

# 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

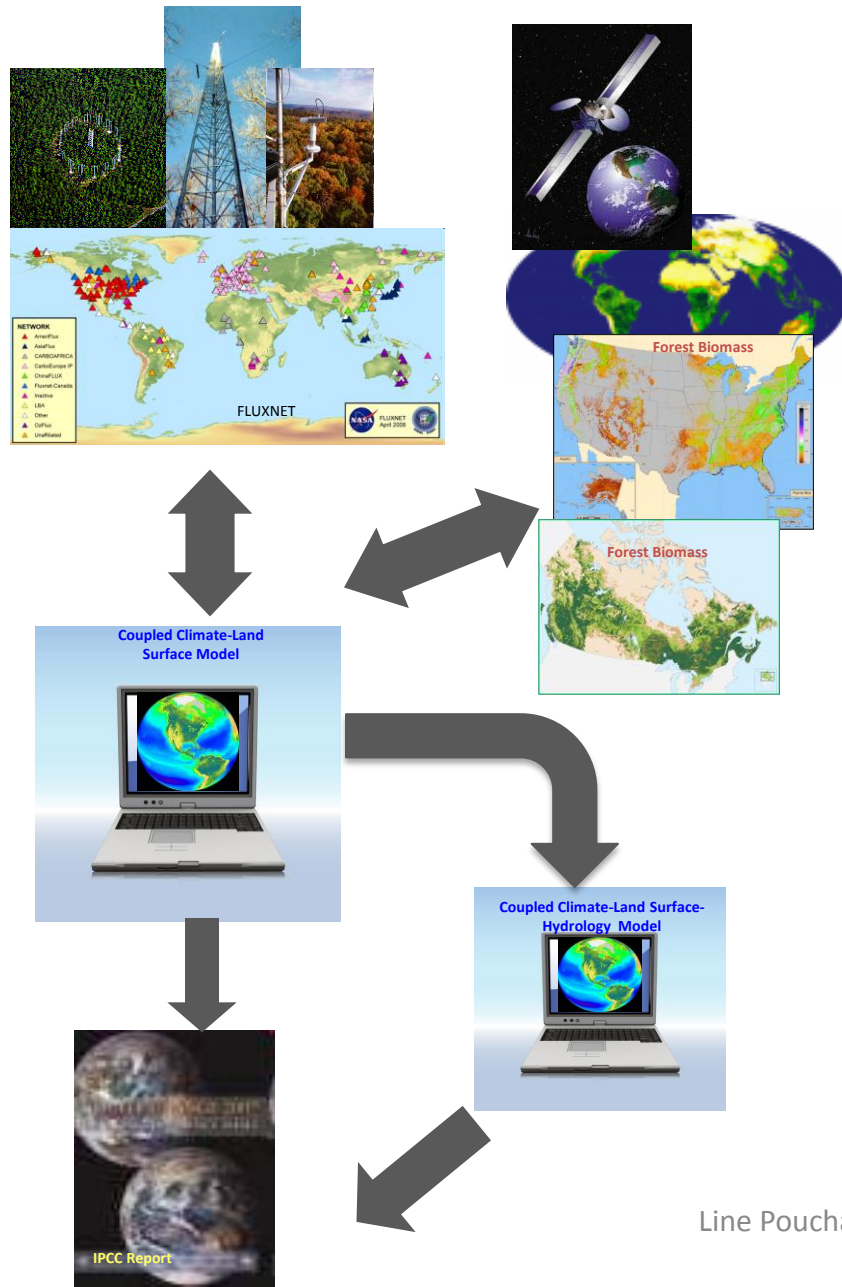


# What is Linked Science?

- Scientific collaborations are growing and becoming more interdisciplinary, Linked Science is:
  - A practice that accounts for objects of study from an end-to-end perspective
  - exhibits key trends in the role of data
  - means of communication between stakeholders are key
  - characterized by experimental, theoretical and simulation methods
  - systematic emphasis on validated datasets, value-added data products, traceable information and integrated processes
- First Linked Science Workshop, International Semantic Web Conference, October 2011
- Second Linked Science Workshop, ISWC, October 2012, Boston

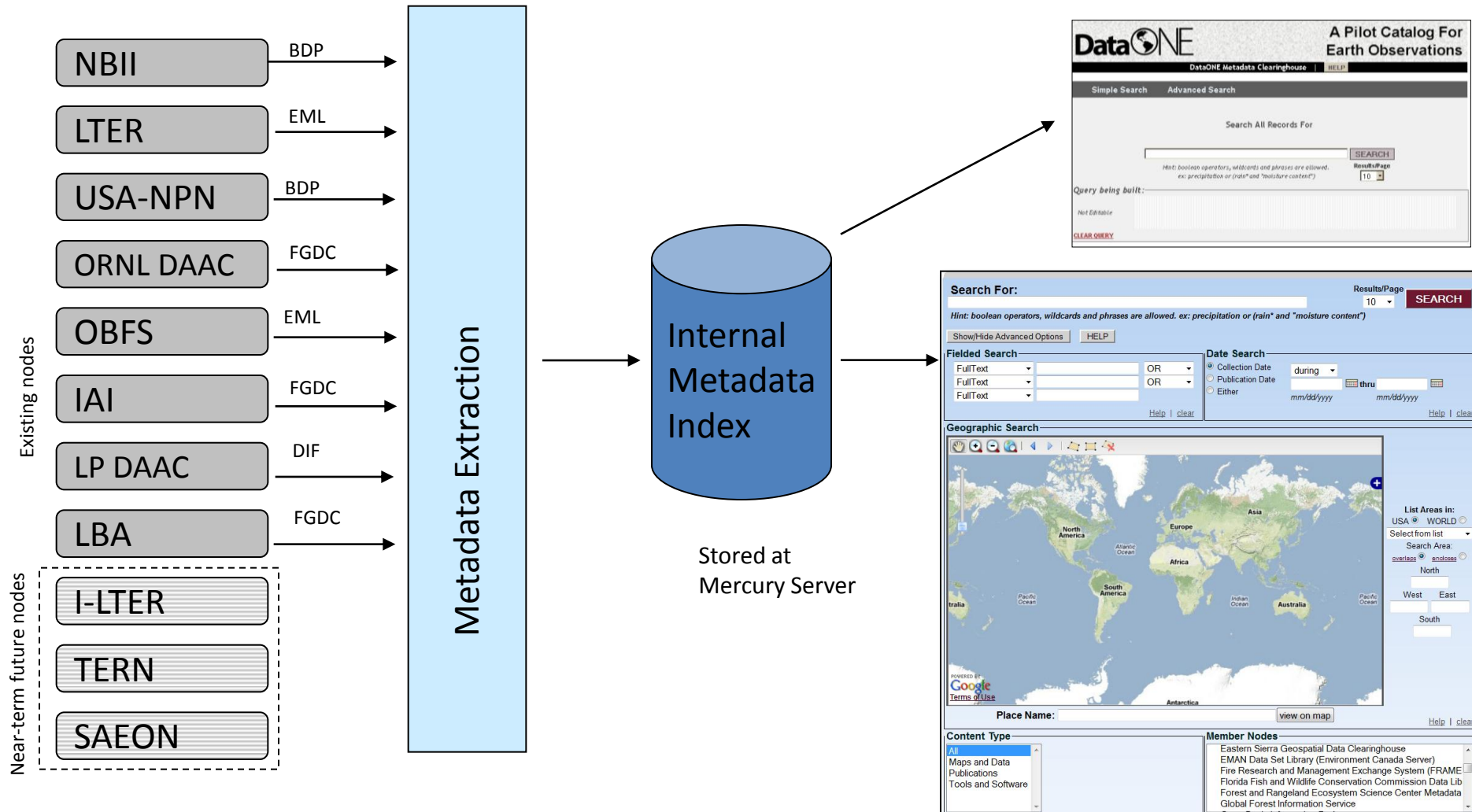


# Linked Science in the Earth Sciences

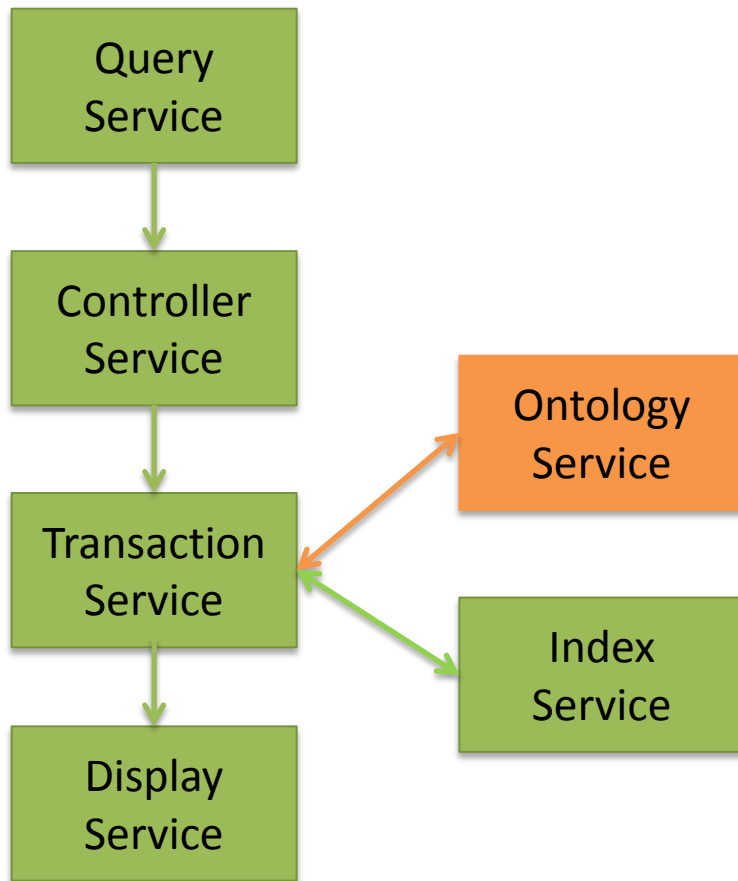


- Discovery and access from heterogeneous sources
  - Simulations, models, experiments, remote sensing, GIS, molecular and omics databases, publications
- Metadata and semantics integration
- Workflows, scenario development, data and process re-use, provenance
- Engaging communities of scientists, educators, 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



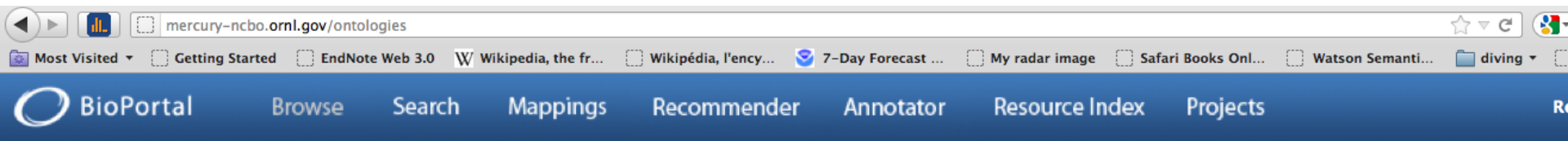
# Coupling the Mercury federated metadata engine and BioPortal



- 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 by group. [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 subscribe to an individual ontology page. Add a new ontology to NCBO BioPortal using the Submit New Ontology link.

FILTER BY CATEGORY	All Categories
FILTER BY GROUP ?	All Groups
FILTER BY 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

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

# Ontology-based search results

**Metadata Summary** [Bookmark](#) [Email](#) [Help](#)

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 "wood"

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

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

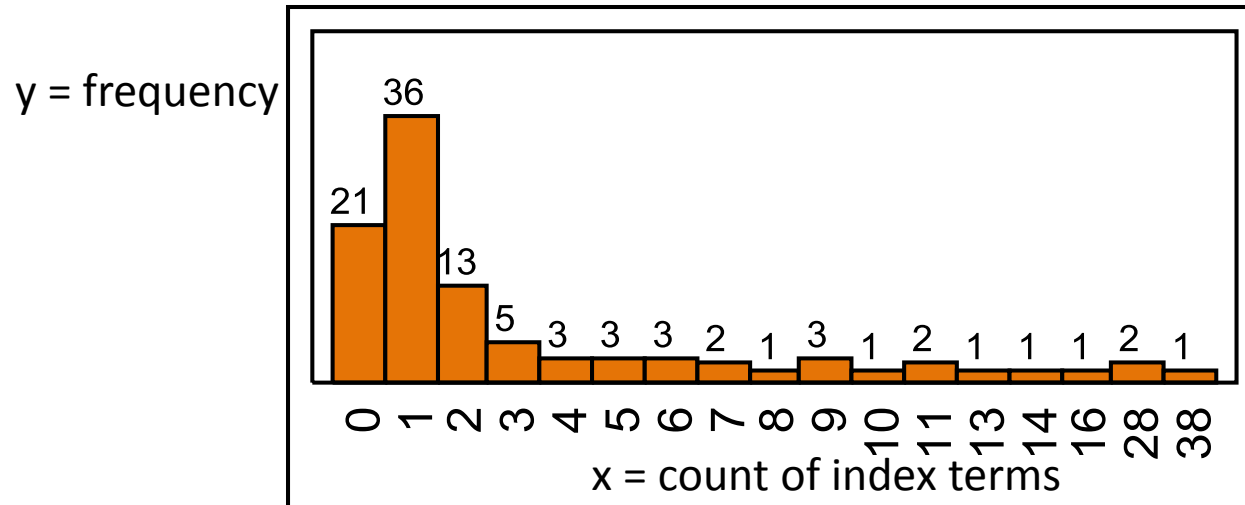
Ontology Concepts	Ontology SuperClasses	Ontology SubClasses	Filter by (keywords) AND (all SubClasses)
matrBiomass.owl#Biomass matrEnergy.owl#Biomass	energy storage living entity	animal waste detritus fertilizer food gelbstoff humus litter manure	forests ( 20 ) soils ( 36 ) europe ( 33 ) biomass ( 19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