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

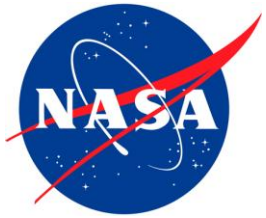
Line C. Pouchard,* Robert B. Cook,* Jim Green,*
Natasha Noy,** Giri Palanisamy*

Oak Ridge National Laboratory*

Stanford Center for Biomedical Informatics Research **

Presented to the Ontology Summit 2012

March 8, 2012

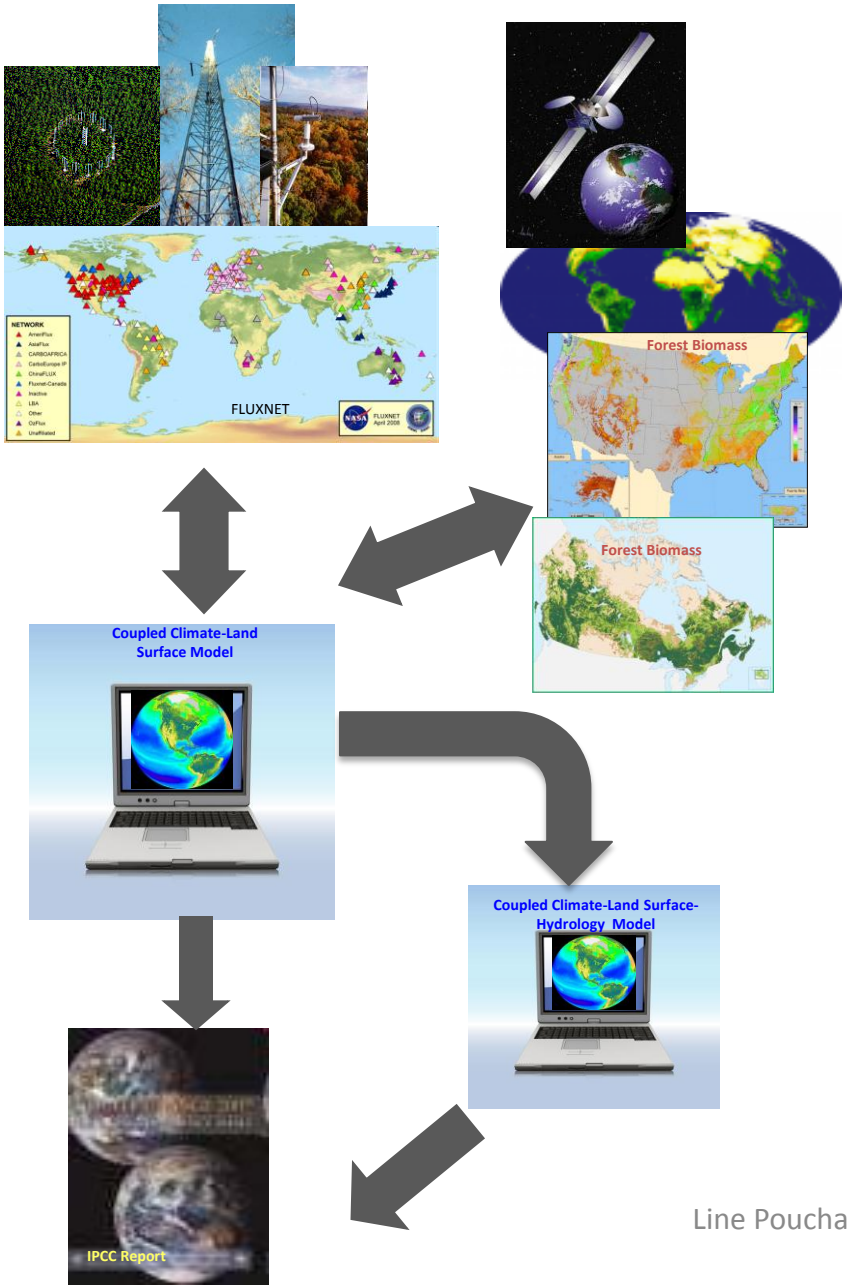


What is Linked Science?

- 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

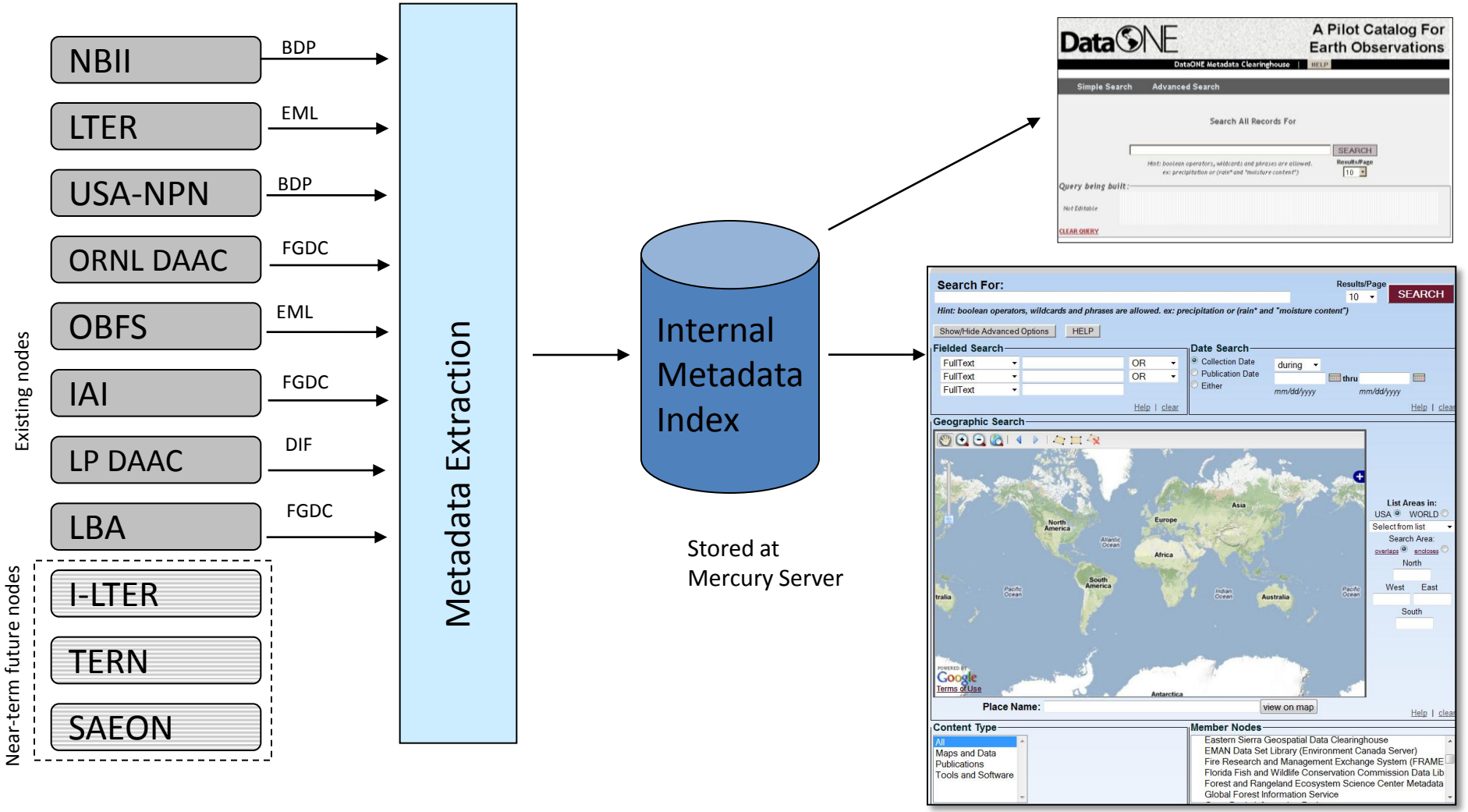
Line Pouchard, pouchardlc@ornl.gov

Linked Science in the Earth Sciences

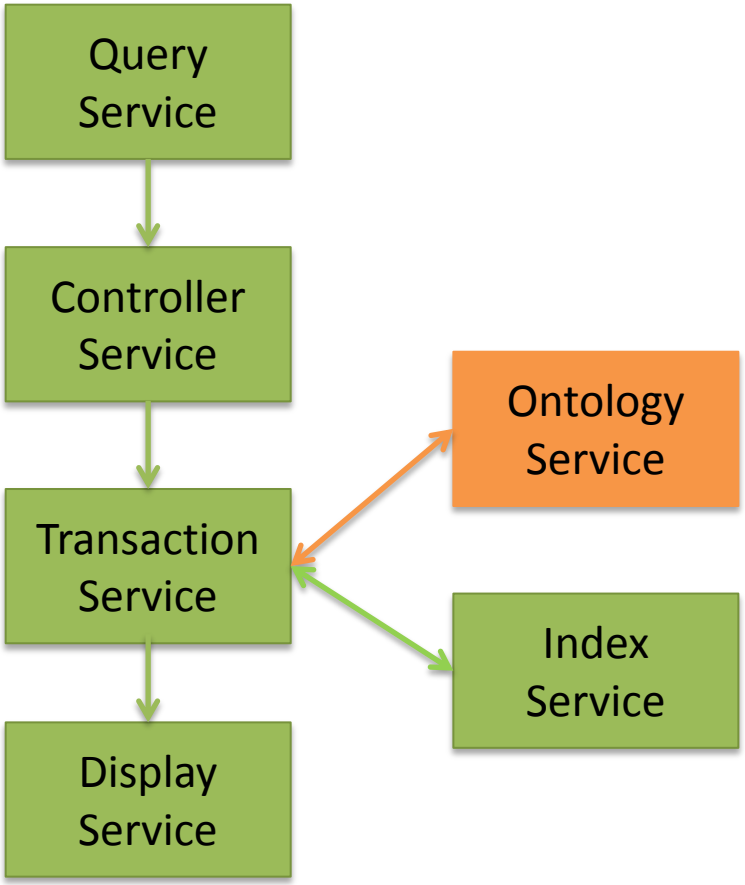


- Discovery and access from heterogeneous sources
  - Simulations, models, experiments, remote sensing, GIS, molecular and omics databases, publications
- Metadata and semantics integration
- Workflows, scenario development, data and process re-use, provenance
- Engaging communities of scientists, educators,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

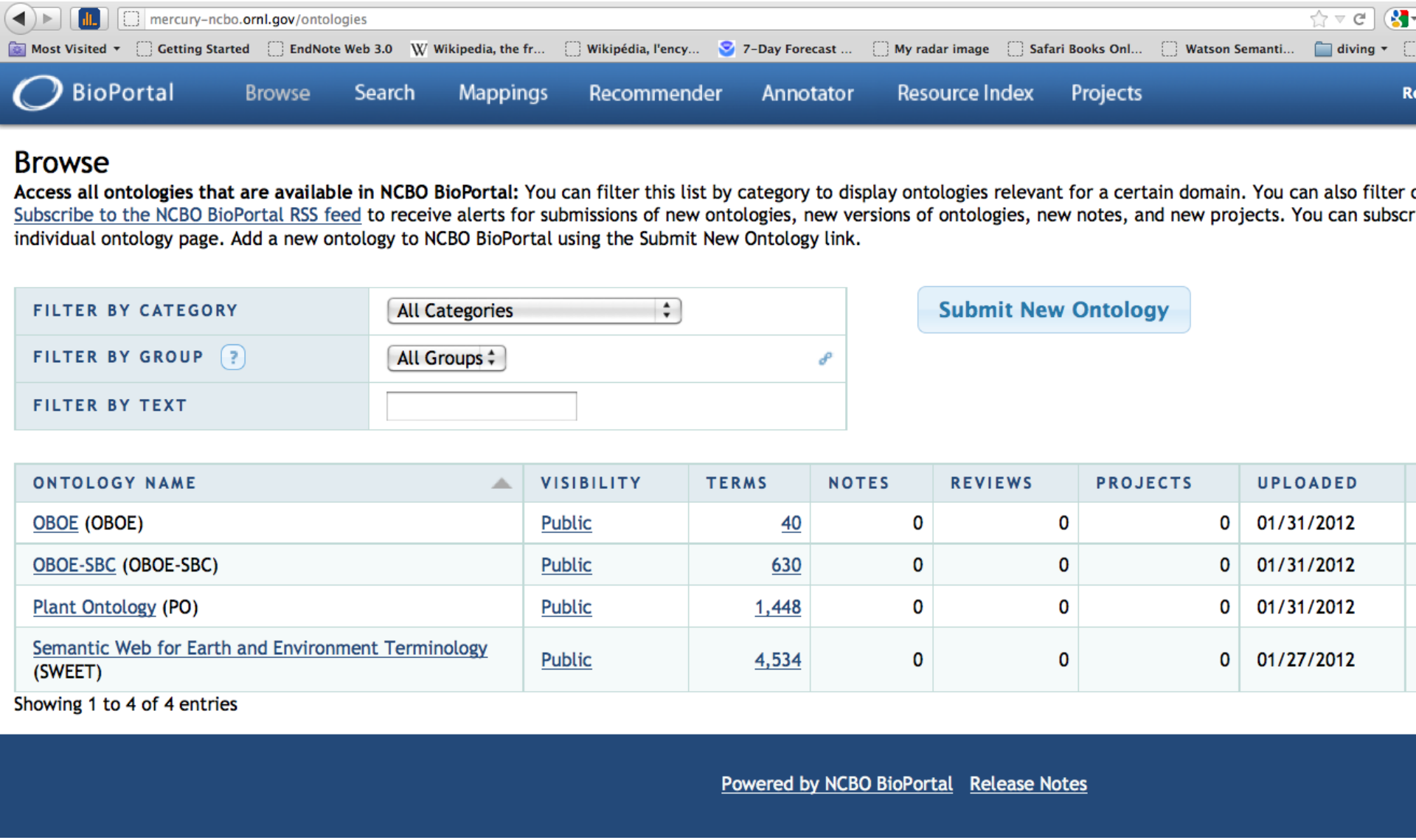


Coupling the Mercury federated metadata engine and BioPortal



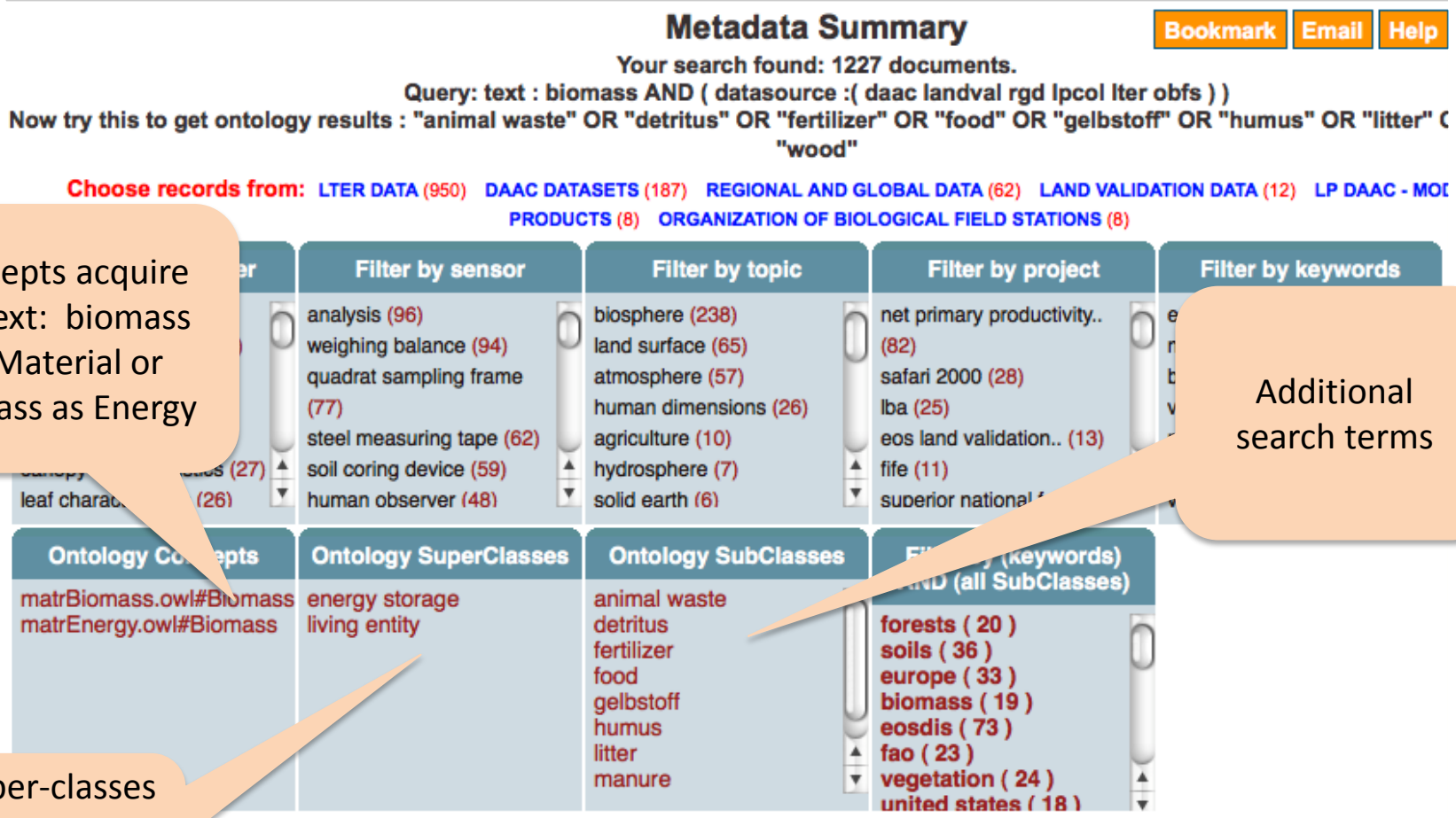
- Uses BioPortal Rest Services for programmatic access
- Returns ontology concepts, super- and sub-classes
- Provides additional keywords
- Provides context
- Uses these for new searches

BioPortal provides access to ontologies



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

Ontology-based search results

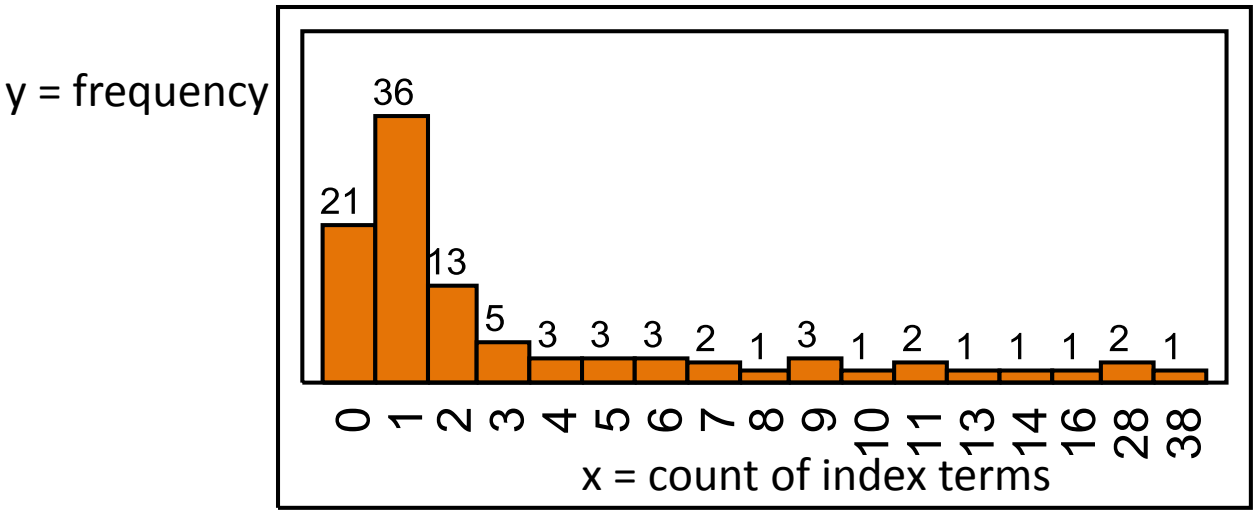


Concepts acquire context: biomass as Material or biomass as Energy

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