



# Open Knowledge Networks for Geosciences, Sustainability, and Convergent Research in Natural-Human Systems

**Yolanda Gil**

Information Sciences Institute  
and Department of Computer Science  
University of Southern California

<http://www.isi.edu/~gil>

@yolandagil  
gil@isi.edu

4 October 2017

# Outline

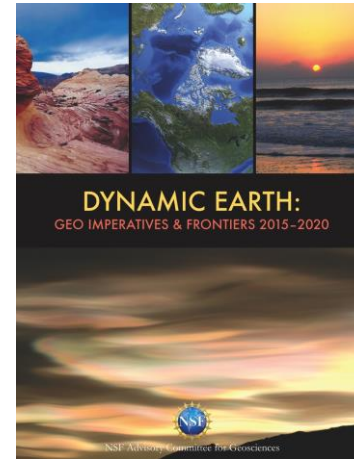
---

1. The geosciences landscape
2. Ontologies, vocabularies, standards
3. Infrastructure
  - Data centers
  - Tools
4. Modeling
  - Model repositories
5. Model integration and convergence research
  - The need for semantics

# Geosciences

## ■ The Earth as a **system**

- Earth (surface and subsurface)
- Ocean
- Atmospheric
- Polar
- Geospace



Land/Ocean Processes

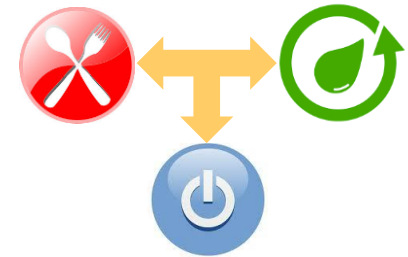
Ocean-Atmosphere-Ice  
Ecosystems

Geo-Bio-Chem-Phys-  
Human Processes in  
Ecosystems

Urban Geosystem  
Science

## ■ Natural processes/resources interact with human activities

- Water
- Food: Agriculture, fisheries
- Energy: Manufacturing, infrastructure



## ■ Ecosystems and sustainability

Introduced in Senate (05/06/2015)

Nexus of Energy and Water for Sustainability Act of 2015 or the NEWS Act of 2015

# Geosciences Landscape

## ■ Funding agencies



## ■ Foundations

- Sloan



## ■ Community organizations

- Earth Science Information Partners (ESIP)



## ■ Scientific organizations and publishers

- AGU, GSA, ASLO, CEDAR



## ■ Standards organizations

- Open Geospatial Consortium (OGC)



## ■ Industry

- Esri (ArcGIS)
- Oil and gas
- Mapping (Google, Microsoft)

## ■ Many groups outside US



# Outline

---

1. The geosciences landscape
2. Ontologies, vocabularies, standards
3. Infrastructure
  - Data centers
  - Tools
4. Modeling
  - Model repositories
5. Model integration and convergence research
  - The need for semantics

# Ontologies, Vocabularies, and Standards

---

- General geosciences vocabularies
  - SWEET (Semantic Web for Earth and Environmental Terminology)
  - EML (Ecological Metadata Language)
  - ENVO (Environmental Ontology) in BioPortal & OBO
- Space and time
  - W3C Space and Time Ontology (builds on GeoSPARQL, KML,...)
  - Open Geospatial Consortium standards (eg, SensorML)
  - ISO 19115 (geospatial data)
- Maps:
  - Gazetteers (e.g., Geonames), USGS Geographic Names Information System (GNIS), NGA GEOnet Names Server, etc.
- Specialized ontologies:
  - WaterML, CF (Climate and Forecast) conventions, land cover,...

# Outline

---

1. The geosciences landscape
2. Ontologies, vocabularies, standards
3. Infrastructure
  - Data centers
  - Tools
4. Modeling
  - Model repositories
5. Model integration and convergence research
  - The need for semantics

# Infrastructure: Data Centers

---

- Federal and state level
  - NASA DAACs: Remote sensing data
  - USGS and state geological surveys
- General repositories
  - Pangea
  - IGSN
- Specialized data centers with semantic APIs:
  - CUAHSI (hydrology)
  - IEDA (geology)
  - IRIS (seismology)
  - NSIDC (polar)
  - BCO-DMO (ocean)
  - Madrigal (geospace)






# Infrastructure: NSF's DataONE



B. Michener, 2017, from <https://www.slideshare.net/aspecht/michener-workshop-montpellier>

# Infrastructure: ESIP Semantics and Ontology Working Group



Community  
Ontology  
Repository

v3.6.5   Help   Contact us

Term Search   SPARQL Search

[Sign in](#)   [Create account](#)

Filters: ×

**Owner**

[Most recent ontology](#) ☐

☐ esip

☐ gcmd

☐ obo

☐ testorg

☐ ...

**Status**

☐ draft

☒ stable

☐ testing

☐ unstable

**Resource type**

☐ ...

Global filter

<http://cor.esipfed.org/>



# Infrastructure: The NSF EarthCube Initiative

## ■ Many projects use ontologies:

- CINERGI, OntoSoft, Linked Earth, Earth System Bridge, EarthCollab, GeoDeepDive, GeoSemantics, X-DOMES, ...
- See roster at: <https://www.earthcube.org/info/about/funded-projects>



## ■ Infrastructure and tools:

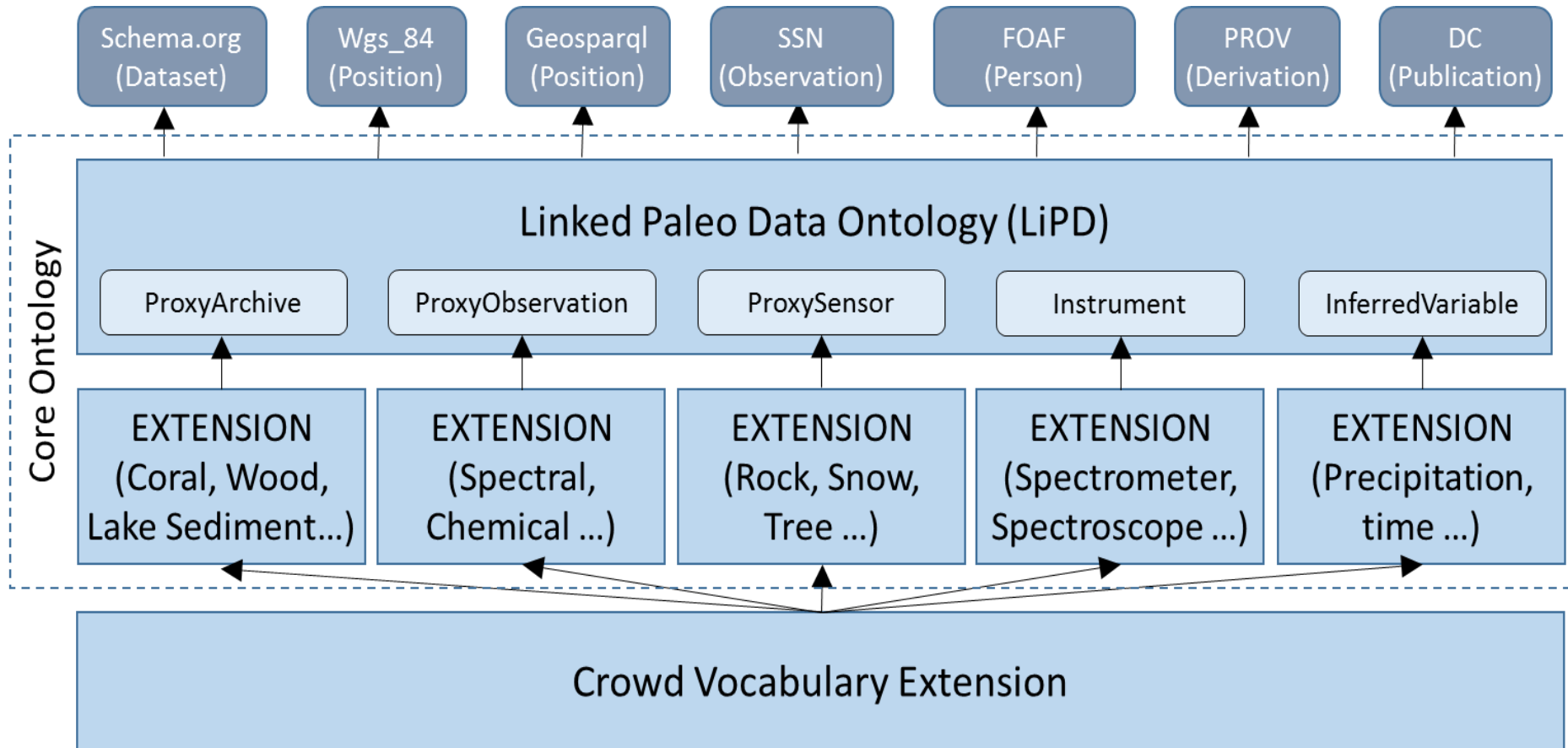
- Text extraction, resource inventory, ontology inventory, data integration, model integration, mediators, semantic services, metadata crowdsourcing, ...
- See <https://www.earthcube.org/tools-inventory>
- Ongoing development of integrated architecture

## ■ Council of Data Facilities

- Includes major data centers in geosciences
- See <https://www.earthcube.org/group/council-data-facilities>

# EarthCube's Linked Earth Project: Controlled Crowdsourcing for Paleoclimate Metadata

*Work with D. Garijo, J. Emile-Geay, D. Khider, V. Ratnakar (USC); N. McKey (NAS)*



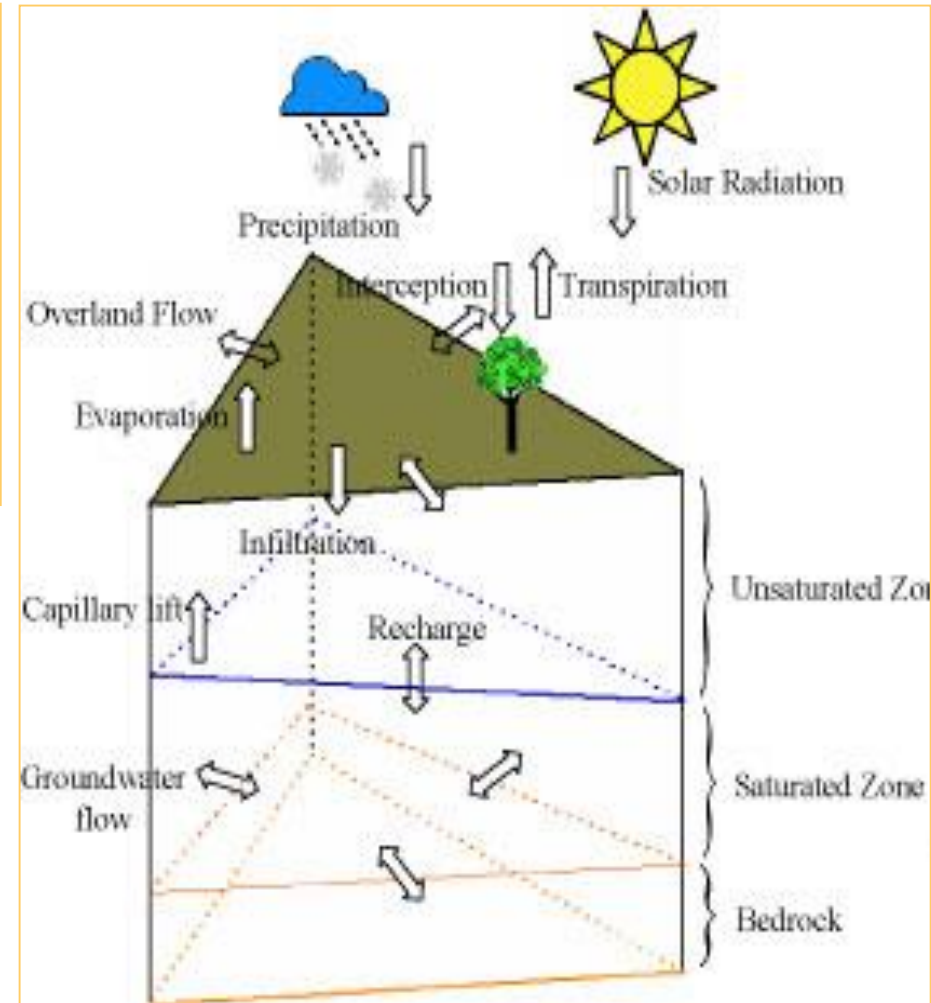
# Outline

---

1. The geosciences landscape
2. Ontologies, vocabularies, standards
3. Infrastructure
  - Data centers
  - Tools
4. Modeling
  - Model repositories
5. Model integration and convergence research
  - The need for semantics

# Modeling in Geosciences: Models of Dynamical Systems

- Historical observational data for calibration
- Forecast data for prediction
- Observational data for evaluation



<http://www.pihm.psu.edu/>

Process	Governing equation/model	Original governing equations	Semi-discrete form ODEs
Channel Routing	St. Venant Equation	$\frac{\partial h}{\partial t} + \frac{\partial (qh)}{\partial x} = q$	$\left( \frac{dh_i}{dt} = P_i - \sum Q_{o,i} + \sum Q_{u,i} + Q_{in,i} - Q_{out,i} - E_i \right)_i$
Overland Flow	St. Venant Equation	$\frac{\partial h}{\partial t} + \frac{\partial (qh)}{\partial x} + \frac{\partial (yh)}{\partial y} = q$	$\left( \frac{dh_i}{dt} = P_i - I - E_i - Q_{oc} + \sum_{j=1}^N Q_{ij}^o \right)_i$
Unsaturated Flow	Richard Equation	$C(\psi) \frac{\partial \psi}{\partial t} = \nabla \cdot (K(\psi) \nabla (\psi + Z))$	$\left( \frac{d\psi_i}{dt} = I - q^o - ET_i \right)_i$
Groundwater Flow	Richard Equation	$C(\psi) \frac{\partial \psi}{\partial t} = \nabla \cdot (K(\psi) \nabla (\psi + Z))$	$\left( \frac{d\psi_i}{dt} = q^s + \sum_{j=1}^N Q_{ij}^s - Q_i + Q_{oc} \right)_i$
Interception	Bucket Model	$\frac{dS_i}{dt} = P - E_i - P_o$	$\left( \frac{dS_i}{dt} = P - E_i - P_o \right)_i$
Snow melt	ISNOBAL	$\frac{dS_{snow}}{dt} = P - E_{snow} - \Delta w$	$\left( \frac{dS_{snow}}{dt} = P - E_{snow} - \Delta w \right)_i$
Evapotranspiration	Pennman-Monteith Method	$ET_o = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_o}}{\Delta + \gamma \left(1 + \frac{r_o}{r_a}\right)}$	$\left( ET_i = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_o}}{\Delta + \gamma \left(1 + \frac{r_o}{r_a}\right)} \right)_i$



# Infrastructure: Model Repositories

---



MATLAB Central



# EarthCube's OntoSoft Project: A Software Metadata Registry [Gil et al eScience 2016]



*Work with D. Garijo, J. Emile-Geay, D. Khider, V. Ratnakar (USC); N. McKey (NAS)*

## Searchable metadata in OntoSoft

## Codes in shared software repositories

**Software Repository**  
Describe your software so others can find and use it

**Software List** [COMPARE]

- CSDMS 1D Hillslope MCMC**  
The model evolves a 1D hillslope according to a non-linear diffusion rule [e.g. Roering et al. 1999] for varying boundary conditions idealised as a gaussian pulse of baselevel fall through time. A Markov Chain Monte Carlo inversion finds the most likely boundary condition parameters when compared...  
Author: Martin Hurst  
Posted by: admin at 2015-09-21 08:05
- CSDMS 2DFLOWVEL**  
2D unsteady nonlinear tidal & wind-driven coastal circulation  
Author: Rudy Slingerland  
Posted by: admin at 2015-09-21 08:05
- C4P 2SAMPLES**  
A software code for estimating difference in location and scale between two climate data samples...  
Posted by: admin at 2015-09-21 08:09
- 3DDY**  
3DDY is a set of scripts that transform geospatial datasets into multiple formats, including GeoJSON, TopoJSON, and STL. The JSON formats are appropriate for interactive visualization of datasets across platforms and applications, while STL formats enable 3D printing for integrating tangible objects...  
Author: Suzanne A Pierce  
Posted by: admin at 2015-09-21 08:03
- C4P A Practical Guide to Wavelet Analysis**

**Filter Software List**

- Search
- Author
- Keywords
- Language
- License
- Operating System
- Publisher

**github**  
SOCIAL CODE HOSTING

**CSDMS**  
COMMUNITY SURFACE DYNAMICS MODELING SYSTEM



(Can export metadata in HTML/XML/RDF/JSON and put in code sharing site)

<http://www.ontosoft.org/>



# Standard Names for Model Variables

## [Peckham iEMSs 2014]

### Model A

Output variables:

- streamflow
- rainrate

### Model B

Input variables:

- discharge
- precip\_rate

## Geoscience Standard Names

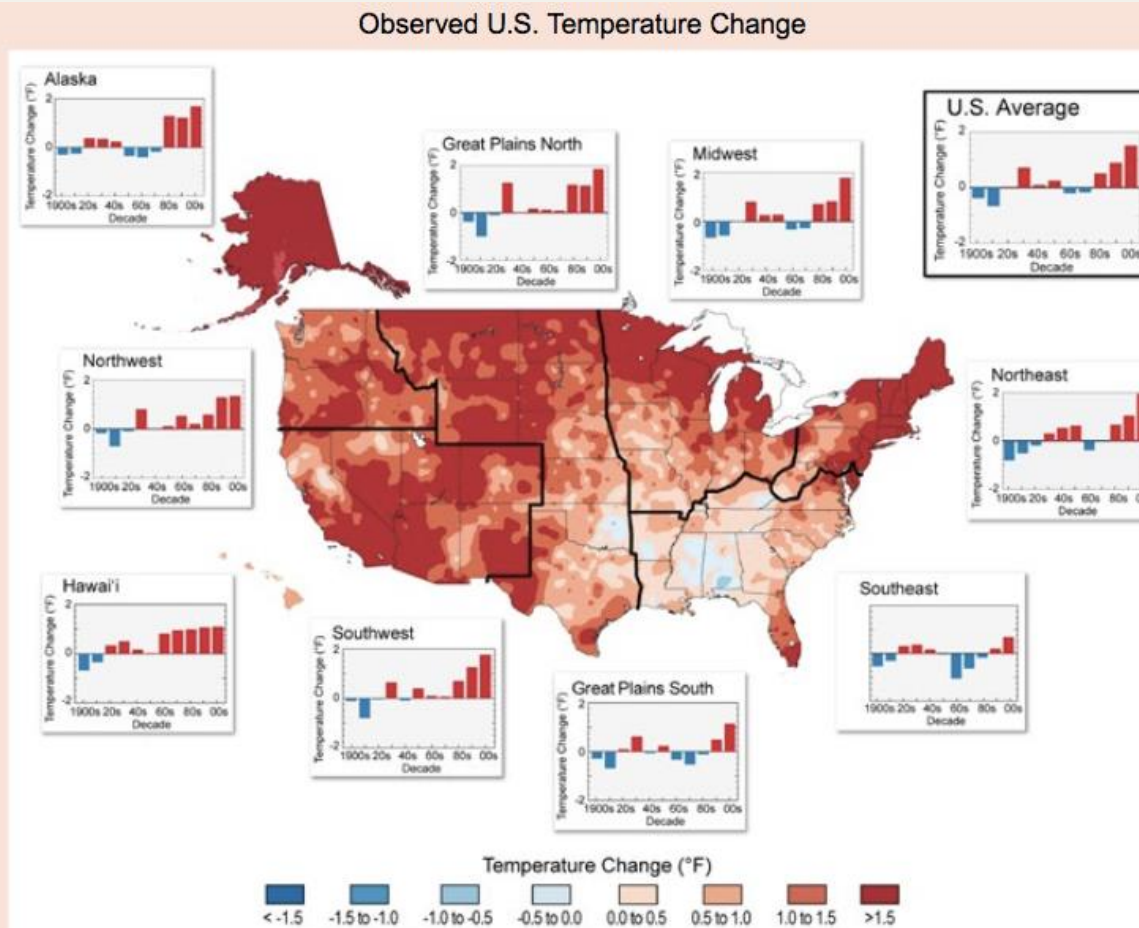
- watershed\_outlet\_water\_\_  
volume\_outflow\_rate
- atmosphere\_water\_\_liquid\_  
equivalent\_precipitation\_rate

atmosphere\_air\_\_increment\_of\_temperature  
glacier\_bottom\_ice\_\_magnitude\_of\_shear\_stress  
atmosphere\_air\_flow\_\_east\_derivative\_of\_pressure  
atmosphere\_air\_flow\_\_elevation\_angle\_of\_gradient\_of\_pressure  
atmosphere\_air\_flow\_\_magnitude\_of\_gradient\_of\_pressure  
...

<http://www.geoscienceontology.org/>



# Transparency and Reproducibility: Third Global Climate Assessment through GCIS [Tilmes 2014]



**Figure 2.7.** The colors on the map show temperature changes over the past 22 years (1991-2012) compared to the 1901-1960 average, and compared to the 1951-1980 average for Alaska and Hawai'i. The bars on the graphs show the average temperature changes by decade for 1901-2012 (relative to the 1901-1960 average) for each region. The far right bar in each graph (2000s decade) includes 2011 and 2012. The period from 2001 to 2012 was warmer than any previous decade in every region. (Figure source: NOAA NCDC / CICS-NC).

<http://nca2014.globalchange.gov/downloads>

W3C<sup>®</sup> PROV

Data  
+ Models  
+ Software  
+ Workflow



## Modern Paper

### Text:

Narrative of the method, some data is in tables, figures/plots, and the software used is mentioned

### Data:

Include data as supplementary materials and pointers to data repositories

## Open Science

### Sharing:

Deposit data and software (and provenance/workflow) in publicly shared repositories

### Open licenses:

Open source licenses for data and software (and provenance/workflow)

### Metadata:

Structured descriptions of the characteristics of data and software (and provenance/workflow)

## Reproducible Publication

### Software:

For data preparation, data analysis, and visualization

### Provenance and methods:

Workflow/scripts specifying dataflow, codes, configuration files, parameter settings, and runtime dependencies

## Digital Scholarship

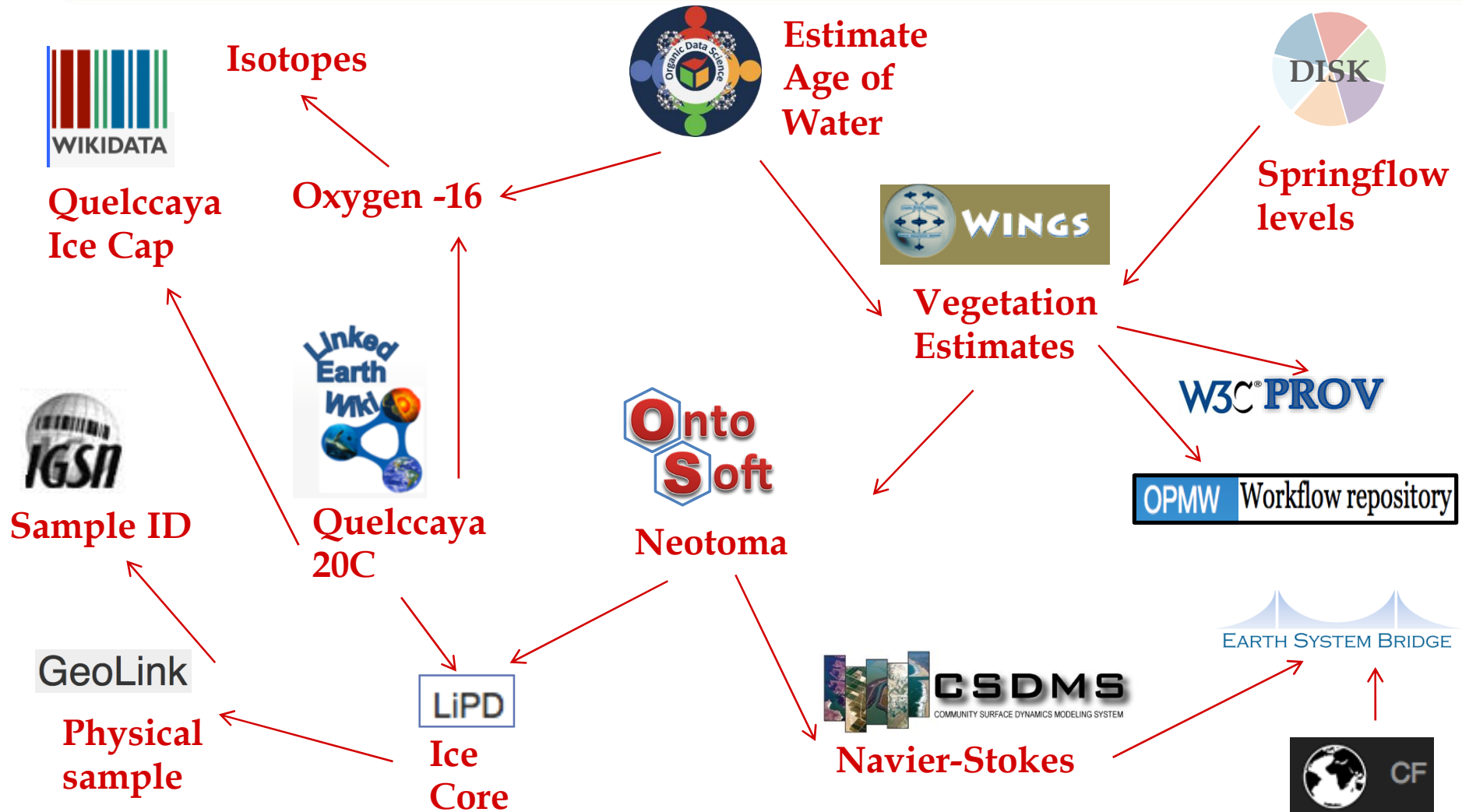
### Persistent identifiers:

For data, software, and authors (and provenance/workflow)

### Citations:

Citations for data and software (and provenance/workflow)

# Linked Data and Knowledge in Geosciences: Data + Models + Software + Workflows



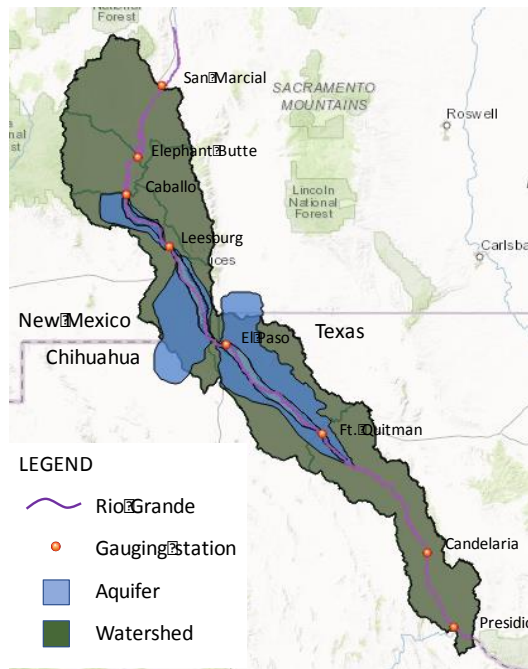
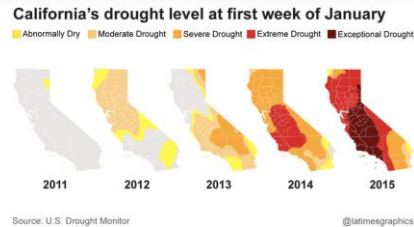
# Outline

---

1. The geosciences landscape
2. Ontologies, vocabularies, standards
3. Infrastructure
  - Data centers
  - Tools
4. Modeling
  - Model repositories
5. Model integration and convergence research
  - The need for semantics

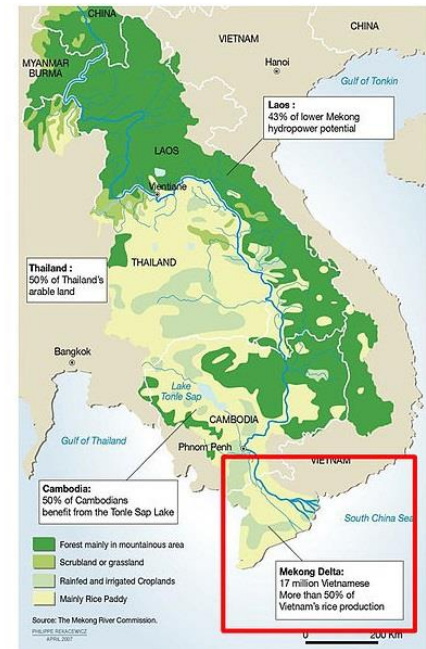


# Model Integration Is Needed to Understand Water Use, Land Cover Changes, Food Insecurity,...



Credit: Deana Pennington, Cybershare Project, UT El Paso

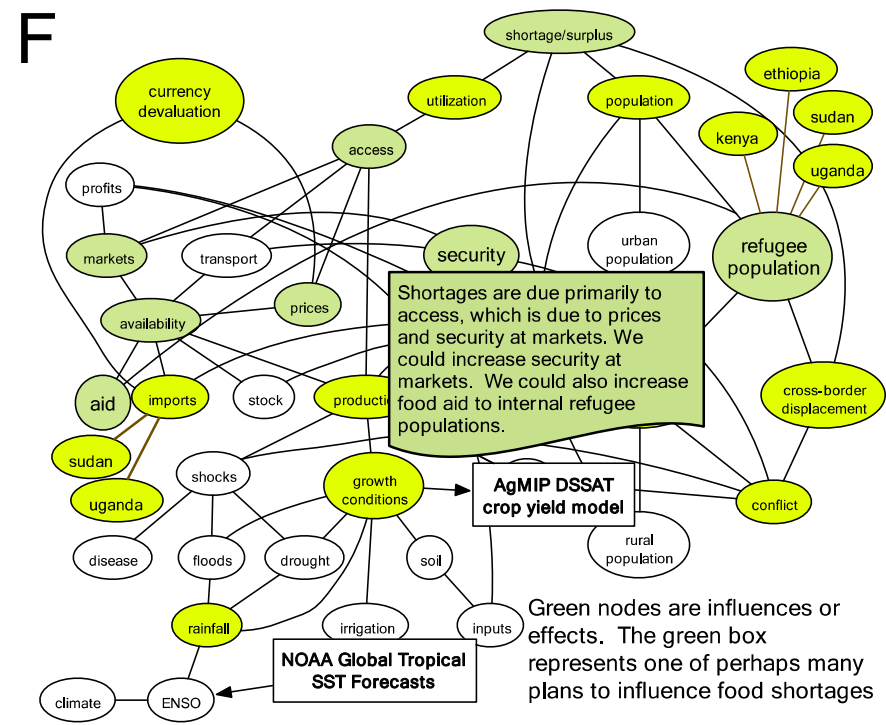
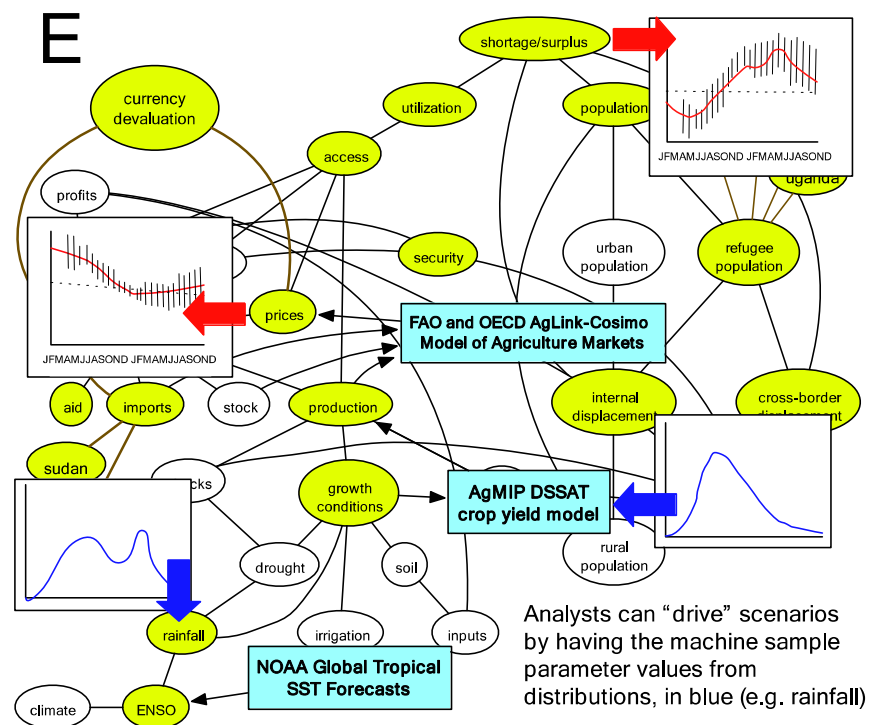
Pecan crops have greatest value but are high water users. Economic value of agriculture is much less than industrial uses, but first in time/first in right in U.S. precludes water allocations to these uses; Mexico has reallocated all surface water to industry



<https://news.mongabay.com/2016/10/vietnam-sweats-bullets-as-china-laos-dam-the-mekong/>

Extends through extends through Tibet, South China, Thailand, Laos, Myanmar, Cambodia, and Vietnam. More than 70 dams are planned in several nations. Recorded deeper droughts and bigger floods than ever before. 2M tons of fish and 500,000 tons of other aquatic animals. Forest cover has decreased from 73 percent in 1973 to 63 percent in 1993. Rice in Cambodia...

# New DARPA World Modelers Program

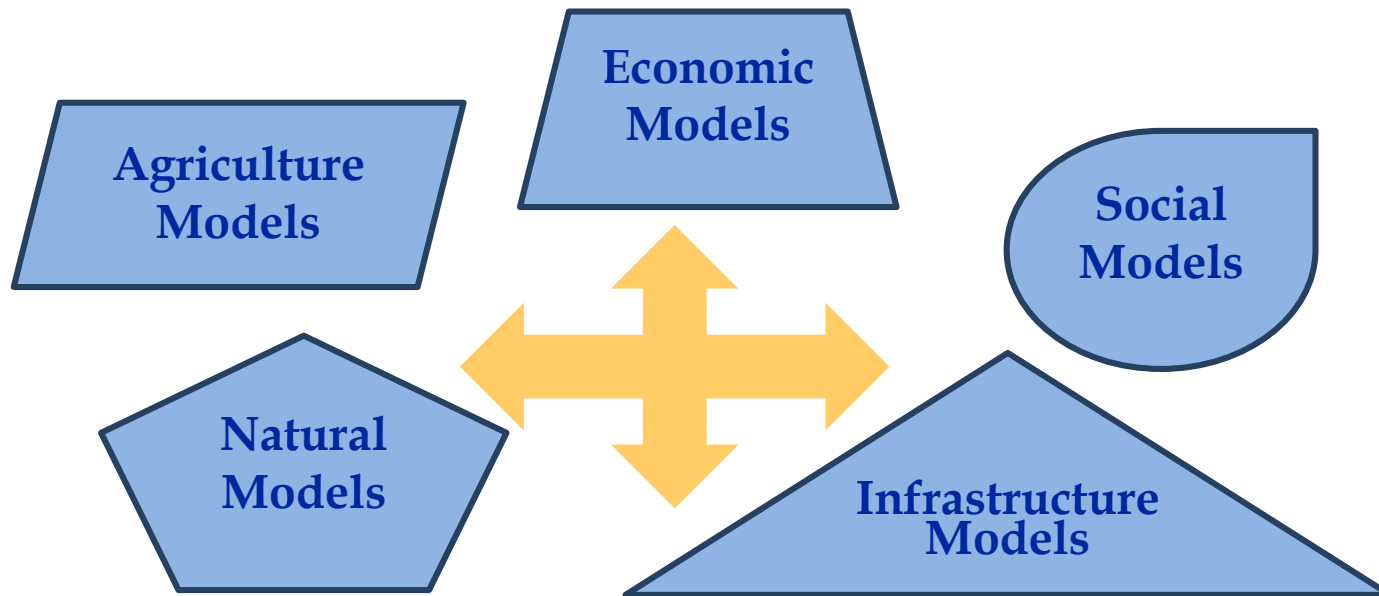


*"World Modelers aims to develop technologies to facilitate analyses that are comprehensive, targeted, causal, quantitative, probabilistic, and timely enough to recommend specific actions that could avert crises."*

# New DARPA World Modelers Program: Semantic Challenges in Model Integration



- **A challenging aspect is mapping model variables**
  - **Standard ontologies needed to describe diverse models**





# New DARPA World Modeler Program: Model INTegration (MINT) Project [Gil et al 2017]



Economic  
Models

Agriculture  
Models

- **Modeling methodology**
  - Empirical (from prior data)
  - Mechanistic (first principles)
  - Mixed
- **Representative models**
  - Many dimensions of diversity
- **Model variables**
  - Ontologies in some domains
- **Integration approaches**
  - Very diverse

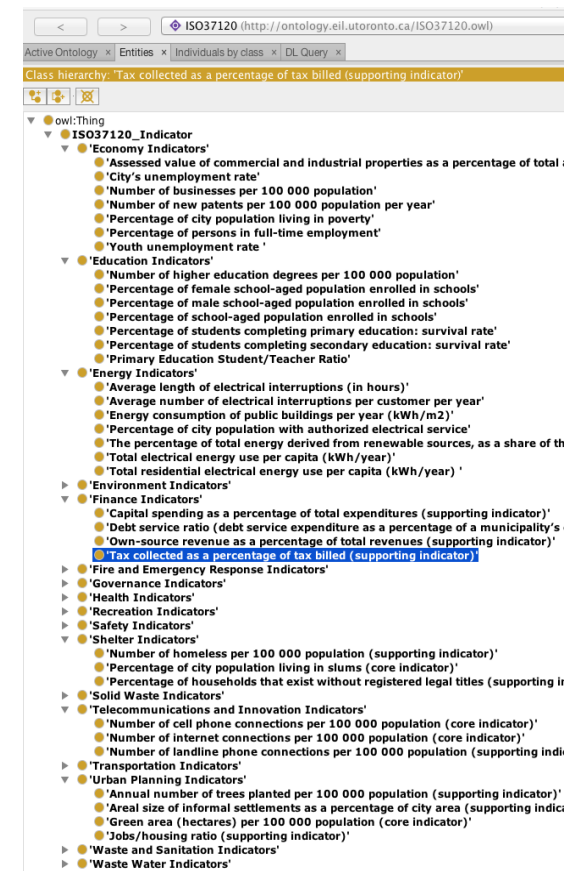
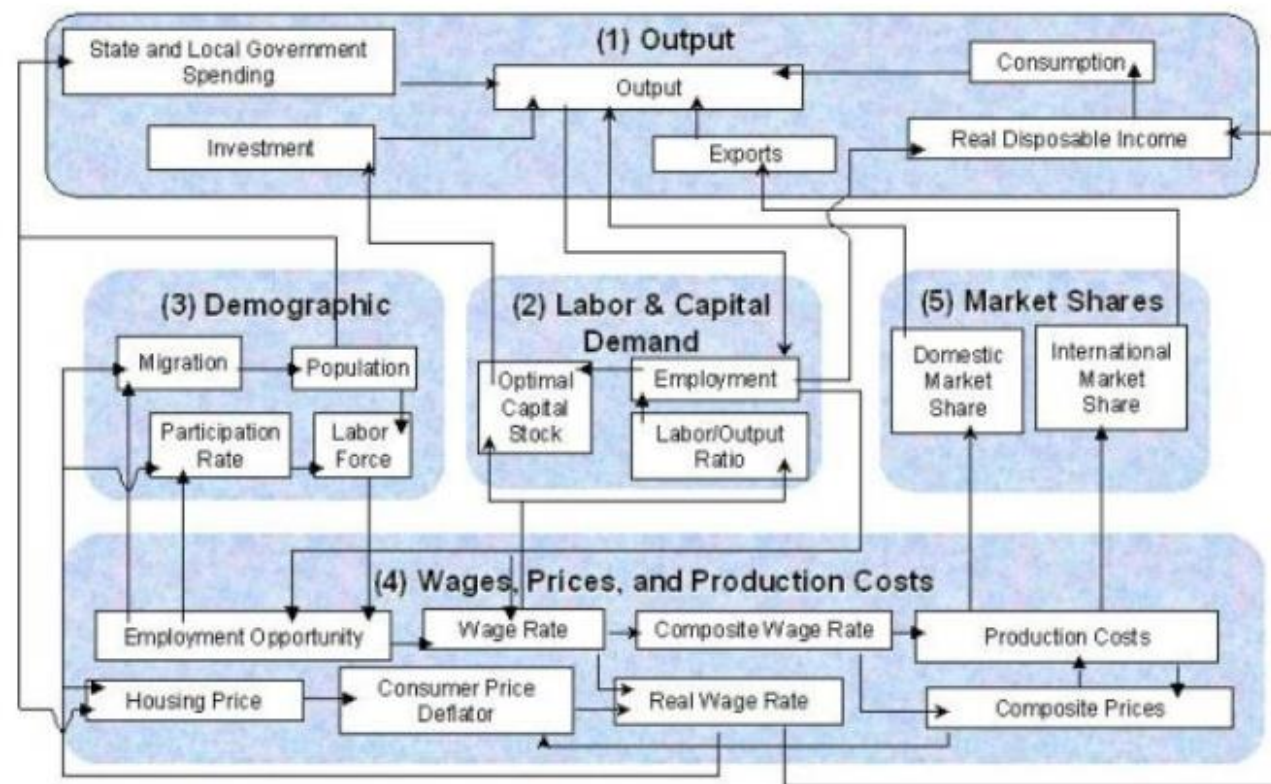
Social  
Models

Natural  
Models

# Economic Models: Very Different from Natural Models, Difficult to Reuse and Integrate



## City Indicators (ISO-37120)



# Summary

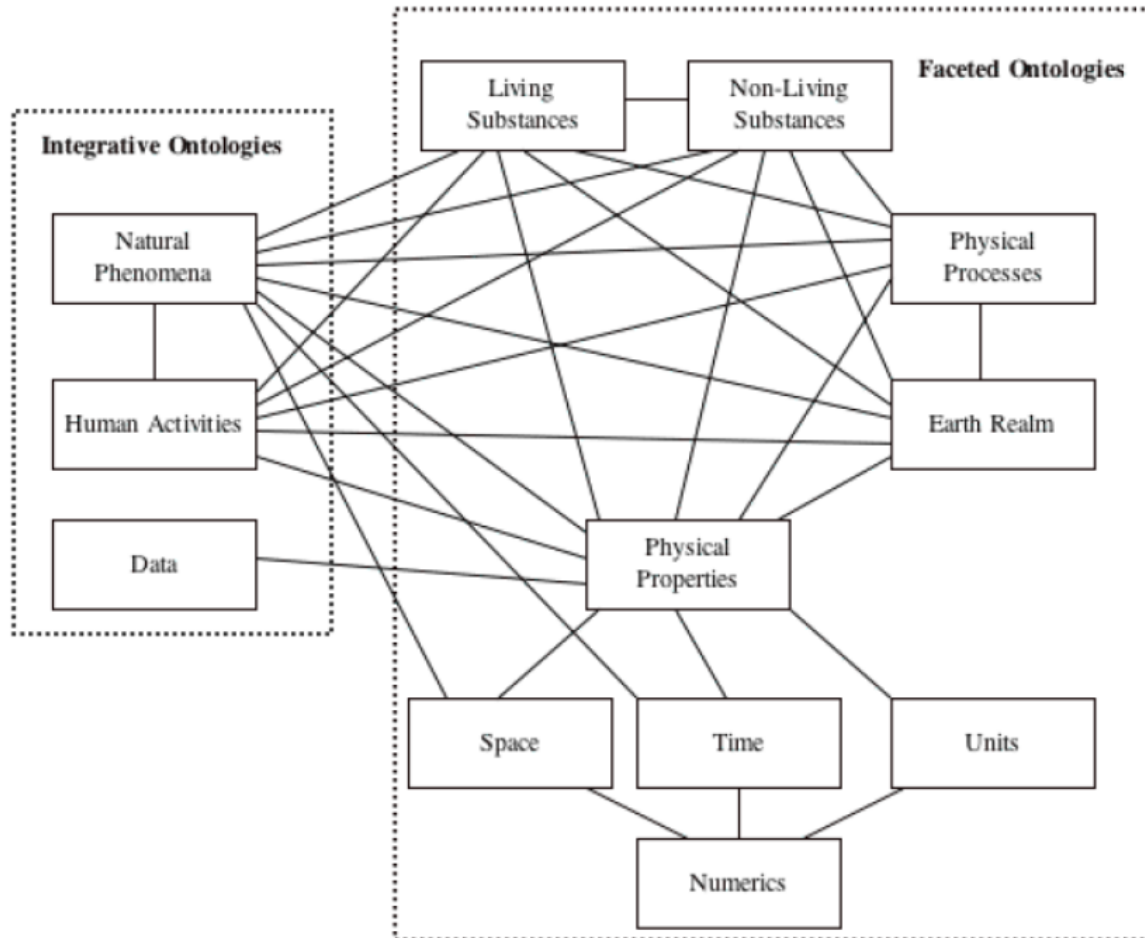
---

1. The geosciences landscape
  - Initial focus could be the NSF CISE-GEO EarthCube initiative
2. Ontologies, vocabularies, standards
  - ESIP Community Ontology Repository
3. Infrastructure
  - Data centers generally speak RDF
  - Tools developed in many EarthCube projects
4. Modeling
  - Model repositories
5. Model integration
  - Model reuse and integration requires semantics
  - Model integration is at the heart of convergent research in geosciences with great societal impact

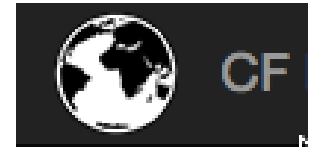


# Ontologies and Vocabularies: Examples

## SWEET Ontologies [Raskin and Pan 2005]



<https://doi.org/10.1016/j.cageo.2004.12.004>



Name:

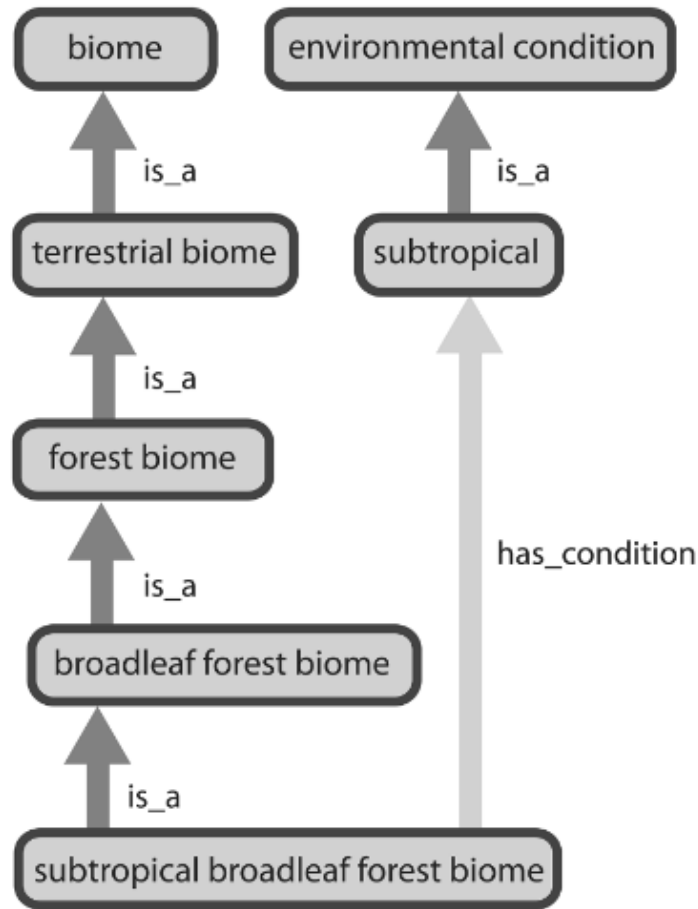
change\_over\_time\_in\_surface  
\_snow\_amount

Description: The surface  
called "surface" means the  
lower boundary of the  
atmosphere.

"change\_over\_time\_in\_X"  
means change in a quantity X  
over a time-interval, which  
should be defined by the  
bounds of the time  
coordinate. "Amount" means  
mass per unit area. Surface  
amount refers to the amount  
on the ground, excluding that  
on the plant or vegetation  
canopy.

Canonical units: kg m-2

# ENVO [Buttigieg et al 2013]



<https://doi.org/10.1186/2041-1480-4-43>

# EarthCube's Linked Earth Project: Creating New Metadata Properties as Needed



## Lake Bosumtwi Sediments Dataset

Data
• <b>DOWNLOAD</b>
From: <a href="http://www.organicdatapublishing.org/index.php/Lake_Bosumtwi_Sediments">http://www.organicdatapublishing.org/index.php/Lake_Bosumtwi_Sediments</a>

### Structured Properties

Add
[x] Archive LakeSediments (By)
[x] Doi 10.1126/science.1166352 (By)
[x] Domain(s) Paleolimnology, Geochemis... (By)
[x] Interpretation Lake Level (By)
[x] Is a Paleoclimate dataset (By)
[x] Measurement Delta18O (By)
[x] MeasurementMaterial Authigenic Calcite (By)
[x] MeasurementStandard VPDB (By)
[x] MeasurementUnits Permil (By)
[x] Reference Shanahan et al. 2009 (By)
[x] Resolution 5 (By)
[x] ResolutionUnits Years (By)
[x] SiteLatitude 6.5 (By)
[x] SiteLatitudeUnits Degrees North (By)
[x] SiteLongitude 358.58 (By)
[x] SiteLongitudeUnits Degrees East (By)
[x] SiteName Bosumtwi (By)
[x] TimeUnits Years CE (By)
[x] WhoAnalysed Tim Shanahan (By)

## Palmyra coral 20C

Data
• <b>DOWNLOAD</b>
From: <a href="http://www.ncdc.noaa.gov/paleo/metadata/noaa-coral-1865.html">http://www.ncdc.noaa.gov/paleo/metadata/noaa-coral-1865.html</a>

### Structured Properties

Add
[x] Archive Coral (By Julien)
[x] Doi 10.1029/2001GL012919 (By Julien)
[x] Domain(s) Climate,geochemistry (By Julien)
[x] Forward model 10.1029/2011GL048224 (By Julien)
[x] Genus Porites (By Julien)
[x] Interpretation SST,SSS (By Julien)
[x] Is a Paleoclimate dataset (By Gil)
[x] Measurement Delta18O (By Julien)
[x] MeasurementMaterial Skeletal aragonite (By Julien)
[x] MeasurementStandard VPDB (By Julien)
[x] MeasurementUnits Permil (By Julien)
[x] Reference Cobb et al, 2001 (By Julien)
[x] Resolution 0.0833 (By Julien)
[x] ResolutionUnits Years (By Julien)
[x] SiteLatitude 5.883 (By Julien)
[x] SiteLatitudeUnits Degrees North (By Julien)
[x] SiteLongitude 197.917 (By Julien)
[x] SiteLongitudeUnits Degrees East (By Julien)
[x] SiteName Palmyra (By Julien)
[x] Species Blue porites (By Anonymous)
[x] Species Lutea (By Julien)

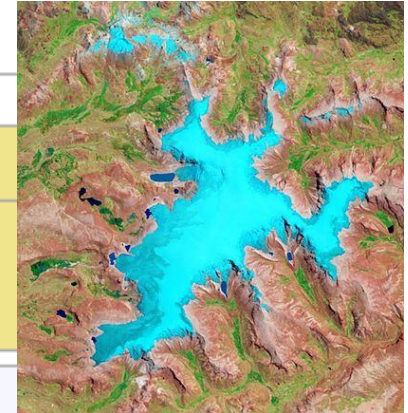


# EarthCube's Linked Earth Project: Promoting Property Normalization and Standards

## Quelccaya Ice Core Dataset

### Data

- [DOWNLOAD](#)
- [Data Types](#)



### Structured Properties

Add

[x]	Archive	IceCore	(By Julien)
[x]	Doi	10.1073/pnas.0603900103	(By Julien)
[x]	Domain(s)	Climate,Glaciology,Geoche...	(By Julien)
[x]	Is a	Paleoclimate dataset	(By Gil)
[x]	Measu		

Measurement

MeasurementMaterial

MeasurementStandard

MeasurementUnits



# EarthCube's Linked Earth Project: Connecting to Other Ontologies/Data

**Palmyra Atoll** [edit]

Structured Properties

main type (GND)	geographical feature	[edit]
is in the administrative unit	United States Minor Outlying Islands	[edit]



## Palmyra coral 20C

### Data

#### • DOWNLOAD

From: <http://www.ncdc.noaa.gov/paleo/metadata/noaa-coral-1865.html>

### Structured Properties

[x]	SiteName	Palmyra	(By Julien)
[x]	Archive	Coral	(By Julien)
[x]	Domain(s)	Climate,geochemistry	(By Julien)
[x]	Forward model	10.1029/2011GL048224	(By Julien)
[x]	Genus	Porites	(By Julien)
[x]	Interpretation	SST,SSS	(By Nick)
[x]	Measurement	Delta18O	(By Julien)
[x]	MeasurementMaterial	Skeletal aragonite	(By Julien)
[x]	MeasurementStandard	VPDB	(By Nick)
[x]	MeasurementUnits	Permil	(By Julien)
[x]	Reference	Cobb et al, 2001	(By Julien)
[x]	Species	Lutea	(By Julien)

- NCBITaxon ontology: porites  
<http://bioportal.bioontology.org/ontologies/NCBITAXON?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FNCBITAXON%2F46719>

## Porites

### Structured Properties

[x]	Property Name	Topic:Finger Coral	[hide]
		<ul style="list-style-type: none"><li>[x] <a href="http://dbpedia.org/resource/Porites">http://dbpedia.org/resource/Porites</a></li><li>[add source]</li></ul>	

Wikipedia Entry  go to original Wikipedia article

**Porites** is a genus of stony coral; they are SPS (Small Polyp Stony) corals. They are characterised by a finger-like morphology. Members of this genus have widely spaced



### Web of Data

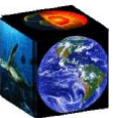


hide	dbpedia
class	Coral
classis	Coral
familia	Poritidae
family	Poritidae
genus	Porites
kingdom	Animal
name	Finger Coral
order	Fungina
order	Scleractinia
ordo	Scleractinia
phylum	Cnidaria

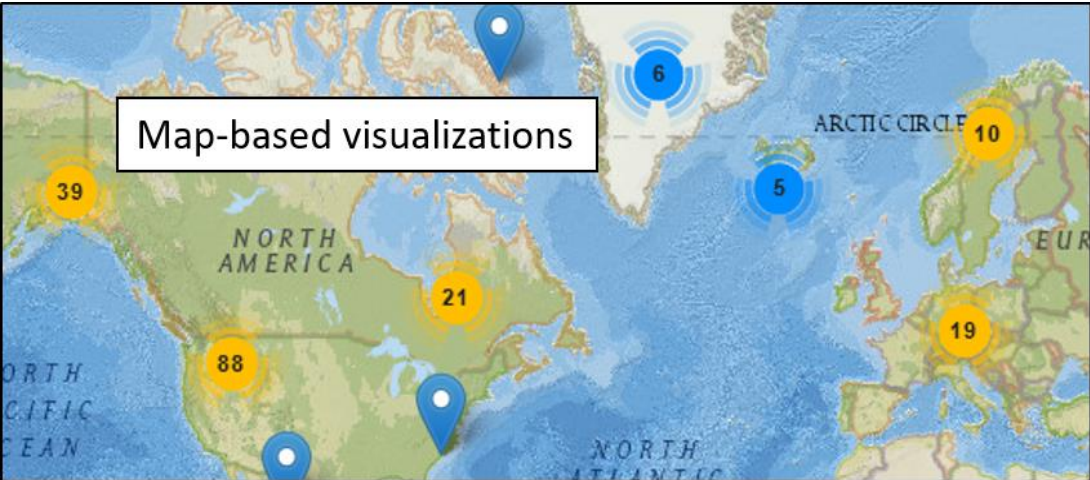
### Credits

Users who have contributed to this Page:

- Julien (43 Edits)
- Nick (34 Edits)



# EarthCube's Linked Earth Project: Social Aspects of Vocabulary Crowdsourcing



Map-based visualizations

Africa-LakeTanganyi.Tierney.2010  
( Dataset (L) )

Download (LPD)

- Author (L)
- CollectedFrom (L)
- Contributor (L)
- FunderBy (L)
- IncludesChromData (L)
- IncludesPaleoData (L)
- PlanOfCompilation (L)
- PublishedIn (L)
- Classification (L)
- DatasetCite (L)
- DatasetLicense (L)
- DatasetVersion (L)
- HasDOI (L)
- Notes (L)

Extra information


- ArchiveType
- CollectionName
- EarliestSampleDate
- GoogleDataURL

BMC  
Africa-LakeTanganyi.Tierney.2010 Location  
Not defined  
Not defined  
Africa-LakeTanganyi.Tierney.2010 PaleoData1  
Not defined  
Publication 10.1038/Nature08865  
lake sediment  
NP04-KH1 (piston core) and NP04-MC1 (multicore)  
504  
<https://docs.google.com/spreadsheets/d/1LH0CjgFvYNgKZedZv20Jp03Wxy14QpEMz7EwOc>

Dataset annotation interface

Varun Ratnakar  
( Person (L) )

Author credit



Contributions:  
Edited 64 pages Imported 633 LPD Created 2

Working Groups:  
Wiki Administration Test Working Group

Wiki Privileges:  
bureaucrat editorial-board sysop

Email (L)  
OrcidNumber (L)  
Name (L)

Subcategories

Ice Cores Working Group  
Lake Sediments Working Group

Pages in category "Working Group"

The following 2 pages are in this category, out of 2 total.

Cross Archive Metadata  
OC3

Community discussions

Should we keep these two separate concepts?  
Please vote below. Results will be shown when you have voted.  
You are not entitled to view results of this poll before you have voted.

☐ Yes  
☐ No

Polls for decision making

There was one vote since the poll was created on 17:59, 15 March 2017.

# Modeling in Geosciences: Models of Dynamical Systems

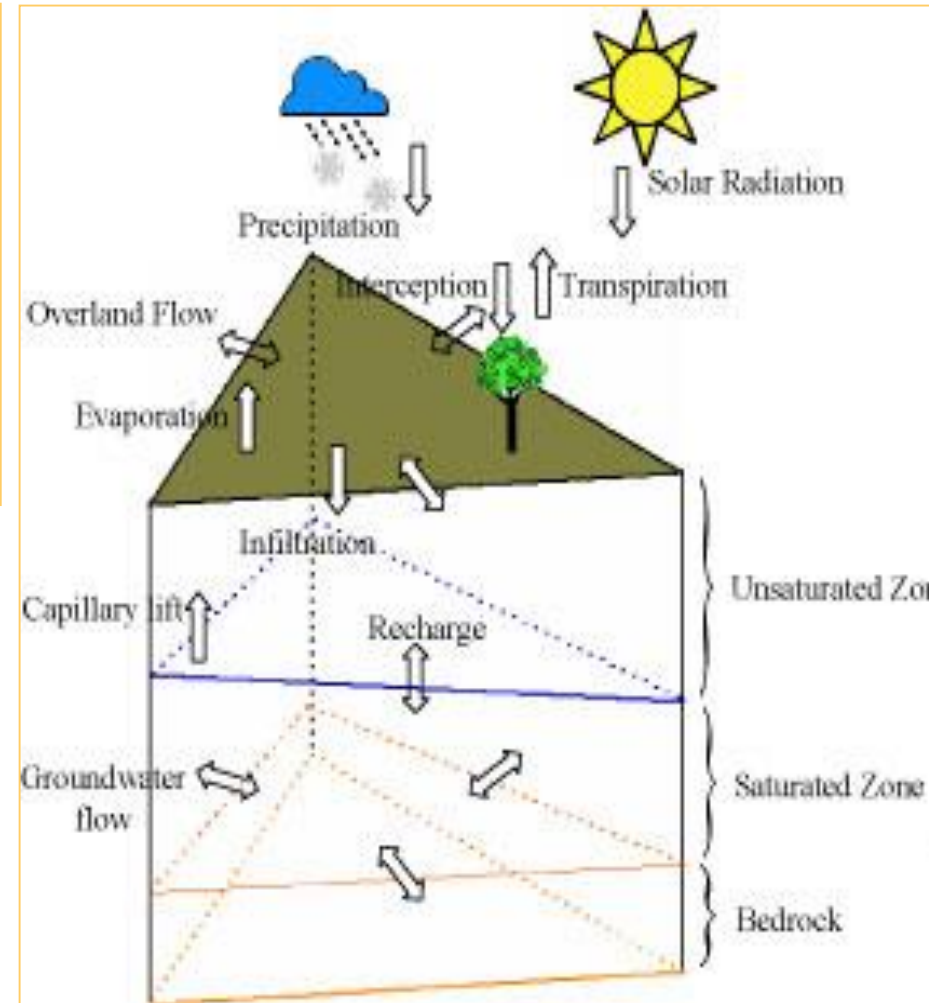
---

- A **simulation model** of a **dynamical system** captures the relationships and **dependencies** between a set of **variables** used to describe it
- Models of dynamical systems are framed in **time and space**
- **Output** variables depend on the **input** variables, internal **state** variables, **exogenous** variables, and **random** variables
- Models can have **parameters** that can be adjusted to fit the empirical observations taken on the system being modeled
- Models are used to **make predictions** about hypothetical configurations and future states of the target system

# Example:

## The PIHM Hydrology Model [Duffy et al 2015]

- Historical observational data for calibration
- Forecast data for prediction
- Observational data for evaluation

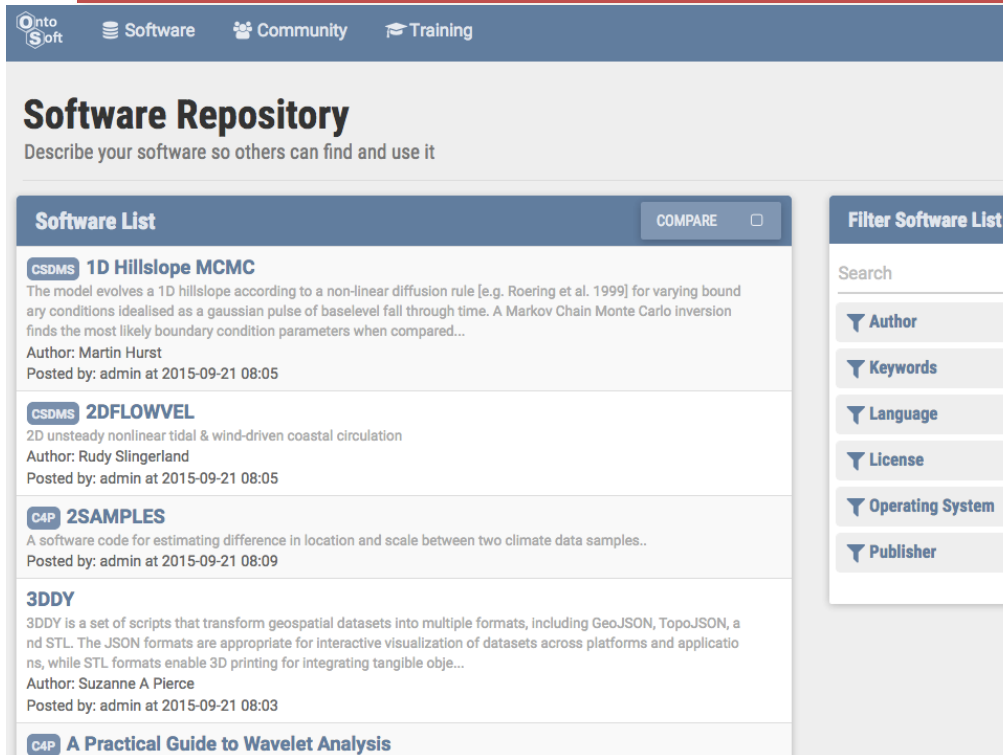


<http://www.pihm.psu.edu/>

Process	Governing equation/model	Original governing equations	Semi-discrete form ODEs
Channel Routing	St. Venant Equation	$\frac{\partial h}{\partial t} + \frac{\partial (qh)}{\partial x} = q$	$\left( \frac{dh_i}{dt} = P_i - \sum Q_{o,i} + \sum Q_{u,i} + Q_{in,i} - Q_{out,i} - E_i \right)_i$
Overland Flow	St. Venant Equation	$\frac{\partial h}{\partial t} + \frac{\partial (qh)}{\partial x} + \frac{\partial (yh)}{\partial y} = q$	$\left( \frac{\partial h}{\partial t} = P_i - I - E_i - Q_{oc} + \sum_{j=1}^3 Q_{s,j}^0 \right)_i$
Unsaturated Flow	Richard Equation	$C(\psi) \frac{\partial \psi}{\partial t} = \nabla \cdot (K(\psi) \nabla (\psi + Z))$	$\left( \frac{d\psi_i}{dt} = I - q^0 - ET_i \right)_i$
Groundwater Flow	Richard Equation	$C(\psi) \frac{\partial \psi}{\partial t} = \nabla \cdot (K(\psi) \nabla (\psi + Z))$	$\left( \frac{d\psi_i}{dt} = q^0 + \sum_{j=1}^3 Q_{g,j}^0 - Q_i + Q_{gc} \right)_i$
Interception	Bucket Model	$\frac{dS_i}{dt} = P - E_i - P_o$	$\left( \frac{dS_i}{dt} = P - E_i - P_o \right)_i$
Snow melt	ISNOBAL	$\frac{dS_{snow}}{dt} = P - E_{snow} - \Delta w$	$\left( \frac{dS_{snow}}{dt} = P - E_{snow} - \Delta w \right)_i$
Evapotranspiration	Pennman-Monteith Method	$ET_o = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_o}}{\Delta + \gamma (1 + \frac{r_o}{r_a})}$	$\left( ET_i = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_o}}{\Delta + \gamma (1 + \frac{r_o}{r_a})} \right)_i$



# EarthCube's OntoSoft Project: Distributed Architecture for Software Registries



**Software Repository**  
Describe your software so others can find and use it

**Software List** [COMPARE] [ ]

**CSDMS 1D Hillslope MCMC**  
The model evolves a 1D hillslope according to a non-linear diffusion rule [e.g. Roering et al. 1999] for varying boundary conditions idealised as a gaussian pulse of baselevel fall through time. A Markov Chain Monte Carlo inversion finds the most likely boundary condition parameters when compared...

Author: Martin Hurst  
Posted by: admin at 2015-09-21 08:05

**CSDMS 2DFLOWVEL**  
2D unsteady nonlinear tidal & wind-driven coastal circulation

Author: Rudy Slingerland  
Posted by: admin at 2015-09-21 08:05

**C4P 2SAMPLES**  
A software code for estimating difference in location and scale between two climate data samples..

Author: Suzanne A Pierce  
Posted by: admin at 2015-09-21 08:09

**3DDY**  
3DDY is a set of scripts that transform geospatial datasets into multiple formats, including GeoJSON, TopoJSON, a nd STL. The JSON formats are appropriate for interactive visualization of datasets across platforms and applications, while STL formats enable 3D printing for integrating tangible objects...

Author: Suzanne A Pierce  
Posted by: admin at 2015-09-21 08:03

**C4P A Practical Guide to Wavelet Analysis**

**Filter Software List**

Search

Author

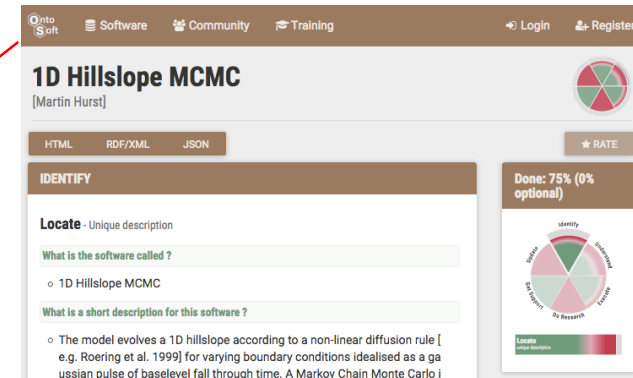
Keywords

Language

License

Operating System

Publisher



**1D Hillslope MCMC**  
[Martin Hurst]

HTML RDF/XML JSON

★ RATE

**IDENTIFY**

**Locate** - Unique description

What is the software called ?

1D Hillslope MCMC

What is a short description for this software ?

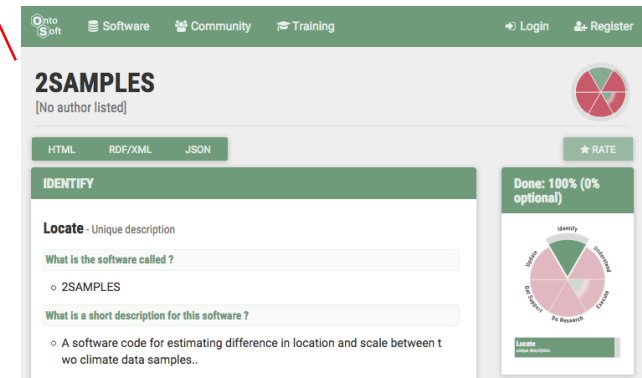
The model evolves a 1D hillslope according to a non-linear diffusion rule [e.g. Roering et al. 1999] for varying boundary conditions idealised as a gaussian pulse of baselevel fall through time. A Markov Chain Monte Carlo inversion finds the most likely boundary condition parameters when compared...

Done: 75% (0% optional)

Identify

Locate

Rate



**2SAMPLES**  
[No author listed]

HTML RDF/XML JSON

★ RATE

**IDENTIFY**

**Locate** - Unique description

What is the software called ?

2SAMPLES

What is a short description for this software ?

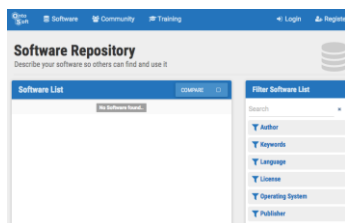
A software code for estimating difference in location and scale between two climate data samples..

Done: 100% (0% optional)

Identify

Locate

Rate



**Software Repository**  
Describe your software so others can find and use it

**Software List** [COMPARE] [ ]

**Filter Software List**

Search

Author

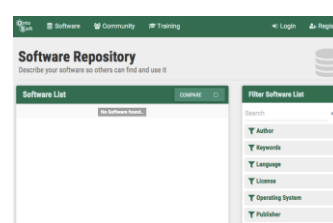
Keywords

Language

License

Operating System

Publisher



**Software Repository**  
Describe your software so others can find and use it

**Software List** [COMPARE] [ ]

**Filter Software List**

Search

Author

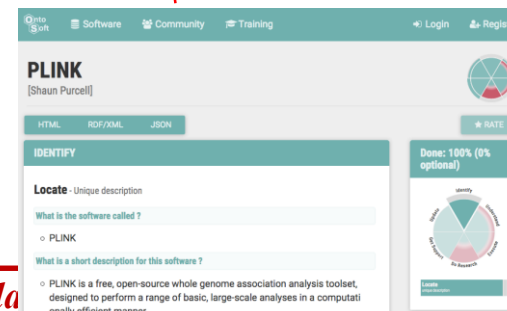
Keywords

Language

License

Operating System

Publisher



**PLINK**  
[Shaun Purcell]

HTML RDF/XML JSON

★ RATE

**IDENTIFY**

**Locate** - Unique description

What is the software called ?

PLINK

What is a short description for this software ?

PLINK is a free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner.

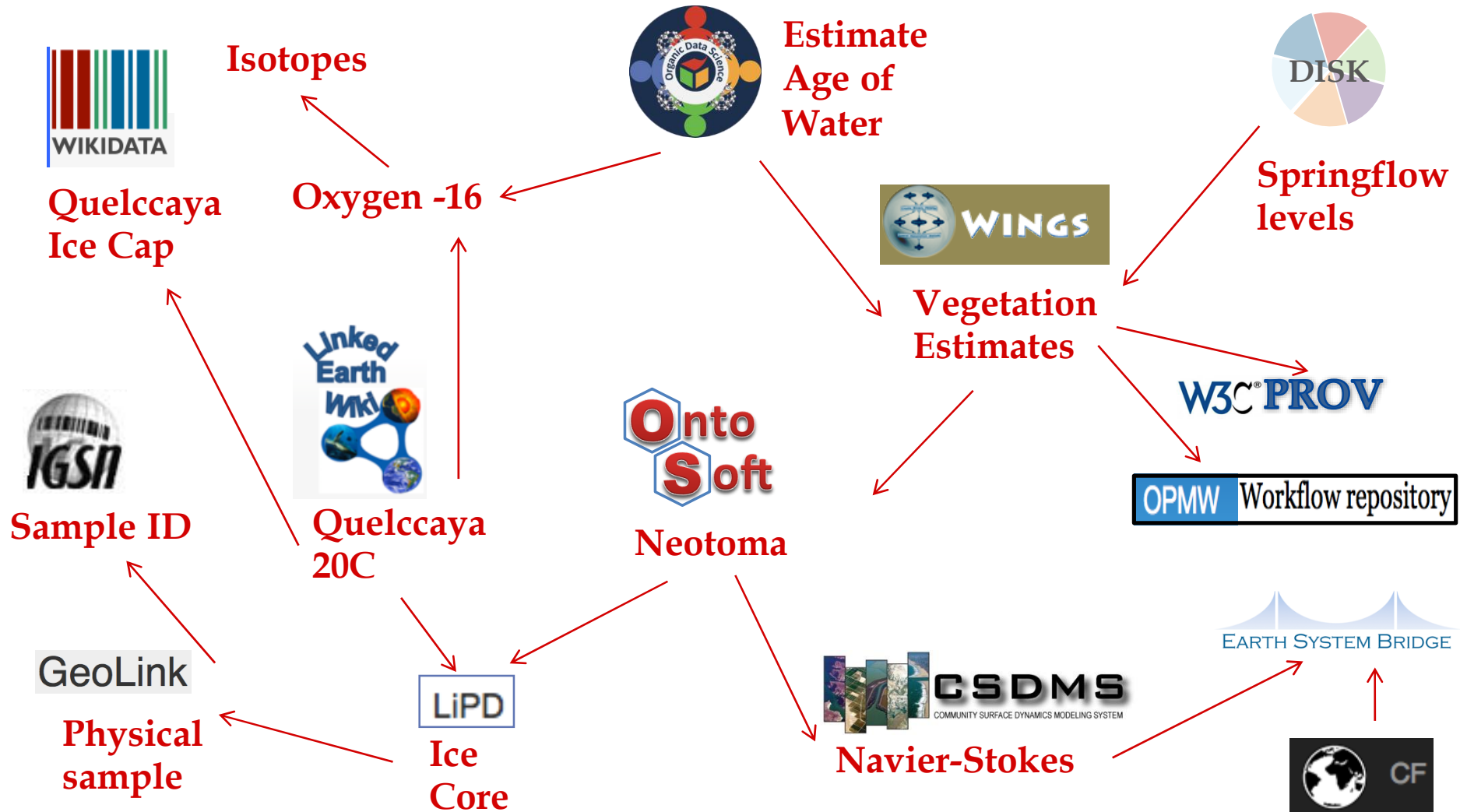
Done: 100% (0% optional)

Identify

Locate

Rate

# Linked Science Data and Knowledge: Data + Models + Software + Workflows



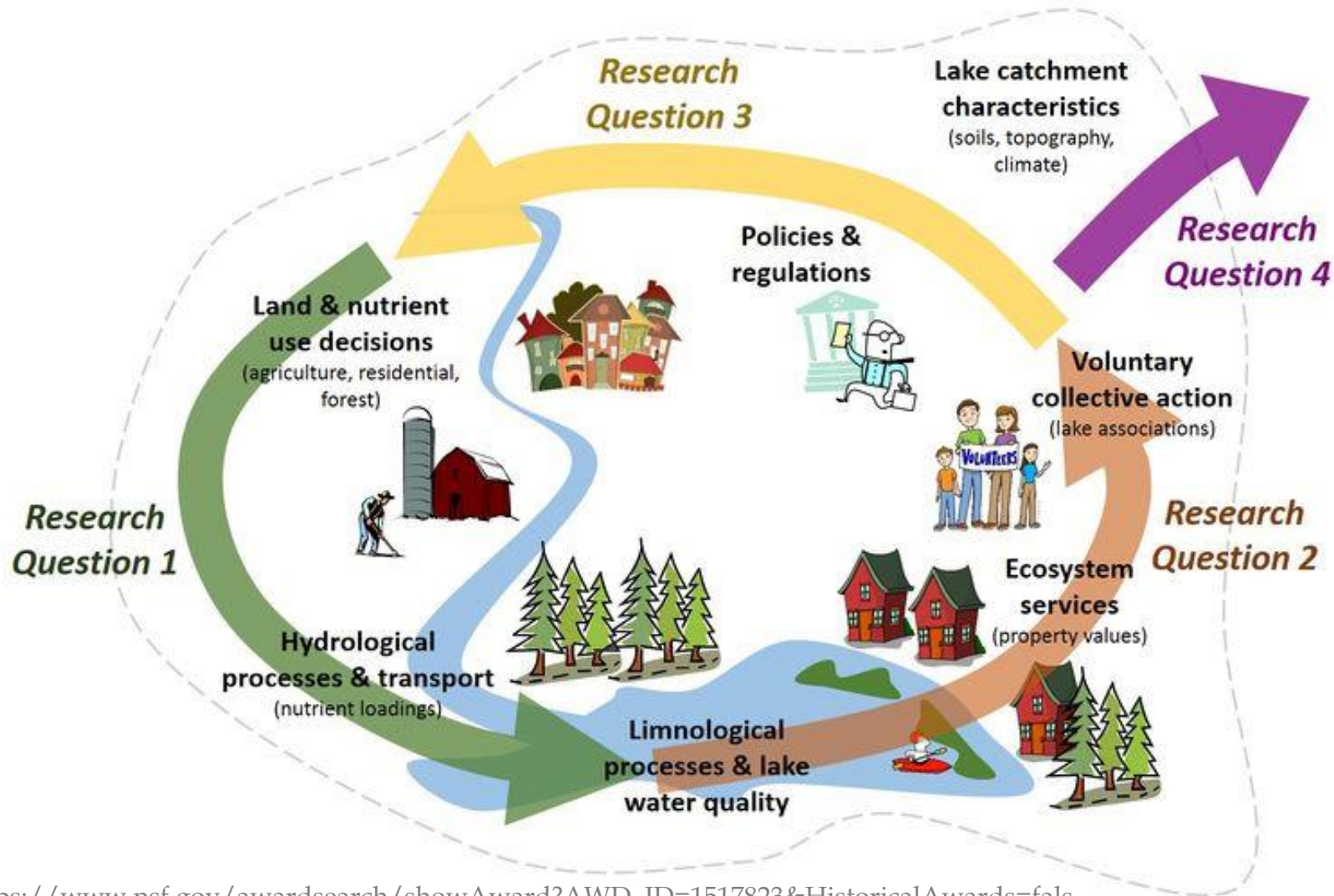
# Modeling in Geosciences: Models of Dynamical Systems

---

- A **simulation model** of a **dynamical system** captures the relationships and **dependencies** between a set of **variables** used to describe it
- Models of dynamical systems are framed in **time and space**
- **Output** variables depend on the **input** variables, internal **state** variables, **exogenous** variables, and **random** variables
- Models can have **parameters** that can be adjusted to fit the empirical observations taken on the system being modeled
- Models are used to **make predictions** about hypothetical configurations and future states of the target system

# Coupling Natural and Human Systems

[Cobourn, Duffy, Hanson, et al 2016]



[https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=1517823&HistoricalAwards=false](https://www.nsf.gov/awardsearch/showAward?AWD_ID=1517823&HistoricalAwards=false)



# Model Integration: Diversity of Strategies

## Model Blending

### Interleaved Execution

*Sliced execution by time tic:  
first one model, then another,  
in a round-robin way*

eg: CSDMS

### Implicit Interleaving

*Collection of equations that  
are designed to be solved together,  
then a solver runs them.*

eg: CGE economic models

### Code Merging

*MPI code to implement all  
models*

eg: earthquake simulations

### Shared Memory

*Models share a R/W memory*

eg: Synthetic Information

## Model Combination

### Result Chaining

*The result of a model is input  
to another model, as in a  
workflow*

eg: pSIMS, CEMSA

### Output Comparison

*Results from several models (or  
the same model) are aggregated  
(eg, an ensemble)*

eg: regional weather prediction

### Output Analysis

*Same model is run with many  
configurations or parameter  
values, to do parameter  
estimation, sensitivity analysis,  
or uncertainty quantification*

## Model Distribution

### Code Parallelization

*The model is implemented  
as parallel code (eg to  
process each grid cell  
separately)*

### Interleaved Behavior

*Individual agents proceed  
based on information made  
available to their  
simulation environment*

Ex: agent-based frameworks

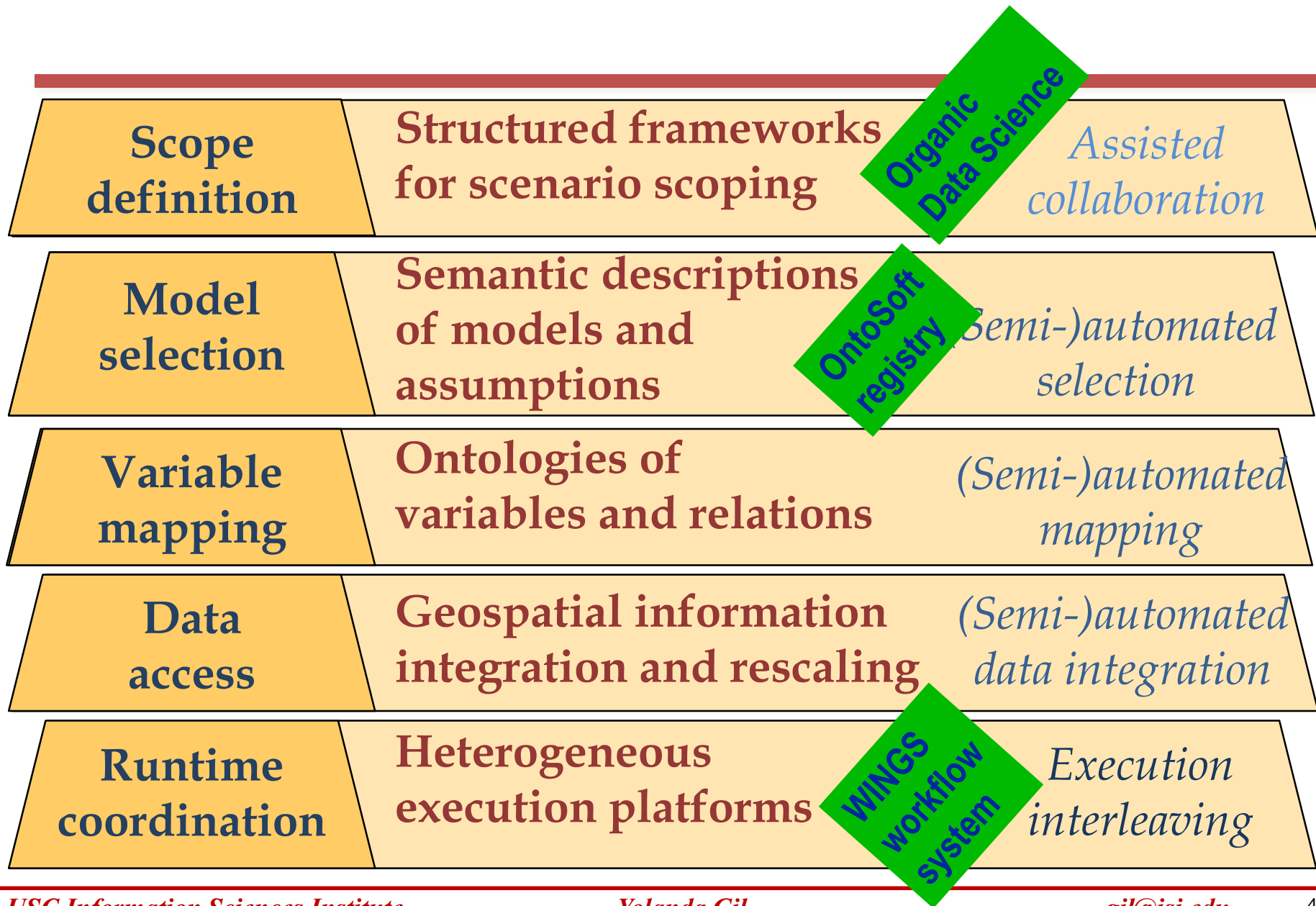
### Integrated Behavior

*Agents are given several  
behavior models that  
determine their actions*

# A Research Agenda for Model Integration

<b>Scope definition</b>	<b>Structured frameworks for scenario scoping</b>	<i>Assisted collaboration</i>
<b>Model selection</b>	<b>Semantic descriptions of models and assumptions</b>	<i>(Semi-)automated selection</i>
<b>Variable mapping</b>	<b>Ontologies of variables and relations</b>	<i>(Semi-)automated mapping</i>
<b>Data access</b>	<b>Geospatial information integration and rescaling</b>	<i>(Semi-)automated data integration</i>
<b>Runtime coordination</b>	<b>Heterogeneous execution platforms</b>	<i>Execution interleaving</i>

# A Research Agenda for Model Integration



# Core Building Blocks

---

**Scope  
definition**

**Organic Data Science (CISE/GEO/BIO)**

**Model  
selection**

**OntoSoft (EC), CNH (GEO)**

**Variable  
mapping**

**Geo Standard Names (EC)**

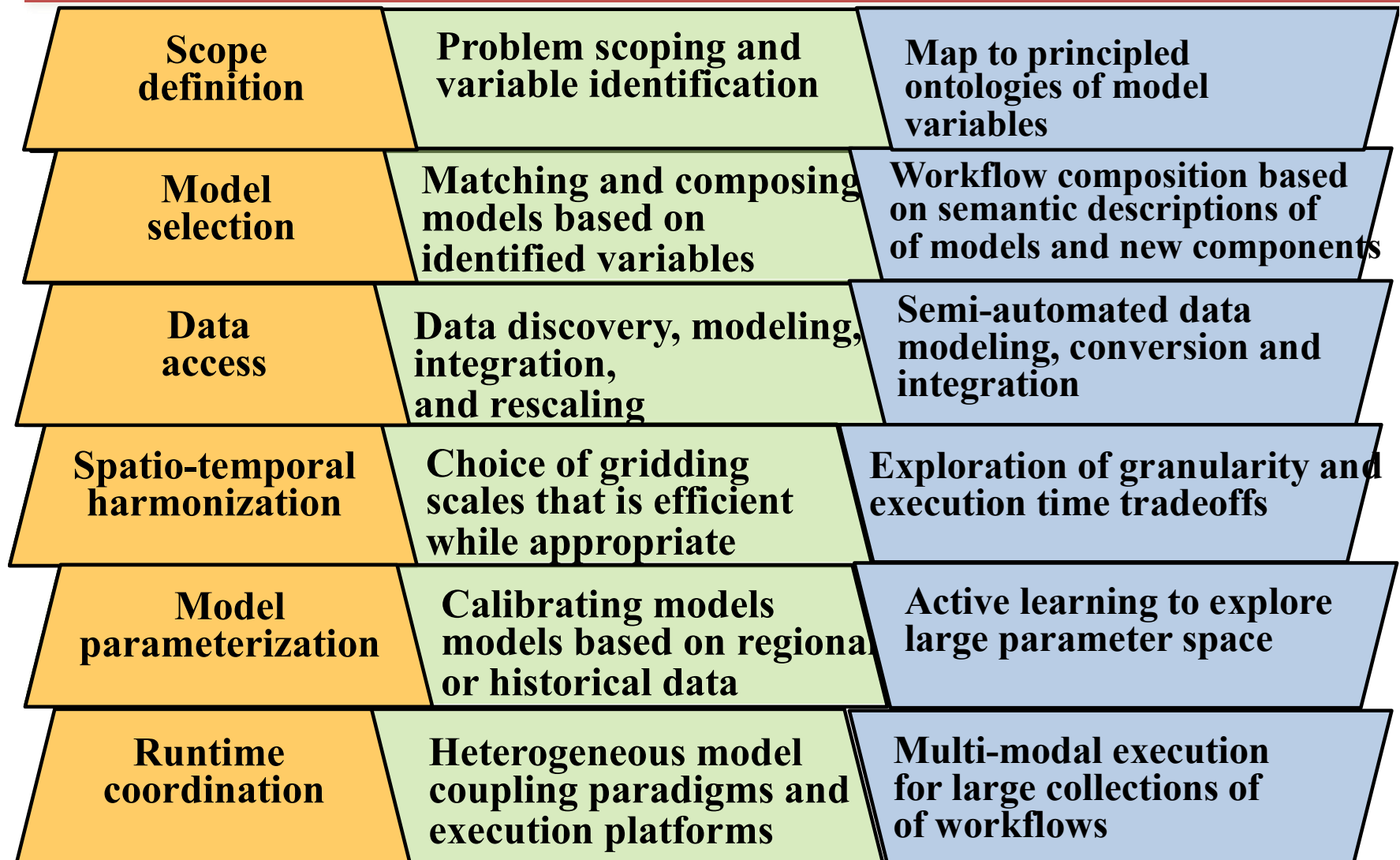
**Data  
access**

**ML-Remote (CISE), Hydroterre (GEO)  
Karma (DARPA)**

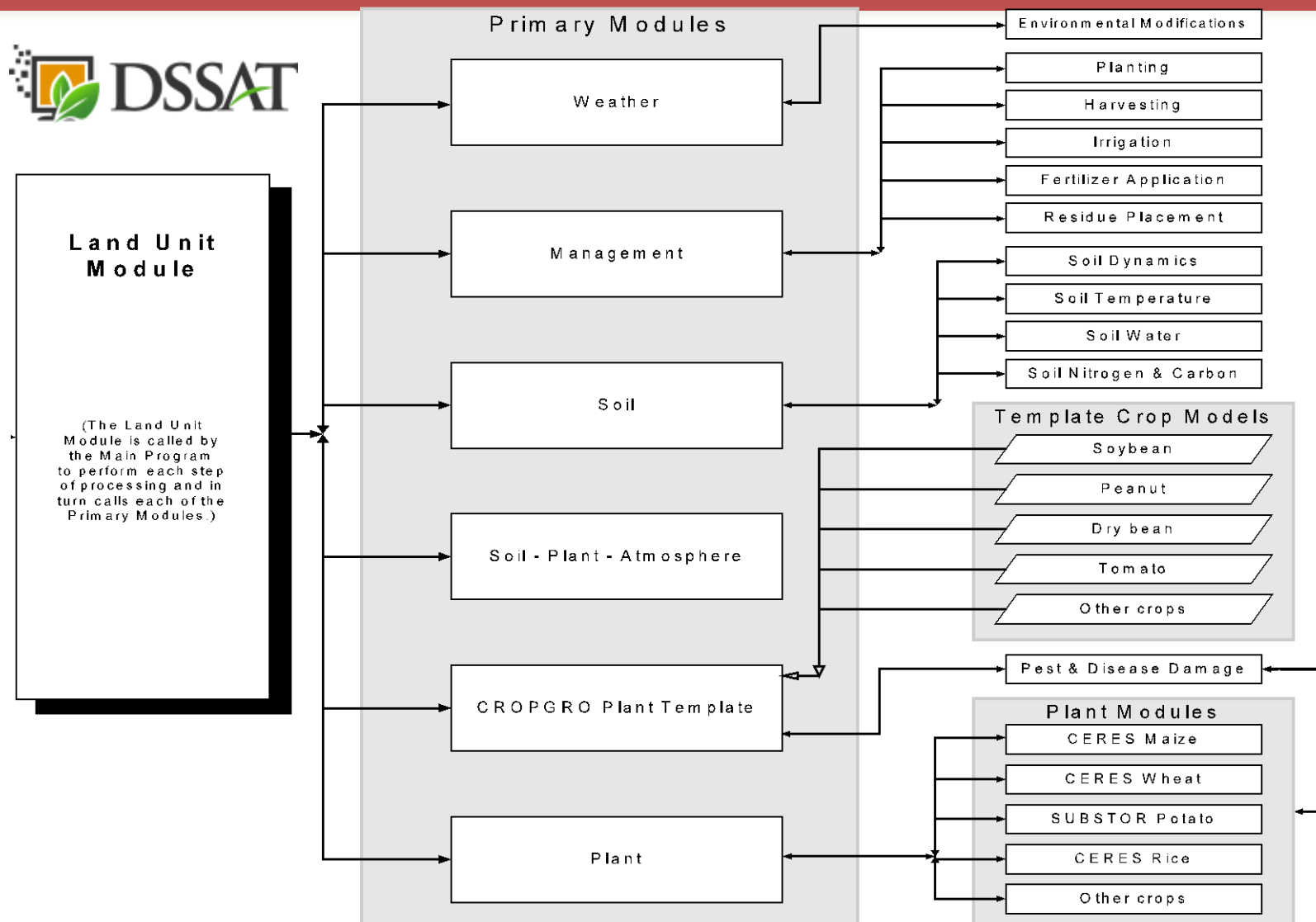
**Runtime  
coordination**

**WINGS/Pegasus/Condor (ACI)**

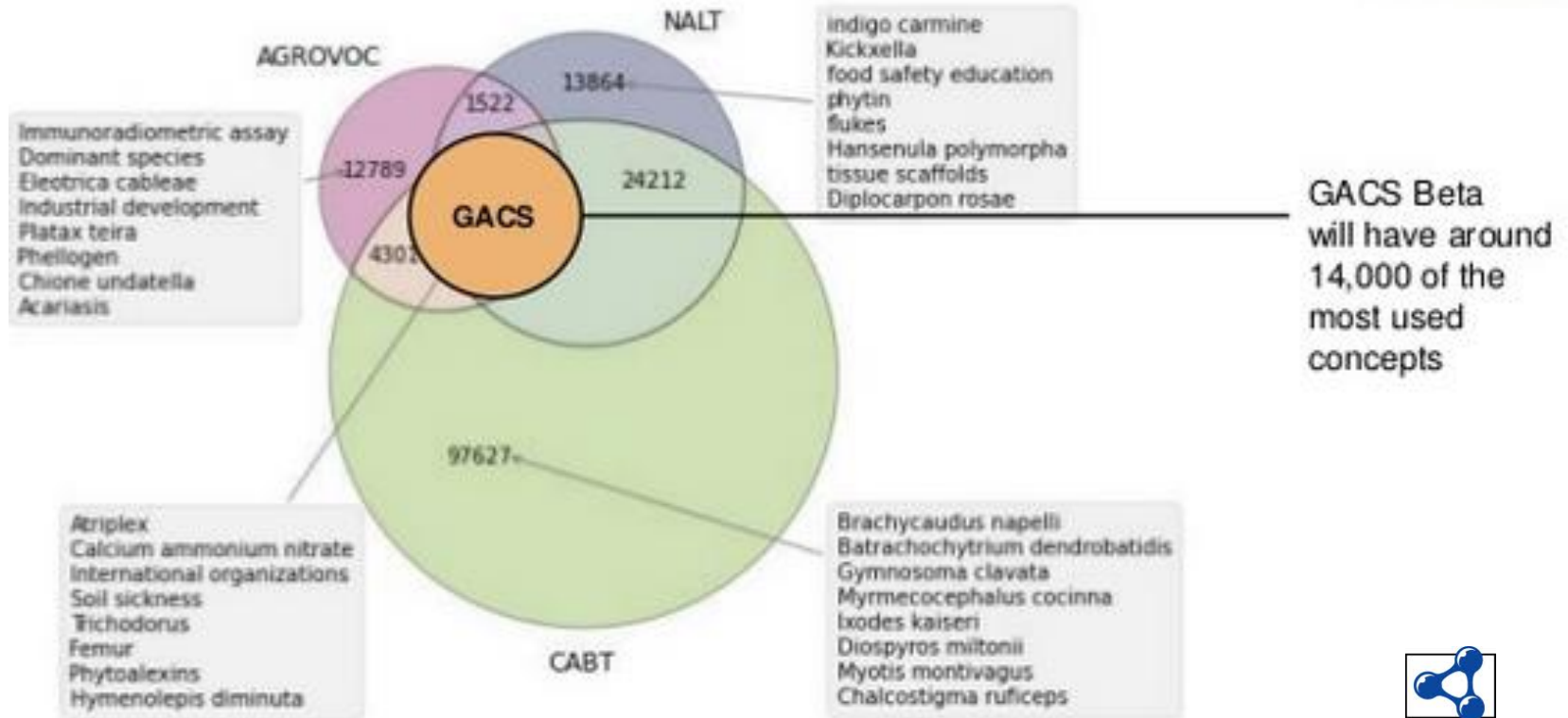
# New Project: MINT (Model INTEgration)



# Agriculture Models: Representative Model



# Agriculture Models: Ontologies of Model Variables





# Social Models: Representative Models

