Adaptable Information Models in the Global Change Information System

Brian Duggan[1,2], Andrew Buddenberg[3], Steve Aulenbach[1,2], Robert Wolfe[4], Justin Goldstein[1,2]

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

July 16, 2014

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

Outline

NCA3 Report Assembly
Implementation
Information Model
Discussion

What
Who
How
Resources
Role of the GCIS
Identifiers

What

- The Third National Climate Assessment
- PDF(s) (http://data.globalchange.gov/report/nca3)
- Website (http://nca2014.globalchange.gov)

NCA3 Report Assembly
Implementation
Information Model
Discussion

What
Who
How
Resources
Role of the GCIS
Identifiers

Who

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

NCA3 Report Assembly
Implementation
Information Model
Discussion

What
Who
How
Resources
Role of the GCIS
Identifiers

How

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

NCA3 Report Assembly
Implementation
Information Model
Discussion

What
Who
How
Resources
Role of the GCIS
Identifiers

Resources

The tools are used to represent and manipulate various resources.

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

NCA3 Report Assembly
Implementation
Information Model
Discussion

What
Who
How
Resources
Role of the GCIS
Identifiers

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

NCA3 Report Assembly
Implementation
Information Model
Discussion

What
Who
How
Resources
Role of the GCIS
Identifiers

Resources

GCIDs

http://data.globalchange.gov

- /article/10.1080/15287390801997625
- /report/usfs-pnw-gtr-855
- /reference/007a7014-723e-4ceb-a395-5c986b1bf884
- /report/nca3/figure/global-temperature-and-carbon-dioxide
- /image/26fc56f4-b4e0-425b-adc8-14c6d961d558
- /report/nca3/table/decisions-scales
- /report/nca3/finding/extreme-precipitation-increase
- /organization/nasa
- /person/0000-0001-6667-7047
- /dataset/nca3-cddv2-r1

NCA3 Report Assembly
Implementation
Information Model
Discussion

Functionality
SPARQL
Testing
Server Architecture
Clients

Functionality

- Support NCA3 report production
- Support NCA3 website (client side jQuery)
- Provide minimal landing pages for resources
- Provide a public JSON API
  http://data.globalchange.gov/api_reference
- Provide semantic information
- Be interoperable (e.g. use existing identifiers)
- Provide a public SPARQL endpoint
  http://data.globalchange.gov/sparql
- JSON, RDF, schema.org, HTML, Turtle, RDF-XML

NCA3 Report Assembly
Implementation
Information Model
Discussion

Functionality
SPARQL
Testing
Server Architecture
Clients

SPARQL

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

NCA3 Report Assembly
Implementation
Information Model
Discussion

Functionality
SPARQL
Testing
Server Architecture
Clients

Testing

- Test driven development (unit tests)
- SPARQL tests
- Continuous Integration Testing (github, travis-ci.org)
- Test driven data acquisition
- Continuous Content Validation
  http://github.com/USGCRP/gcis-qa

NCA3 Report Assembly
Implementation
Information Model
Discussion

Functionality
SPARQL
Testing
Server Architecture
Clients

Server Architecture

- RDBMS (PostgreSQL) for storage
  Fine-grained transaction auditing, referential integrity
- HTML templates
- Turtle templates (and other formats)
- Scrape into triple store (Virtuoso)
- Data structures into JSON, YAML
- nginx reverse proxy cache

NCA3 Report Assembly
Implementation
Information Model
Discussion

Functionality
SPARQL
Testing
Server Architecture
Clients

Clients

- Python (Andrew)
  http://github.com/USGCRP/gcis-py-client
- Perl
  http://github.com/USGCRP/gcis-pl-client
- Javascript (jQuery)
- php (Drupal)

NCA3 Report Assembly
Implementation
Information Model
Discussion

Concepts
Details

Narrative vs structure

NCA3 Report Assembly
Implementation
Information Model
Discussion

Concepts
Details

Semantic vs Relational

NCA3 Report Assembly
Implementation
Information Model
Discussion

Concepts
Details

Resources

NCA3 Report Assembly
Implementation
Information Model
Discussion

Concepts
Details

Identifiers

NCA3 Report Assembly
Implementation
Information Model
Discussion

Concepts
Details

Publications, Contributors (Entities, Agents, Activities)

NCA3 Report Assembly
Implementation
Information Model
Discussion

Concepts
Details

http://data.globalchange.gov/resources

Discussion