

# Working with a community-based organization to support an ontology infrastructure

Dr. Line Pouchard, Purdue University Libraries

Professor Michael Huhns, University of South Carolina

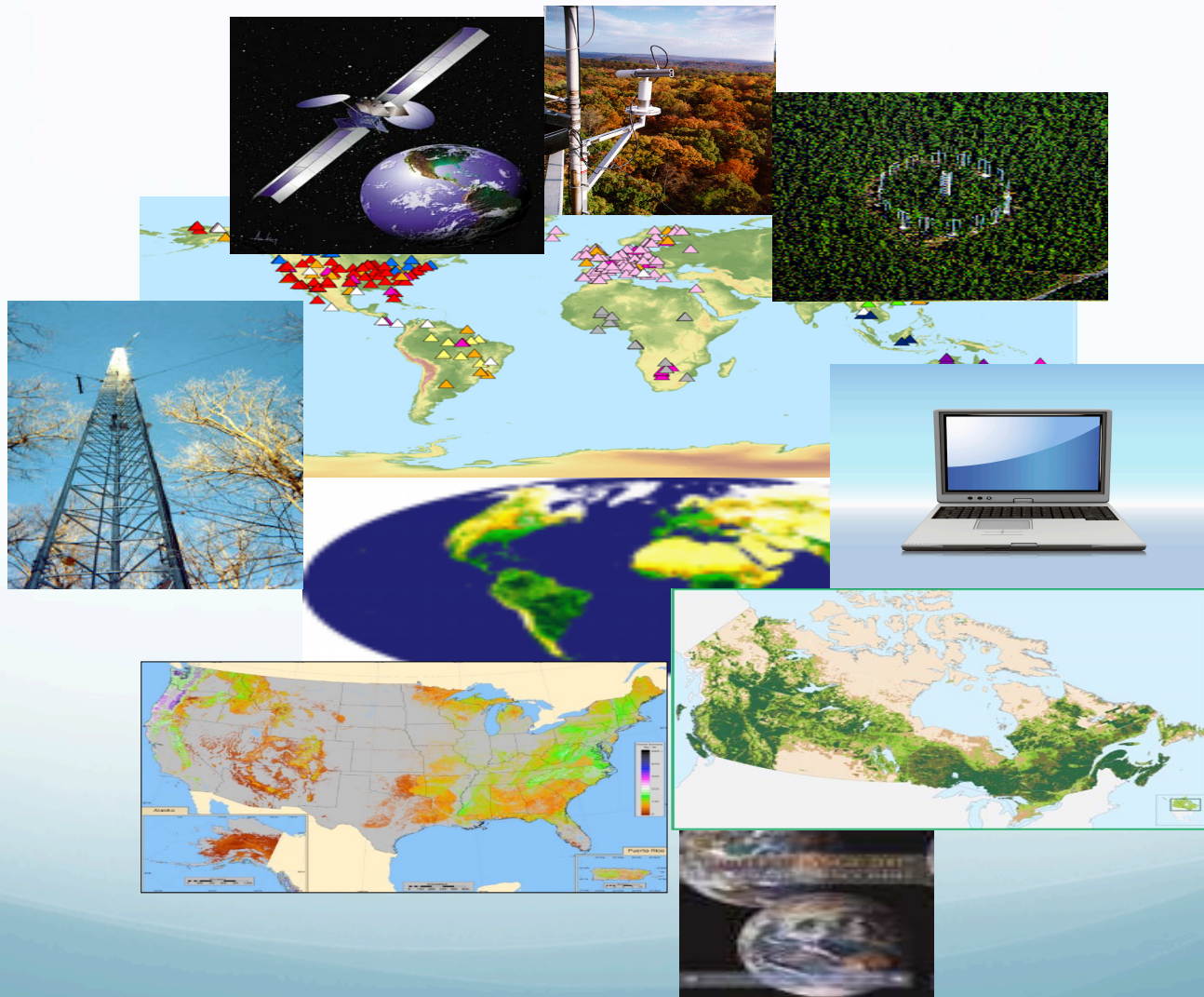
Dr. Thomas Narock, Marymount University

Erin Robinson, ESIP Executive Director

CNI Spring Member Meeting, San Antonio, TX, April 4-5, 2016



# A very diverse data and metadata ecosystem



# Data deluge in Earth Science

Sensors, sensor networks, and remote sensing gather observations;  
Data management and stewardship

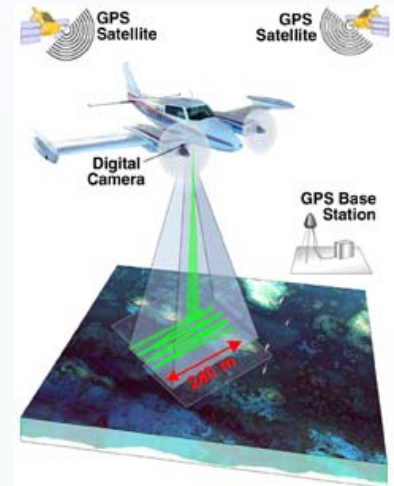
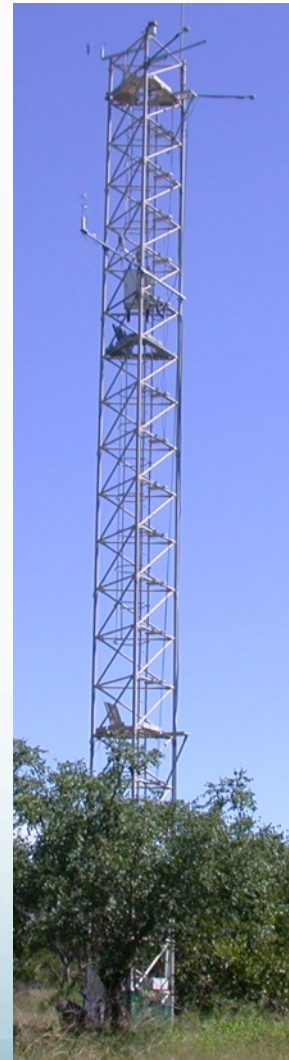
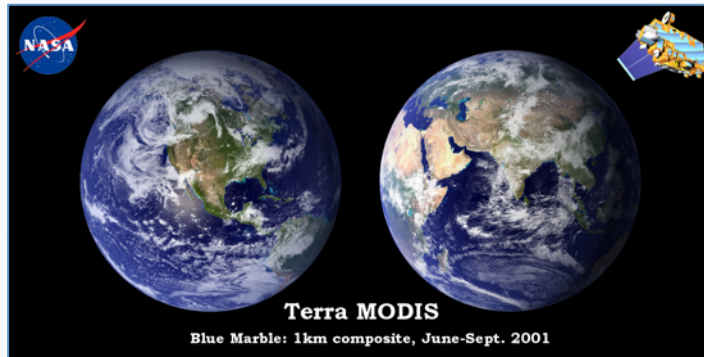


Photo courtesy of [www.carboafrika.net](http://www.carboafrika.net)

# Motivation: what we are trying to solve

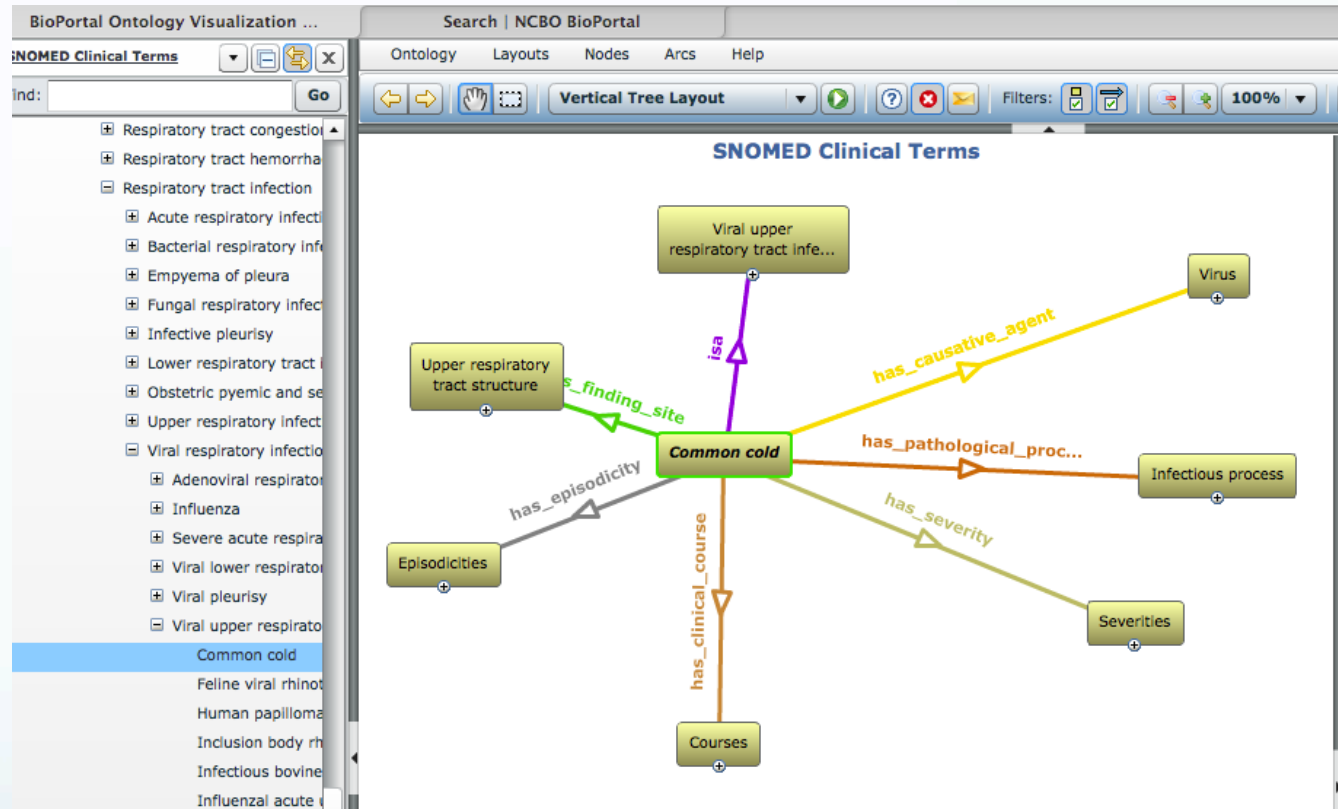
- Data held in data centers focused on broad data types or earth realms
  - Ocean, ice, hydrology, climate, etc.
- But phenomena of interest require data to be integrated from all these data types for meaningful analysis
  - Predicting flash floods
  - Visualizing hurricanes
  - Containing forest fires
  - Studying the North Pacific algae bloom
- Between various sources, data and metadata may overlap or be orthogonal

# Our proposed solution

- Provide ontologies to help with detailed annotations or descriptions of data elements and text
- Improve discoverability of ontologies to support information integration
  - Ontologies are currently living on project web sites
- Enhance the quality of data annotations
  - Evaluate ontology scope
  - Improve ontology re-use
  - Enable the re-use of minimal amount of concepts
  - Apply ontology design patterns

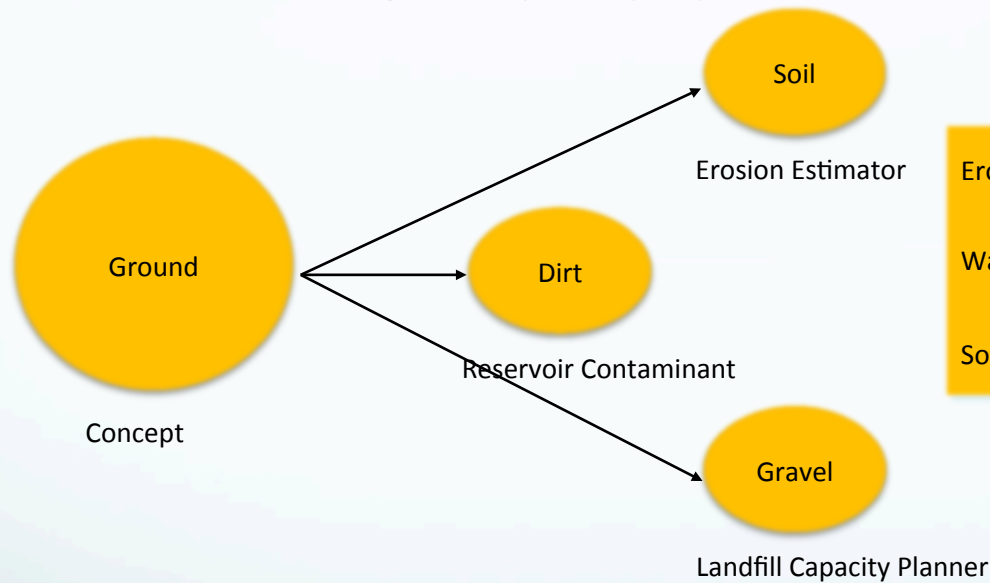
# What are ontologies?

- A formal specification of the basic concepts in a topic area, the relationships between these concepts, and the rules constraining the instances of a concept.
- A set of formal machine-readable definitions for the terms comprising a vocabulary
- The rules for combining terms and relations

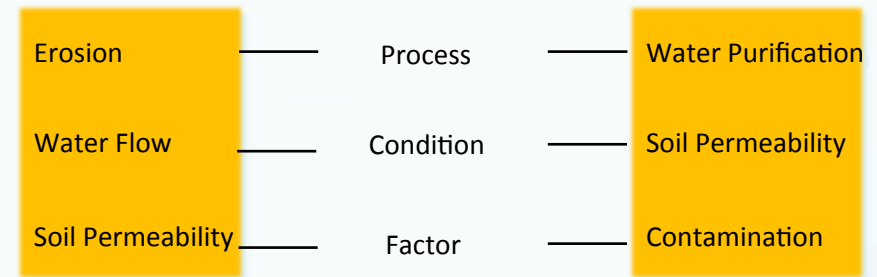


# Challenges to human communication and software interoperability

Semantic challenge 1: synonymy



Semantic challenge 2: polysemy and term ambiguity



**Most effort – and success – in relating human and system concepts**  
**Little effort – and success – in relating human and system processes**

# Ontologies are relevant to Big Data

- Massive-scale raw data must be highly structured to be useful to downstream users in the Earth Science domain
- Data produced at great effort and expense are only useful as researchers' ability to locate, integrate and access them
- This ability depends upon proper annotation of the data with tags. Ontologies are a good source of agreed-upon tags for specific domains
- The creation and curation of ontologies in any scientific domain is resource-intensive and may be best achieved through community curation and crowd-sourcing

"Big Data: the future of biocuration." Doug Howe et al. *Nature* 455, 47-50 (2008)

"Omics data sharing." Dawn Field et al. *Science* 326(5950), 234-236(2009)

"Motivating online publication of data." Mark Costello. *Bioscience* 59(5) (2009)



# Our plan

- Deploy an ontology repository
  - Provide a centralized location for earth science ontologies
- Create mappings between ontologies in the repository
- Place the project under the auspices of the ESIP federation

# The ESIP federation

<http://esipfed.org>



FRONTIERS IN EARTH SCIENCES

# BIG DATA

Summer Meeting  
Chapel Hill, N.C.  
19-22 July 2016

# ESIP community



Multiple Science  
Domains



Corporate

users

data providers

application  
developers

data archives

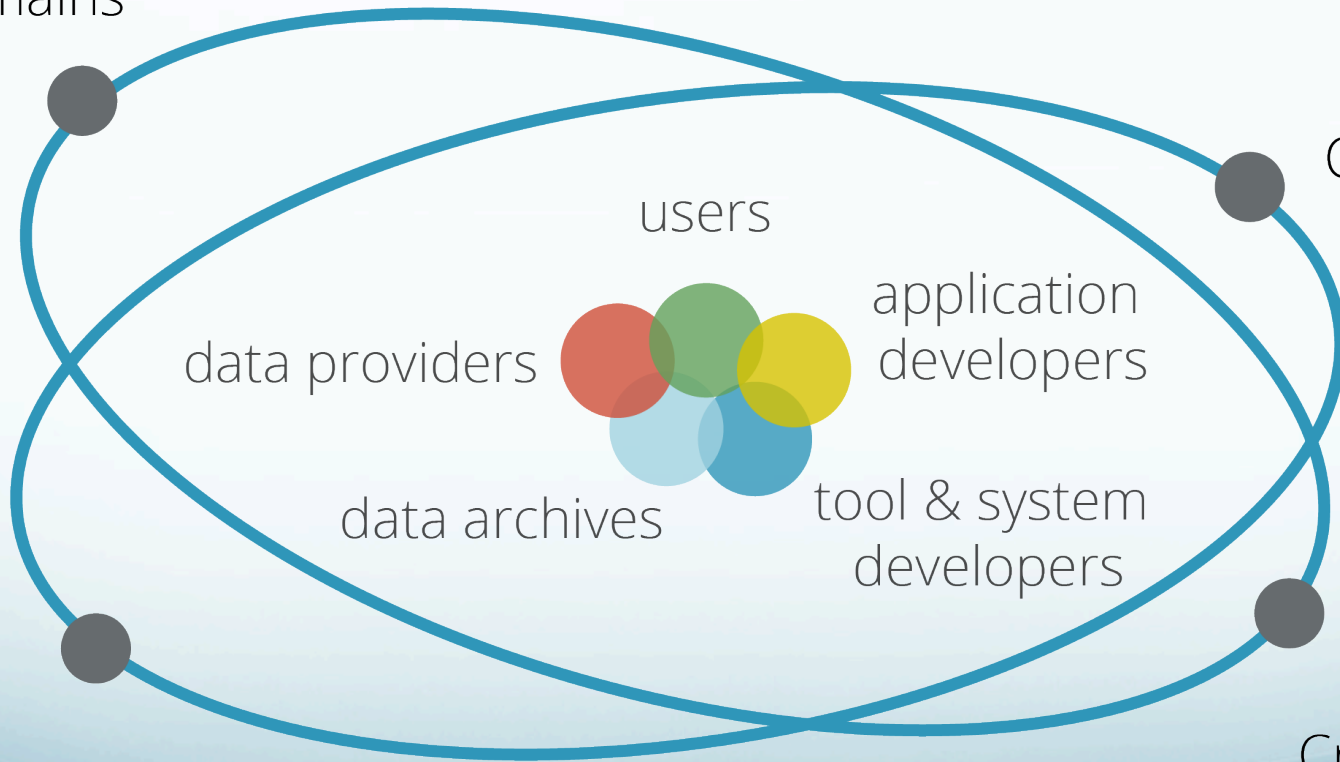
tool & system  
developers



Academia



Cross-  
Agency



# Collaboration Area Structure

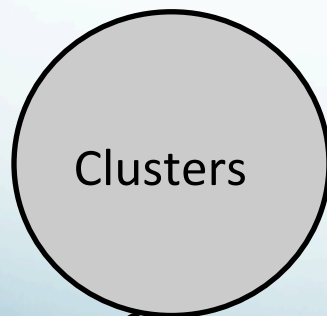
## Features:

- Forms by sending VP an email
- Ends when last person hangs up

<http://esipfed.org/collaboration-areas>

## ESIP Clusters:

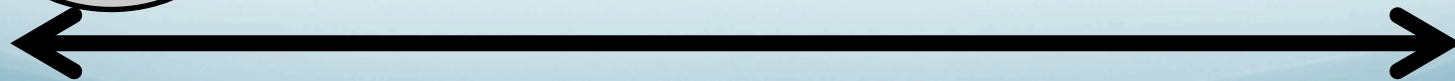
- Agriculture and Climate
- Cloud Computing
- Discovery
- Disaster Lifecycle
- Documentation
- Drones
- Drupal
- Earth Science Data Analytics
- Envirosensing
- Information Quality
- Geospatial
- Science Software
- Sustainable Data Management
- Web Services



Least

Formality

Most



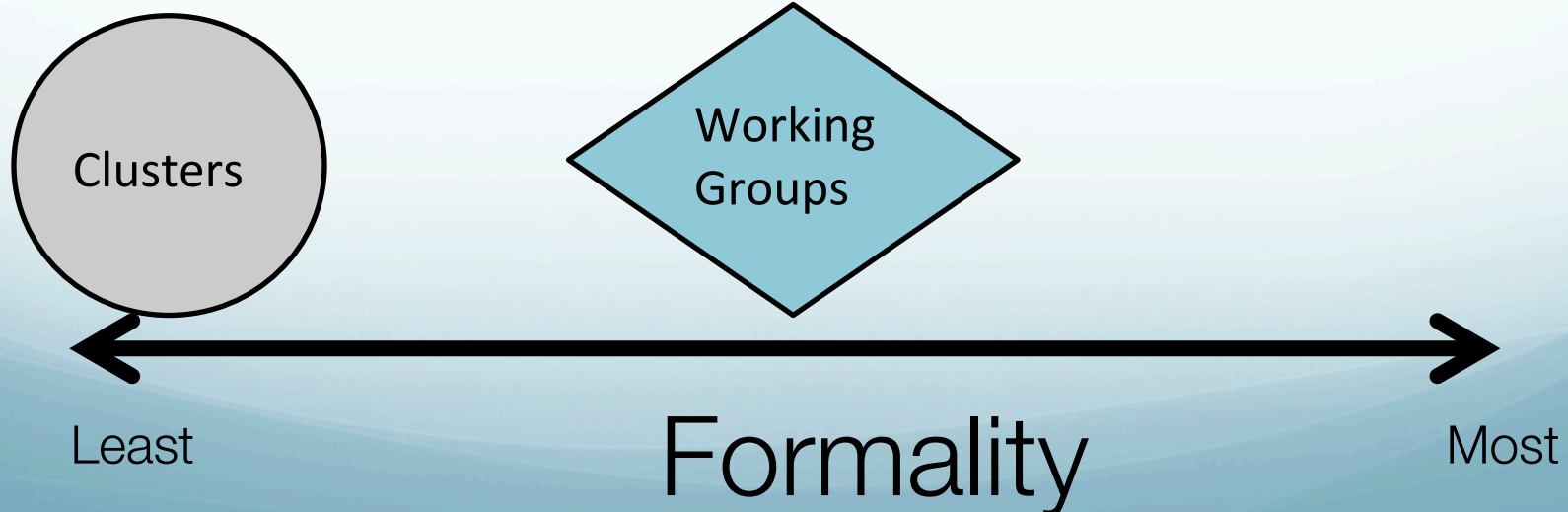
# Collaboration Area Structure

## Features:

- Created by Assembly or Committee
- Task-oriented
- Budget\*

## ESIP Working Groups:

- Data Study
- Drupal
- Climate Education
- Visioneers



# Collaboration Area Structure

## Features:

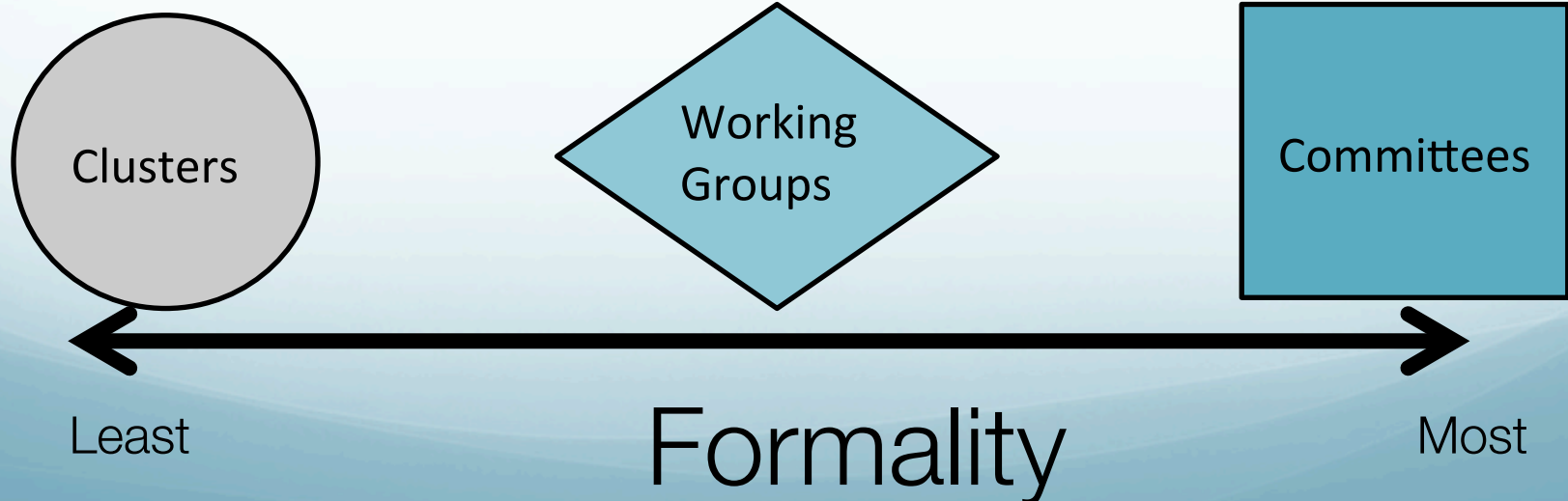
- Created by Assembly or Committee
- Task-oriented
- Budget\*

## Administrative Committees:

- Constitution and Bylaws
- Finance and Appropriations
- Partnership

## Standing Committees:

- Data Stewardship
- Education
- Information Technology and Interoperability
- Products and Services
- Semantic Technologies



# Governance: types of ESIP members

## Type I: Data Centers

- NASA DAACs
- NOAA (NGDC, NODC, NCDC)

## Type II: Researchers and Tool Developers

- Academia
- Government labs

## ESIP Assembly:

- One partner, one vote
- Annual business meeting at ESIP Winter Meeting
- Leadership elected from Assembly representatives

## Type III: Application Developers

- Commercial
- Nonprofit
- Educational

## Type IV: Strategic Partners

- NASA
- NOAA

# ESIP Groups

## Standing Committees

- Data Stewardship
- Education
- Information Technology and Interoperability
- Products and Services
- Semantic Web Technologies

## Administrative Committees

- Constitution and Bylaws
- Finance and Appropriations
- Partnership
- Nominations

## Working groups

- Visioneers
- Climate Education
- Energy & Climate

## Clusters

- Agriculture and Climate
- Cloud Computing
- Discovery
- Disaster Lifecycle
- Documentation
- Drones
- Drupal
- Earth Science Collaboratory
- Envirosensing
- Information Quality
- Geospatial
- Science Software
- Sustainable Data Management
- Web Services

<http://esipfed.org/collaboration-areas>



# Things ESIP Does

- Community-generated best practices
- Testbed funds several proposals per year
- Professional development
- Outreach
- Provide virtual and in-person venues for collaboration and connections
- Provide mini-grants to make stuff happen
- Student fellows
- Winter and Summer meetings – everyone welcome



# ESIP Online



**Welcome to the ESIP Federation**

The Federation of Earth Science Information Partners (ESIP) is an open, networked community that brings together science, data and information technology practitioners. Participation in the ESIP Federation is beneficial because it provides an intellectual commons to expose, gather and enhance in-house capabilities in support of an organization's own mandate.

[Read More](#)

**Agenda**

**Recent Updates**

**Quick Links**

<http://esipfed.org>  
<http://wiki.esipfed.org>  
<http://commons.esipfed.org>  
<http://testbed.esipfed.org>  
<http://tinyurl.com/esip-facebook>



**Welcome to the ESIP Commons**

The ESIP Commons is a knowledge repository created by members of the ESIP community. Content on this site is licensed under Creative Commons and freely available to the public for use, sharing, repurposing and remixing.

**Type**

- Author (883)
- Meeting Session (498)
- Poster (334)
- Feed Item (131)
- Blog entry (70)
- Esip Story (News) (68)
- Esip Basic Page (46)
- Book page (41)
- Education Modules (35)
- Encyclopedia (26)

[Show more](#)

**Collaboration Area**

- Data Preservation (112)

**Organization**

- NASA (202)

**15 YEARS**  
MAKING DATA MATTER

TWEETS: 3,736 | FOLLOWING: 874 | FOLLOWERS: 857 | FAVORITES: 677 | LISTS: 17



**ESIP Federation** @ESIPFed

ESIP Federation is a community [gov/acad/ind] that produces & interprets Earth science data & develops applications for data use. [Administered by ER, AB, RF]

Virtual [esipfed.org](#)  
 Joined November 2010

152 Photos and videos Photos and Videos

**Tweets** | Tweets & replies | Photos & videos

ESIP Federation retweeted **FES** @EarthSciBound · 7h  
 Last day for Early ESIPFed Summer Meeting Registration-epurt.com/bqQOD!

ESIP Federation retweeted **ESIP Federation** @ESIPFed · Jun 30  
 2-week countdown! ...on.s3-website-us-west-2.amazonaws.com/#CleanTheWiki#ESIPSummerMeeting

ESIP Federation @ESIPFed · Jun 29  
 Please finalize all content for the #ESIPFed Summer Meeting by Wed, July 1st. Check out the agenda + posters here: bit.ly/1Fw4A26

ESIP Federation retweeted **CIESIN** @ciesin · Jun 29  
 Job opening @CIESIN: remotensing #geospatial processing #R, pubs and proj mgt bkg bit.ly/1GBX3Y

ESIP Federation @ESIPFed · Jun 26  
 Travel funds available for the Conceptual Design #EarthCube Workshop co-located w/ our summer meeting. Apply today! bit.ly/1TRK12T



**SUMMER 2015**  
 The ESIP Federation & Community Resilience: Coming Together

**15 YEARS**  
MAKING DATA MATTER

Federation of Earth Science Information Partners (ESIP) Community

[Create Call to Action](#) [Share](#)

**Timeline** | About | Photos | Monday Update Signup | More

179 likes +3 this week  
 141 post reach this week

**Promote Your Page**  
 Connect with more of the people who matter to you  
[Promote Page](#)

**ABOUT**

The Federation of Earth Science Information Partners (ESIP) is a community comprising government, academic & industry partners that produce & interpret...  
[READ MORE](#)

<http://esipfed.org/> [Promote Website](#)

**APPS**

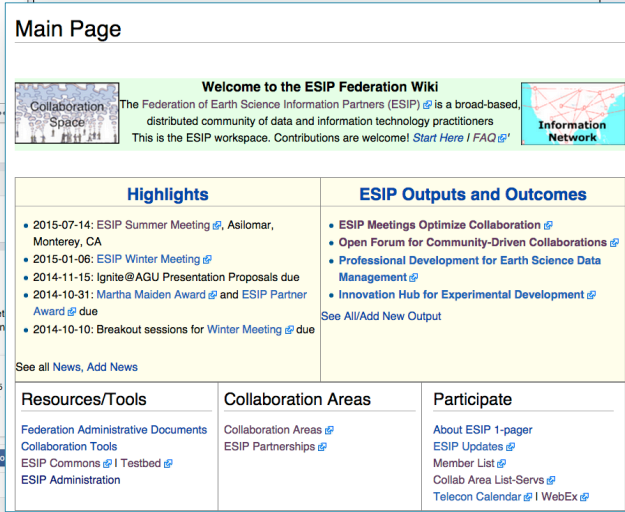
**Monday Update Signup**

44 people reached [Boost Post](#)

Like · Comment · Share

Chris Mattmann and Wenyu Gong like this.

Write a comment...



**Main Page**

**Welcome to the ESIP Federation Wiki**

The Federation of Earth Science Information Partners (ESIP) is a broad-based, distributed community of data and information technology practitioners. This is the ESIP workspace. Contributions are welcome! [Start Here](#) | [FAQ](#) | [G](#)

**Highlights**

- 2015-07-14: ESIP Summer Meeting @, Asilomar, Monterey, CA
- 2015-01-06: ESIP Winter Meeting @
- 2014-11-15: Ignite@AGU Presentation Proposals due
- 2014-10-31: Martha Maiden Award @ and ESIP Partner Award @ due
- 2014-10-10: Breakout sessions for Winter Meeting @ due

[See All/Add New Update](#)

**ESIP Outputs and Outcomes**

- ESIP Meetings Optimize Collaboration @
- Open Forum for Community-Driven Collaborations @
- Professional Development for Earth Science Data Management @
- Innovation Hub for Experimental Development @

[See All/Add New Update](#)

**Resources/Tools**

- Federation Administrative Documents
- Collaboration Tools
- ESIP Commons @ | Testbed @
- ESIP Administration

**Collaboration Areas**

- Collaboration Areas @
- ESIP Partnerships @

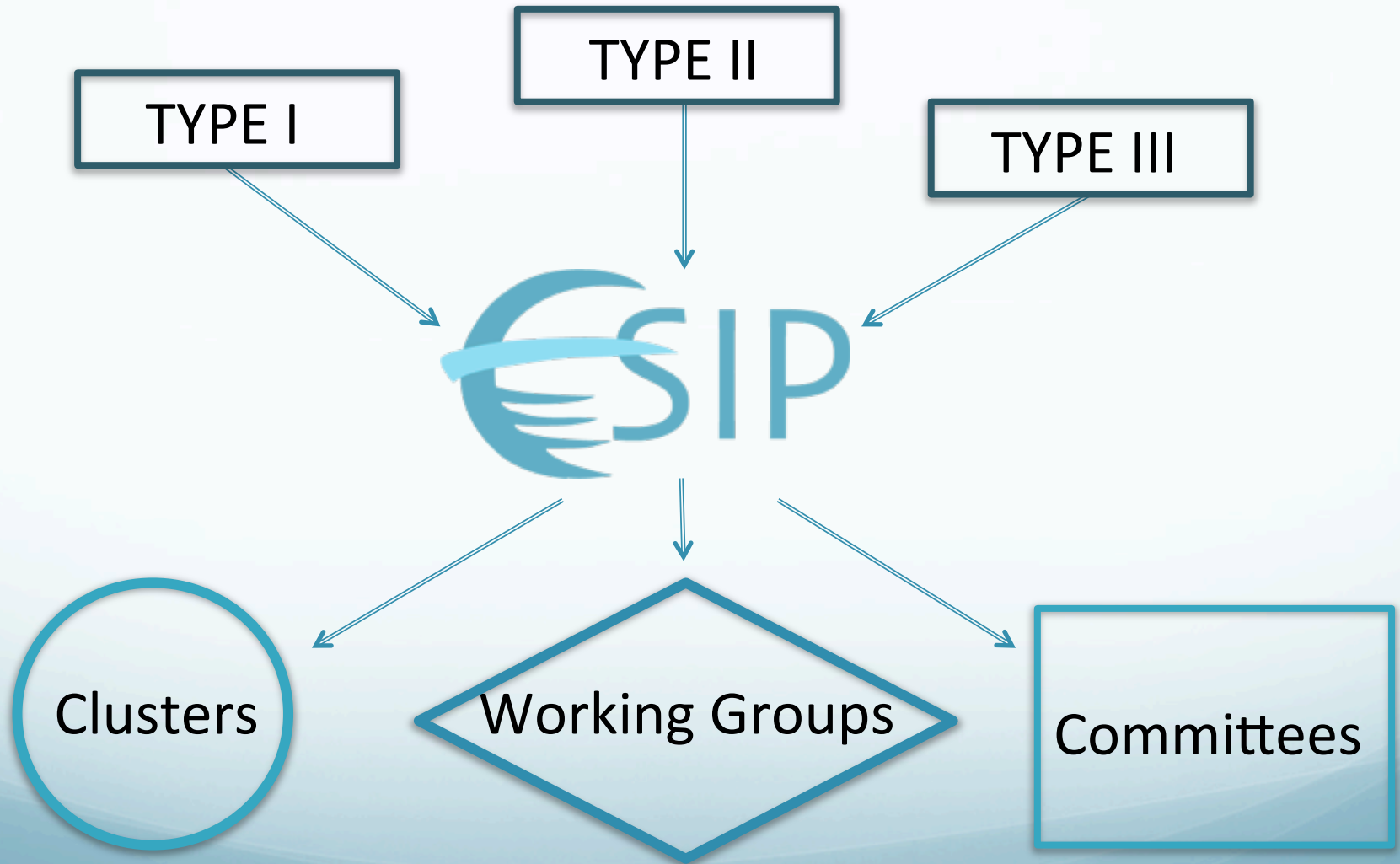
**Participate**

- About ESIP 1-pager
- ESIP Updates @
- Member List @
- Collab Area List-Servs @
- Telecon Calendar @ | WebEx @

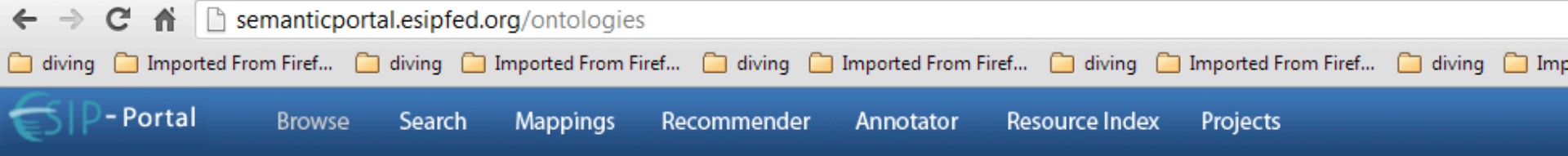


#esipfed  
 @ESIPFed

# COLLABORATION AREA STRUCTURE



# The ESIP Semantic Portal



## Browse

Access all ontologies that are available in the ESIP-Portal: You can filter ontologies that belong to a certain group. Add a new ontology to the ESIP-Portal using the

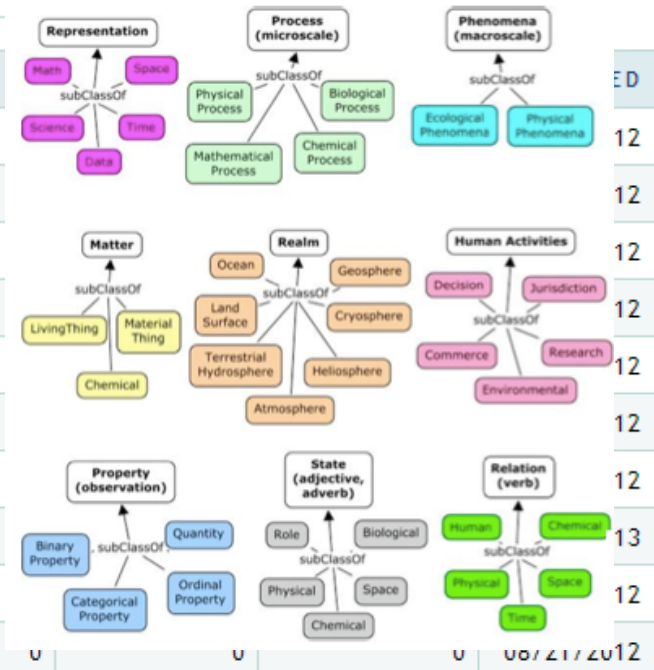
**New:** [Configure](#) which ontologies you see in ESIP-Portal

**FILTER BY GROUP** ?  ⌵ ⌵

**FILTER BY TEXT**

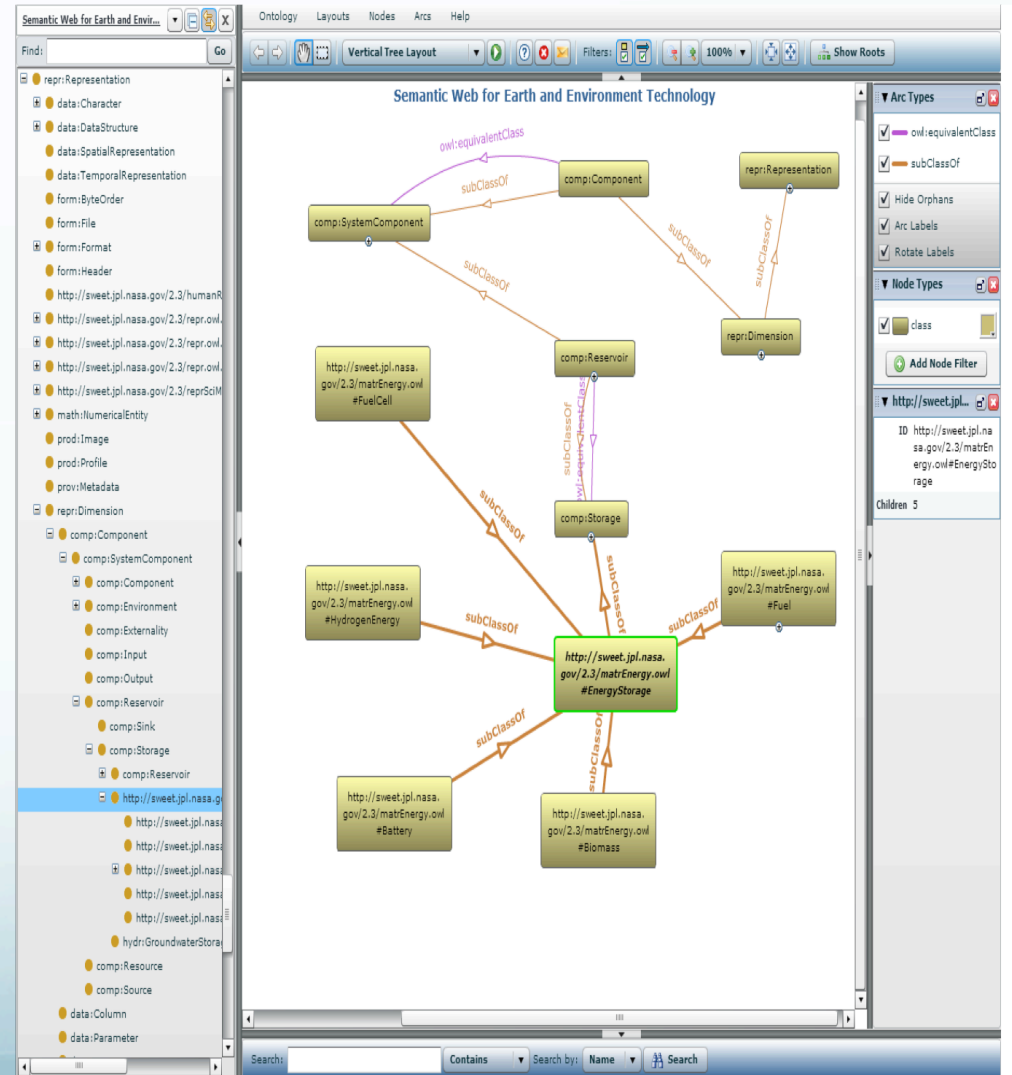
[Submit New Ontology](#)

ONTOLOGY NAME	VISIBILITY	TERMS	NOTES
<a href="#">Basic Formal Ontology (BFO)</a>	<a href="#">Public</a>	<a href="#">39</a>	
<a href="#">Contour Map WDO (CoM-WDO)</a>	<a href="#">Public</a>	<a href="#">65</a>	
<a href="#">Crustal Modeling WDO (CrM-WDO)</a>	<a href="#">Public</a>	<a href="#">84</a>	
<a href="#">Data Fusion project WDO (DF-WDO)</a>	<a href="#">Public</a>	<a href="#">45</a>	
<a href="#">Eddy Covariance WDO (EC-WDO)</a>	<a href="#">Public</a>	<a href="#">64</a>	
<a href="#">ESIP Data Ontology (ESIPData)</a>	<a href="#">Public</a>	<a href="#">108</a>	
<a href="#">ESIP Service Ontology (ESIPService)</a>	<a href="#">Public</a>	<a href="#">13</a>	
<a href="#">Essential Climate Variables Ontology (ECV)</a>	<a href="#">Public</a>	<a href="#">2</a>	
<a href="#">Hole's Code WDO (HC-WDO)</a>	<a href="#">Public</a>	<a href="#">55</a>	
<a href="#">ISDC ontology (gfz)</a>	<a href="#">Public</a>	<a href="#">185</a>	



# Advantages of the semantic portal

- Programmatic access to ontologies via REST services
- Visualization of ontology entities in their hierarchical context
- Visualization of ontology entities in a graph
- Possibility to enter and edit annotations
- Automatic upload of mappings



# Evaluating the ESIP ontologies

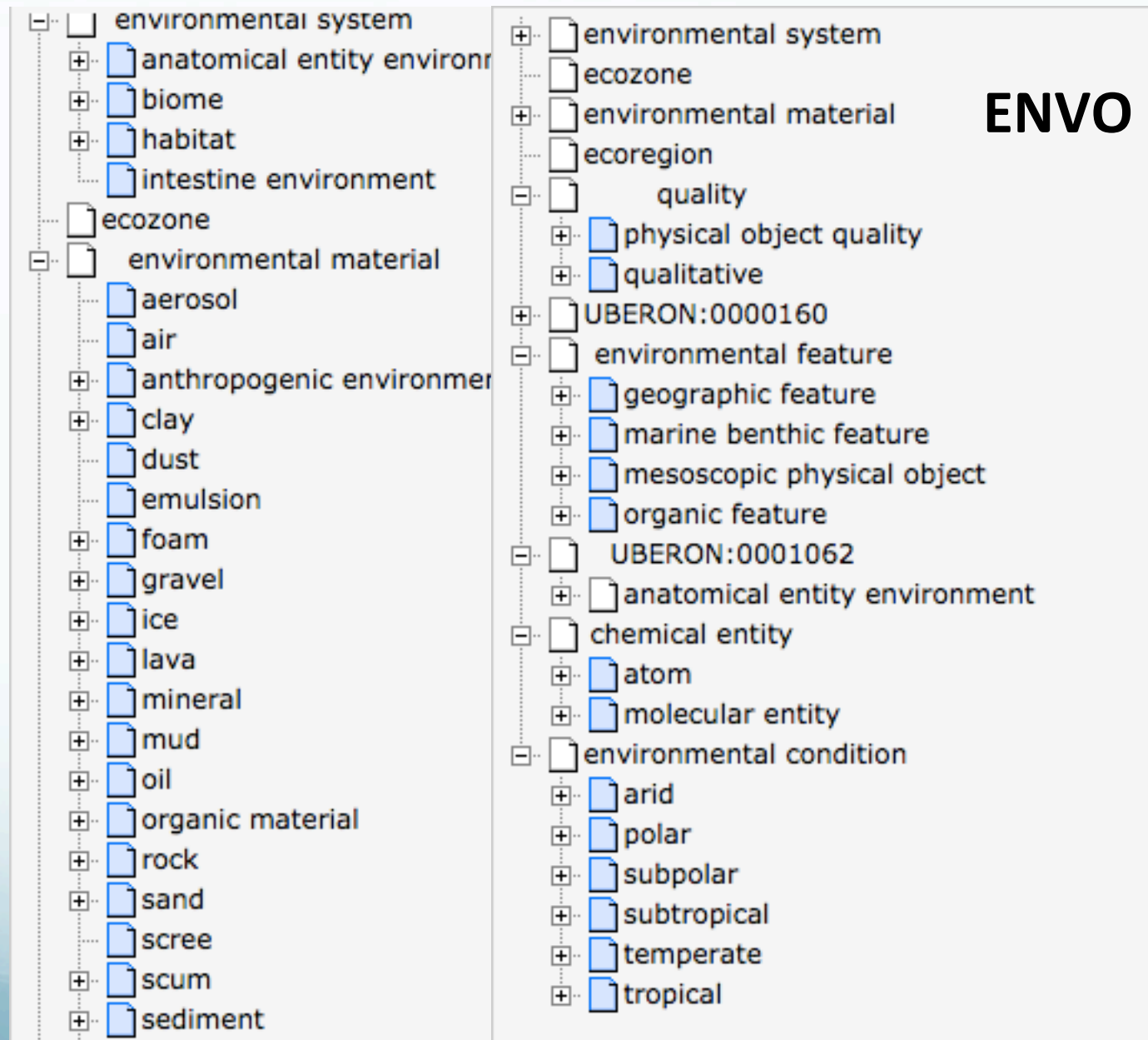
- Very different degrees of quality and descriptions
- One way to improve the degree of quality is to locate terms with similar semantics between two ontologies
- One purpose of the evaluation was to provide backend mappings between entities for semantically similar terms from different ontologies for the same domain so that users can then annotate the terms, add relationships, re-use patterns.

# Agreement Maker Light (AML)

- We use Agreement Maker Light, an algorithm from the Ontology Alignment Evaluation Initiative.
  - Citation: Ontology Maker Light Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I. F., & Couto, F. M. (2013, January). The agreementmakerlight ontology matching system. In *On the Move to Meaningful Internet Systems: OTM 2013 Conferences* (pp. 527-541). Springer Berlin Heidelberg.
- AML uses synonym properties between labels as determined by the TF-IDF measure
- Ranked second in the OAIE annual campaign, 2013
  - Ranked first in 8 ontologies for OAIE 2014 campaign

# Ontologies to map

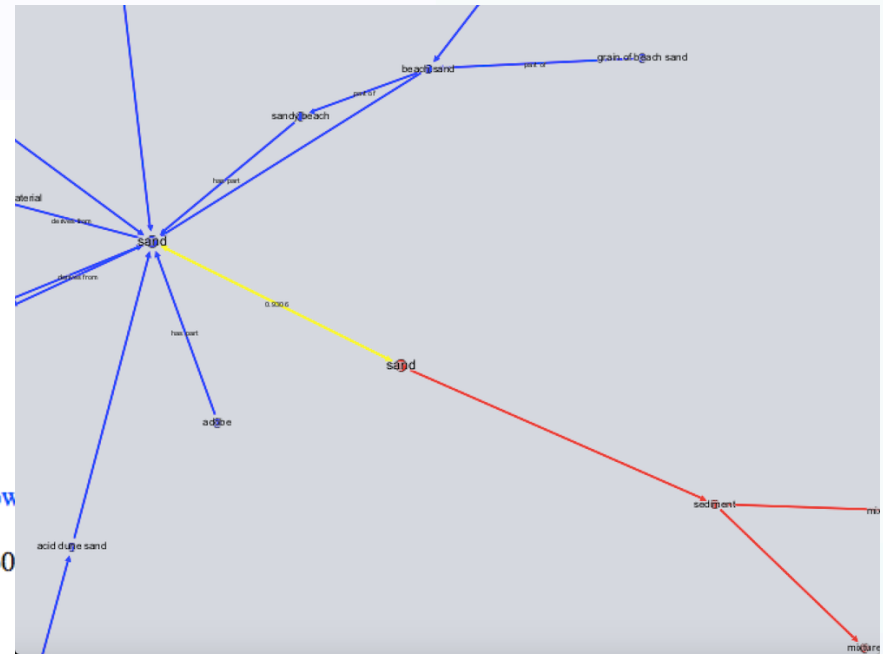
- SWEET:  
Summer  
2014  
version –  
4549  
classes, 360  
properties
- ENVO:  
Environmental  
Ontology –  
over 4700  
classes, 60  
properties





# Results of the mapping exercise

```
- <rdf:RDF alignmentSource="AgreementMakerLight">
- <Alignment>
  <xml>yes</xml>
  <level>0</level>
  <type>??</type>
  <onto1>http://sweet.jpl.nasa.gov/2.3/sweetAll.owl</onto1>
  <onto2>http://purl.obolibrary.org/obo/envo.owl</onto2>
  <uri1>http://sweet.jpl.nasa.gov/2.3/sweetAll.owl</uri1>
  <uri2>http://purl.obolibrary.org/obo/envo.owl</uri2>
- <map>
  - <Cell>
    <entity1 rdf:resource="http://sweet.jpl.nasa.gov/2.3/realmLandOrographic.owl#sand"/>
    <entity2 rdf:resource="http://purl.obolibrary.org/obo/ENVO_00000087"/>
    <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.930</measure>
    <relation>=</relation>
  </Cell>
</map>
- <map>
  - <Cell>
    <entity1 rdf:resource="http://sweet.jpl.nasa.gov/2.3/realmOceanFloor.owl#OceanFloor"/>
    <entity2 rdf:resource="http://purl.obolibrary.org/obo/ENVO_00000426"/>
    <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.9306</measure>
    <relation>=</relation>
  </Cell>
</map>
```



Alignment produced by ALM light: 841 mappings

Term mapping produced by LOOM in the portal instance: 242

# Semantic Sea Ice Interoperability Initiative (SSII) ontologies

Filename		
<a href="#">README.txt</a>	SWEET-to-EGG:	18
<a href="#">Sea-Ice-Chart-Regions.rdf</a>	EGG-to-SWEET:	18
<a href="#">annotations.owl</a>		
<a href="#">egg.omn</a>	SWEET-to-Iceoflandorigin:	5
<a href="#">egg.owl</a>	Iceoflandorigin-to-SWEET:	5
<a href="#">ice-of-land-origin.omn</a>		
<a href="#">ice-of-land-origin.owl</a>	SWEET-to-seaiceconcentration:	12
<a href="#">seaice-concentration.omn</a>	seaiceconcentration-to-SWEET:	12
<a href="#">seaice-concentration.owl</a>		
<a href="#">seaice-development.omn</a>	SWEET-to-seaicedevelopment:	1
<a href="#">seaice-development.owl</a>	seaicedevelopment-to-SWEET:	1
<a href="#">seaice-form.omn</a>		
<a href="#">seaice-form.owl</a>	SWEET-to-seaiceform:	9
<a href="#">seaice.omn</a>	seaiceform-to-SWEET:	9
<a href="#">seaice.owl</a>		
<a href="#">sigrid3.omn</a>	SWEET-to-seaice:	15
<a href="#">sigrid3.owl</a>	seaice-to-SWEET:	15
<a href="#">sweet_cryo.owl</a>		
	SWEET-to-sig3:	20
	Sig3-to-SWEET:	20

# ESIP support of the repository

- Through the Semantic Technologies committee
  - Original impulse
  - The audience reached fits the purpose
    - Wide audience of Earth scientists and technologists with the need for ontologies or who create ontologies in multi-disciplinary research
  - Collect user feedback, wish list, guidance
  - Proposals to ESIP for designing a governance model
  - Technology evaluation
- Through the Products and Services committee
  - Initial funding
  - Technology evaluation framework

# Sustainability

- ESIP's AWS cloud micro-instance not sufficient to support the code base
- Currently on departmental server at USC
- Applying to ESIP testbed for infrastructure support
  - Ontology Repositories support a wide range of features
    - Seeking Balance - advanced features require more support (hardware and admin support)
    - Currently evaluating the necessary set of features for this community
  - Governance in discussion
    - Governance refers to both the portal and the ontologies
    - What are the long-term implications of supporting a portal (hardware/software/admin) for ESIP?
    - How to effectively manage ontology changes in 300+ person organization
- Several sustainability studies currently under way within ESIP

# Conclusions

- Our solution supports the integration of data and metadata by providing detailed annotations to data elements using ontology entities
- We provide mappings between the ontology entities that support extension of the ontologies and an estimate of the difference in concepts between ontologies
- Our organizational structure ensures that we reach the intended users and can collect feedback
- We are working with the ESIP foundation to establish a sustainability and governance plan

**Thank you!**