

# Linked Science as a producer and consumer of big data in the Earth Sciences

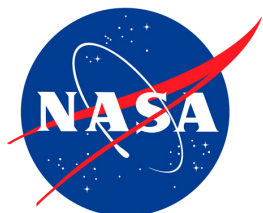
Line C. Pouchard,\* Robert B. Cook,\* Jim Green,\*  
Natasha Noy,\*\* Giri Palanisamy\*

Oak Ridge National Laboratory\*

Stanford Center for Biomedical Informatics Research \*\*

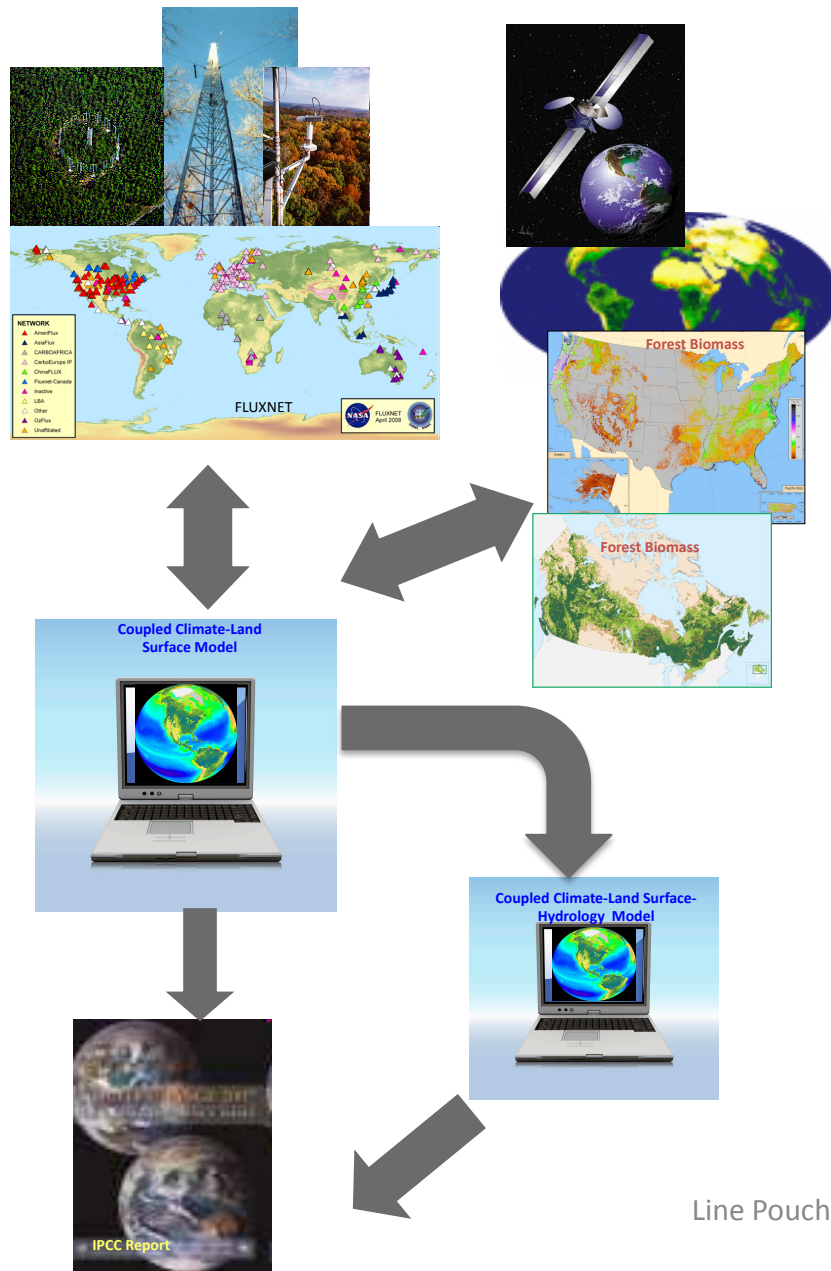
Presented to the Ontology Summit 2012

March 8, 2012



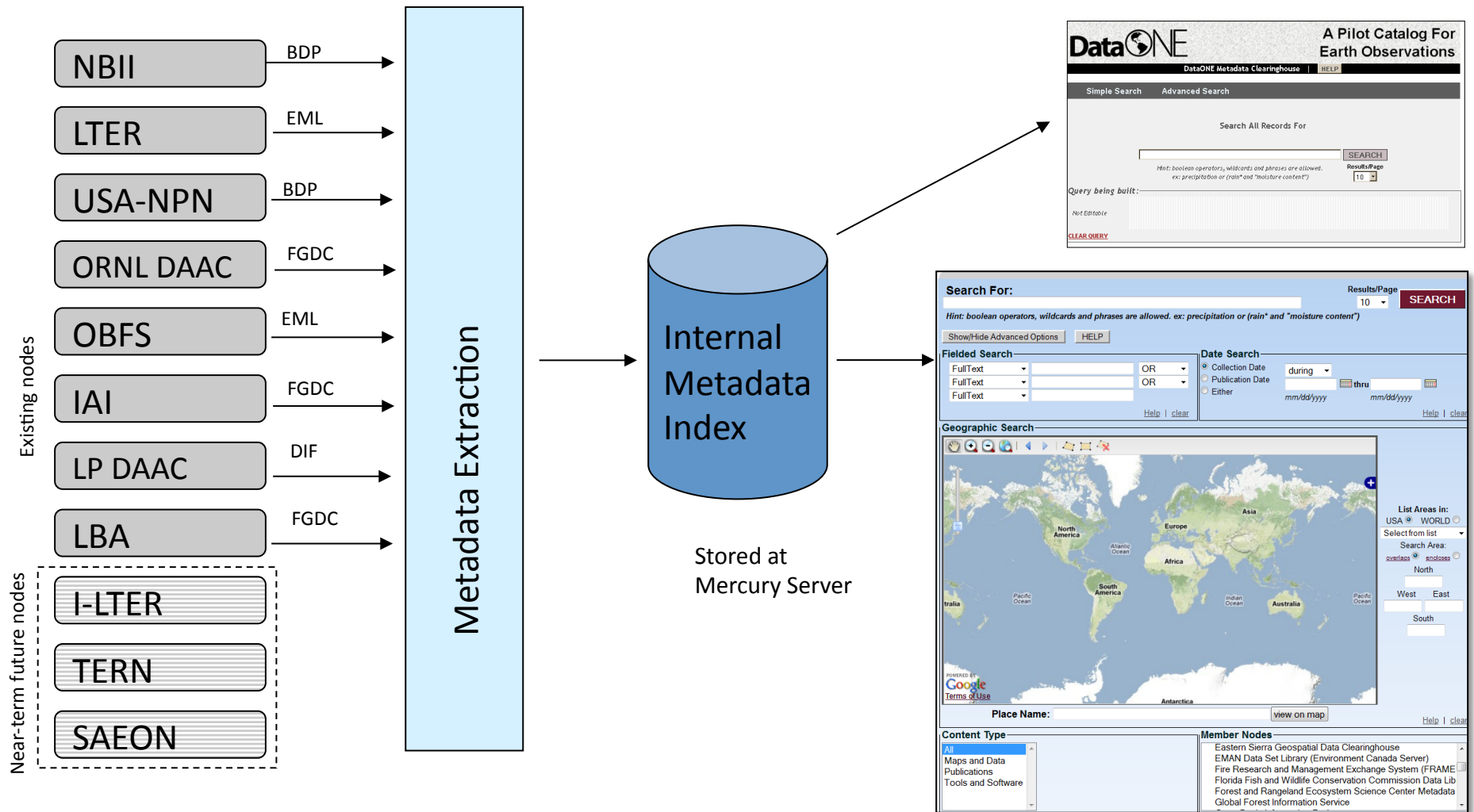


# Linked Science in the Earth Sciences

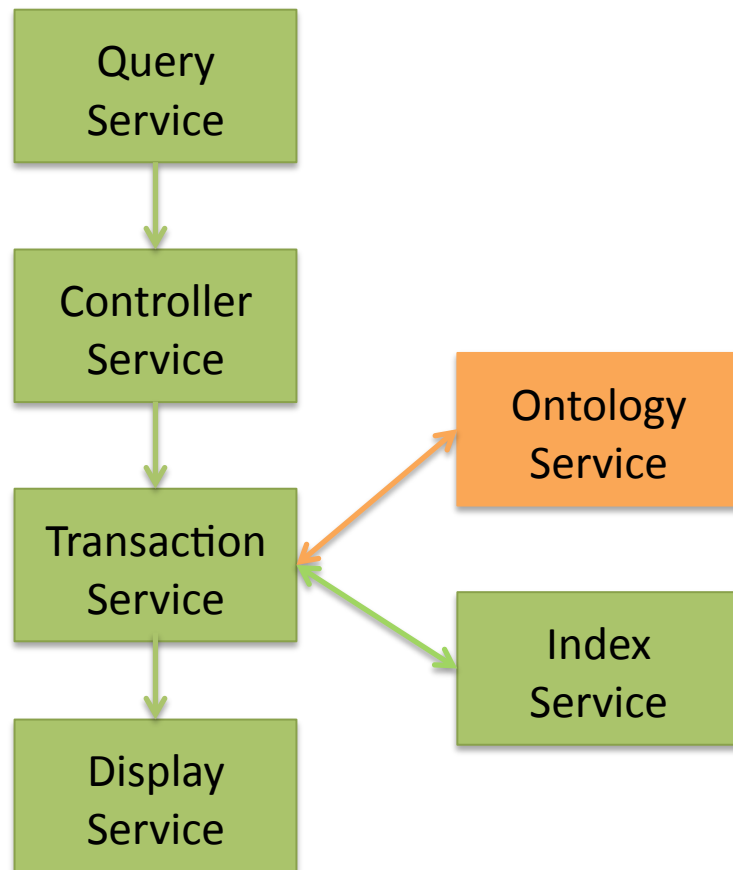


- Discovery and access from heterogeneous sources
  - Simulations, models, experiments, remote sensing, GIS, molecular and omics databases, publications
- Metadata and semantics integration
- Workflows, scenario development, data and process re-use, provenance
- Engaging communities of scientists, educators, librarians, developers, volunteers
- Relies upon cyber-infrastructure promoting open source
- Complex systems of systems, networks of projects, repositories, archives, publishers

# Mercury: federated metadata engine for earth sciences data



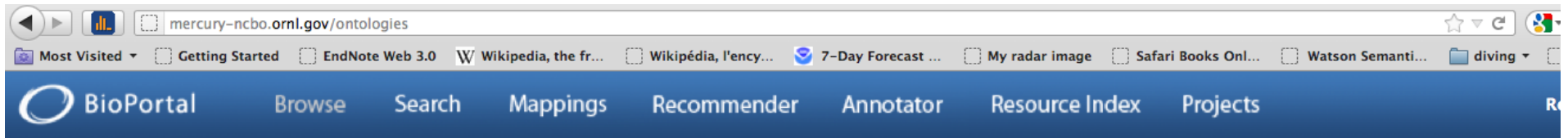
# Coupling the Mercury federated metadata engine and BioPortal



- Uses BioPortal Rest Services for programmatic access
- Returns ontology concepts, super- and sub-classes
- Provides additional keywords
- Provides context
- Uses these for new searches



# BioPortal provides access to ontologies



## Browse

Access all ontologies that are available in NCBO BioPortal: You can filter this list by category to display ontologies relevant for a certain domain. You can also filter c [Subscribe to the NCBO BioPortal RSS feed](#) to receive alerts for submissions of new ontologies, new versions of ontologies, new notes, and new projects. You can subscri individual ontology page. Add a new ontology to NCBO BioPortal using the Submit New Ontology link.

|                                   |                              |
|-----------------------------------|------------------------------|
| FILTER BY CATEGORY                | All Categories               |
| FILTER BY GROUP <a href="#">?</a> | All Groups <a href="#">?</a> |
| FILTER BY TEXT                    | <input type="text"/>         |

[Submit New Ontology](#)

| ONTOLOGY NAME  | VISIBILITY             | TERMS                 | NOTES | REVIEWS | PROJECTS | UPLOADED   |
|--|------------------------|-----------------------|-------|---------|----------|------------|
| <a href="#">OBOE (OBOE)</a>  | <a href="#">Public</a> | <a href="#">40</a>    | 0     | 0       | 0        | 01/31/2012 |
| <a href="#">OBOE-SBC (OBOE-SBC)</a>  | <a href="#">Public</a> | <a href="#">630</a>   | 0     | 0       | 0        | 01/31/2012 |
| <a href="#">Plant Ontology (PO)</a>  | <a href="#">Public</a> | <a href="#">1,448</a> | 0     | 0       | 0        | 01/31/2012 |
| <a href="#">Semantic Web for Earth and Environment Terminology (SWEET)</a> | <a href="#">Public</a> | <a href="#">4,534</a> | 0     | 0       | 0        | 01/27/2012 |

Showing 1 to 4 of 4 entries

Powered by [NCBO BioPortal](#) [Release Notes](#)

Domain ontologies are used within the engineered system to improve faceted search results

# Ontology-based search results

[Bookmark](#) [Email](#) [Help](#)

**Metadata Summary**

Your search found: 1227 documents.  
 Query: text : biomass AND ( datasource : ( daac landval rgd lpcol lter obfs ) )  
 Now try this to get ontology results : "animal waste" OR "detritus" OR "fertilizer" OR "food" OR "gelbstoff" OR "humus" OR "litter" OR "manure" OR "wood"

Choose records from: [LTER DATA \(950\)](#) [DAAC DATASETS \(187\)](#) [REGIONAL AND GLOBAL DATA \(62\)](#) [LAND VALIDATION DATA \(12\)](#) [LP DAAC - MO...](#) [PRODUCTS \(8\)](#) [ORGANIZATION OF BIOLOGICAL FIELD STATIONS \(8\)](#)

| Filter by sensor   | Filter by topic  | Filter by project   | Filter by keywords |
|--|--|---|--------------------|
| analysis (96)<br>weighing balance (94)<br>quadrat sampling frame (77)<br>steel measuring tape (62)<br>soil coring device (59)<br>human observer (48) | biosphere (238)<br>land surface (65)<br>atmosphere (57)<br>human dimensions (26)<br>agriculture (10)<br>hydrosphere (7)<br>solid earth (6) | net primary productivity.. (82)<br>safari 2000 (28)<br>lba (25)<br>eos land validation.. (13)<br>fife (11)<br>superior national |                    |

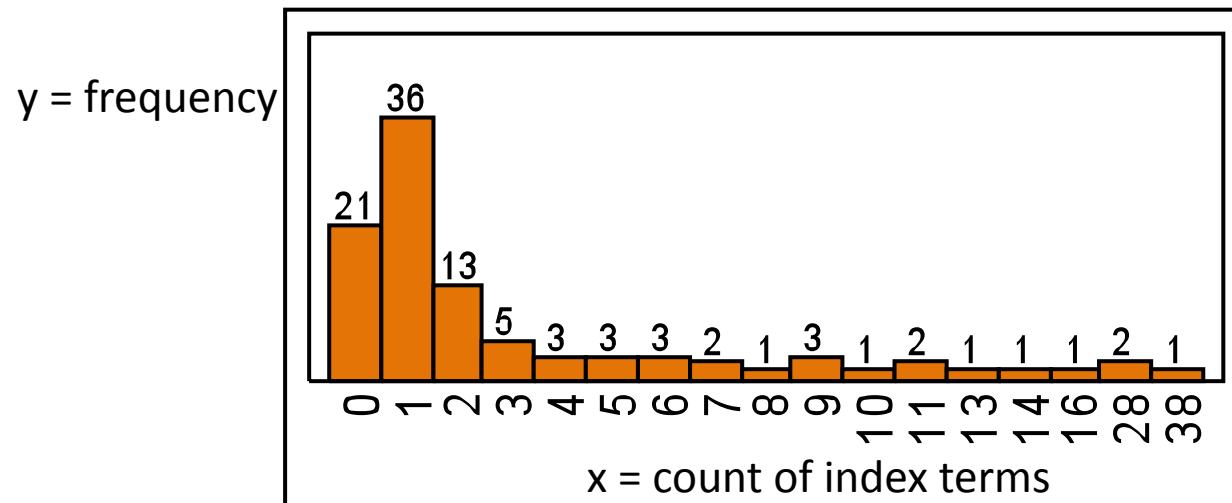
| Ontology Concepts                                 | Ontology SuperClasses           | Ontology SubClasses  | Filter by (keywords) AND (all SubClasses)   |
|---|---------------------------------|--|---|
| matrBiomass.owl#Biomass<br>matrEnergy.owl#Biomass | energy storage<br>living entity | animal waste<br>detritus<br>fertilizer<br>food<br>gelbstoff<br>humus<br>litter<br>manure | forests ( 20 )<br>soils ( 36 )<br>europe ( 33 )<br>biomass ( 19 )<br>eosdis ( 73 )<br>fao ( 23 )<br>vegetation ( 24 )<br>united states ( 18 ) |

Concepts acquire context: biomass as Material or biomass as Energy

Additional search terms

Super-classes may have different properties

# Metrics assessing the impact of using ontologies in search



- Matching the top 100 Mercury parameters to ontology terms
  - Frequency count: 79% of the Top 100 keywords have at least one match in the chosen ontologies
  - N = 99, 2 values missing (plant, leaf)
  - water : 38
  - air, carbon = 28



# Lessons Learned

## User-friendly display

- Current display may be confusing. What are the options?
  - send the user to a new page
  - implement a new display dynamically driven by ontology relationships

## Ontology content

- SWEET provides a good basis, but needs to be further specified for the needs of this Data Center
- Many ontologies provide only few relationships

## Implementation

- Adding ontology entities to a keyword index helps with recall but cannot substitute for semantic annotations of the metadata documents

# Thank you

- ORNL DAAC and Mercury
  - <http://mercury.ornl.gov>
- ORNL DAAC ontology service
  - <http://mercury.ornl.gov/OntologyDemo>
- ORNL DAAC instance of BioPortal
  - <http://mercury-ncbo.ornl.gov>
- Stanford Center for Biomedical Informatics Research BioPortal
  - <http://bioportal.bioontology.org>
- Stanford Center for Biomedical Informatics Research Protégé ontology editor
  - <http://protege.stanford.edu>

The submitted manuscript has been authored by the US Department of Energy, Office of Science of the Oak Ridge National Laboratory, managed for the U. S. DOE by UT-Battelle, LLC, under contract No. DE-AC05-00OR22725.

Line Pouchard, [pouchardlc@ornl.gov](mailto:pouchardlc@ornl.gov)