Global Change Information System

Technical Updates

Presenter: Brian Duggan
US Global Change Research Program
ESIP Summer 2015

Last year..
Global Change Information System

- RESTful API
- SPARQL Endpoint
- Persistent URIs
- http://data.globalchange.gov...
  - /report/nca3
  - /report/nca3/chapter/2
  - /report/nca3/chapter/2/figure/26
  - /person/0000-0001-7599-9750
  - /dataset/nca3-cddv2-r1
Since then...
1,216 commits

$ git rev-list --since="@{"one year ago"}" HEAD --count
1216

https://github.com/USGCRP/gcis
https://travis-ci.org/USGCRP/gcis
https://coveralls.io/github/USGCRP/gcis
SELECT sum(n_live_tup) 
FROM pg_stat_user_tables 
where schemaname='gcis_metadata'

39358

94209

http://data-stage.globalchange.gov/export/gcis.txt
137,683 database changes

```sql
select action, count(1)
from audit.logged_actions
where action_tstamp_tx > now() - interval '1 year'
group by 1;
```

<table>
<thead>
<tr>
<th>action</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>7767</td>
</tr>
<tr>
<td>I</td>
<td>62584</td>
</tr>
<tr>
<td>U</td>
<td>67332</td>
</tr>
</tbody>
</table>

(3 rows)
179,847 new triples

```
SELECT (COUNT(*) AS ?triples)
from <http://data.globalchange.gov>
{ ?s ?p ?o }
```

311407

131560

http://data.globalchange.gov/sparql
More resources

$ curl http://data.globalchange.gov/metrics.yaml

```
counts:
  person: 8867 # 1141
  dataset: 2377 # 23
  article: 2150 # 2086
  organization: 1204 # 845
  report: 821 # 704
  figure: 575 # 490
  image: 579 # 274
  instrument: 530 # 0
  journal: 544 # 536
  book: 226 # 166
  platform: 282 # 0
  model: 33 # 0
  scenario: 3 # 0
```

http://data.globalchange.gov/metrics.yaml
More types of resources

- report: nca3
- chapter, figures, findings, tables, references (report production)
- images, articles, journals, books (nca3 references)
- people, organizations (nca3 authors, report authors)
- datasets (report production, images)
- datasets (journal articles, DAACs)
- platforms, instruments (CEOS)
- models, scenarios, model runs (CMIP3, CMIP5, others)
- people (article authors)

http://data.globalchange.gov/resources
More is not always better.
Lite®
A FINE PILSNER
BEER

Everything you always wanted in a beer. And less.
Linked Data

“Putting datasets on the web ... is actually pretty easy. ... Linking from one data set to another, so that both people and machines understand that a resource in one data set is the same thing as a resource in a second data set is far harder. Those explicit links can be made, but it requires agreement, standards, and an awful lot of hard work.” -- timbl, 2015-07-10
Established Identifiers

- DOI
- ISSN
- ISBN
- ORCID
- URL?
## Lexicons

http://data.globalchange.gov/lexicon/

<table>
<thead>
<tr>
<th>Platform</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>AATSR</td>
<td>/instrument/advanced-along-track-scanning-radiometer</td>
</tr>
<tr>
<td>AltiKa</td>
<td>/instrument/ka-band-altimeter</td>
</tr>
<tr>
<td>AMI</td>
<td>/instrument/active-microwave-instrumentation-image-mode</td>
</tr>
<tr>
<td>AMR</td>
<td>/instrument/advanced-microwave-radiometer</td>
</tr>
<tr>
<td>AMSR</td>
<td>/instrument/advanced-microwave-scanning-radiometer</td>
</tr>
<tr>
<td>AMSR2</td>
<td>/instrument/advanced-microwave-scanning-radiometer-2</td>
</tr>
<tr>
<td>AMSR-E</td>
<td>/instrument/advanced-microwave-scanning-radiometer-ecs</td>
</tr>
<tr>
<td>AQUARIUS_RADIOMETER</td>
<td>/instrument/aquarius-l-band-radiometer</td>
</tr>
<tr>
<td>AQUARIUS_SCATTEROMETER</td>
<td>/instrument/aquarius-l-band-scatterometer</td>
</tr>
<tr>
<td>ASCAT</td>
<td>/instrument/advanced-scatterometer</td>
</tr>
<tr>
<td>ATSR-2</td>
<td>/instrument/along-track-scanning-radiometer-2</td>
</tr>
<tr>
<td>AVHRR-2</td>
<td>/instrument/advanced-very-high-resolution-radiometer-2</td>
</tr>
<tr>
<td>AVHRR-3</td>
<td>/instrument/advanced-very-high-resolution-radiometer-3</td>
</tr>
<tr>
<td>DORIS</td>
<td>/instrument/doris-ng</td>
</tr>
<tr>
<td>ENVISAT RA-2</td>
<td>/instrument/radar-altimeter-2-2</td>
</tr>
</tbody>
</table>

http://events.linkeddata.org/ldow2015/
Merge and redirect

$ curl -vL http://data.globalchange.gov/person/4990.json
...
< HTTP/1.1 302 Found
< Location: /person/1024.json
> GET /person/1024.json HTTP/1.1
< HTTP/1.1 200 OK
{
    "uri" : "/person/1024",
...
}
Work in progress

- Scaling up linking
- Health report
- Controlled vocabularies
- Station data
- Health datasets (EPA, CDC)
- Models
Thank you

Questions:

• What identifiers do you use?
• What identifiers do you create?
• Who uses your identifiers?
• How do you construct and use URIs and URLs?
• Who manages identifier mappings?

http://data.globalchange.gov/about