The Global Change Information System

Reports, Data, APIs, Communication and Information

Brian Duggan[12], Steve Aulenbach[12],
Robert Wolfe[23], Justin Goldstein[12]

[1]UCAR, [2]USGCRP, [3]NASA

July 10, 2014

http://data.globalchange.gov
http://github.com/USGCRP/gcis

Outline

What

- The Third National Climate Assessment
- PDF(s)
- Website
- Information System

Who

- Scientists (Authors)
- Science analysts
- Editors
- Graphic designers
- Web developers
- Project Managers

How

- Spreadsheets
- Google docs
- Email
- Endnote
- Scientific Software
- Graphics Software
- Content Management Systems
- Wikis
- Various miscellaneous desktop and cloud software

Resources

The tools are used to represent and manipulate various resources.

- Journal Articles
- Reports
- References
- Images, Figures, Tables
- Findings
- Organizations
- People
- Datasets

Role of the GCIS

- Common points of reference
- Common vocabulary across teams
- Language, terminology, vocabulary, ontology
- Uniform Resource Identifiers
- URIs are actionable : URLs
- Information manipulation via API or web forms
- Information extraction via API or browsing
- Information modeling with relational or semantic models
- Fine grained tracking of all changes.
- Convenient useful information entry
- Highly scalable information retrieval

Functionality

- Support nca2014.globalchange.gov (jQuery)
- GCIS provides JSON backend
- Figures have client side calls
- Also uses dataset landing pages
- API used for data ingestion and retrieval
- Support semantic queries
- SPARQL endpoint `http://data.globalchange.gov/sparql`
- JSON, RDF, RDF-A, HTML, Turtle, RDF-XML

Architecture

- RDBMS (PostgreSQL) for storage
- HTML templates
- Turtle templates into other formats
- Scrape into triple store
- Data structures into JSON, YAML

Server

- Perl (mojolicious)
- nginx, proxies
- postgres
- Virtuoso
- Caching

Clients

- Python (Andrew)
- Perl
- Javascript (jQuery)
- php (Drupal)

SPARQL

http://bit.ly/1ilgeQz

```
PREFIX dbpediaowl: <http://dbpedia.org/ontology/>
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX gcis: <http://data.globalchange.gov/gcis.owl#>
PREFIX cito: <http://purl.org/spar/cito/>

SELECT DISTINCT ?gcisjournal
FROM <http://data.globalchange.gov/sparql>
WHERE

   SERVICE <http://data.globalchange.gov/sparql>
     ?gcisjournal a bibo:Journal .
     ?gcisjournal bibo:issn ?issn .
     ?gcisarticle gcis:inPublication ?gcisjournal .
     ?gcisarticle cito:isCitedBy <http://data.globalchange.gov/report/nca3> .

   BIND(STRLANG(?issn, "en") AS ?issn_en)

   SERVICE <http://dbpedia.org/sparql>
     ?dbpjournal dbpediaowl:frequencyOfPublication "Monthly"@en .
     ?dbpjournal dbpediaowl:issn ?issn_en .
     FILTER(STR(?issn_en) = ?issn)
```

Versioning

- git
- Postgres audit triggers

Testing

- Test driven development (unit tests)
- SPARQL tests
- Continuous Integration Testing
- Test driven data acquisition
- Continuous Content Validation (gcis-qa)

Narrative vs structure

Semantic vs Relational

Resources

Identifiers

Publications, Contributors

http://data.globalchange.gov/resources

Discussion