   - Mission
   - Support NCA3 report production
   - Serve as the NCA3 website backend
   - Provide provenance of resources
   - Be a source of reliable information
   - Connect disparate sources of information
2. Information Model
   - Relational
   - Semantic
   - Example
3. System Architecture
   - Diagram
   - Schema Changes
   - Ontology Changes
   - Updating Content
4. Conclusion, Ongoing Work, Future Plans