Domain Vocabularies
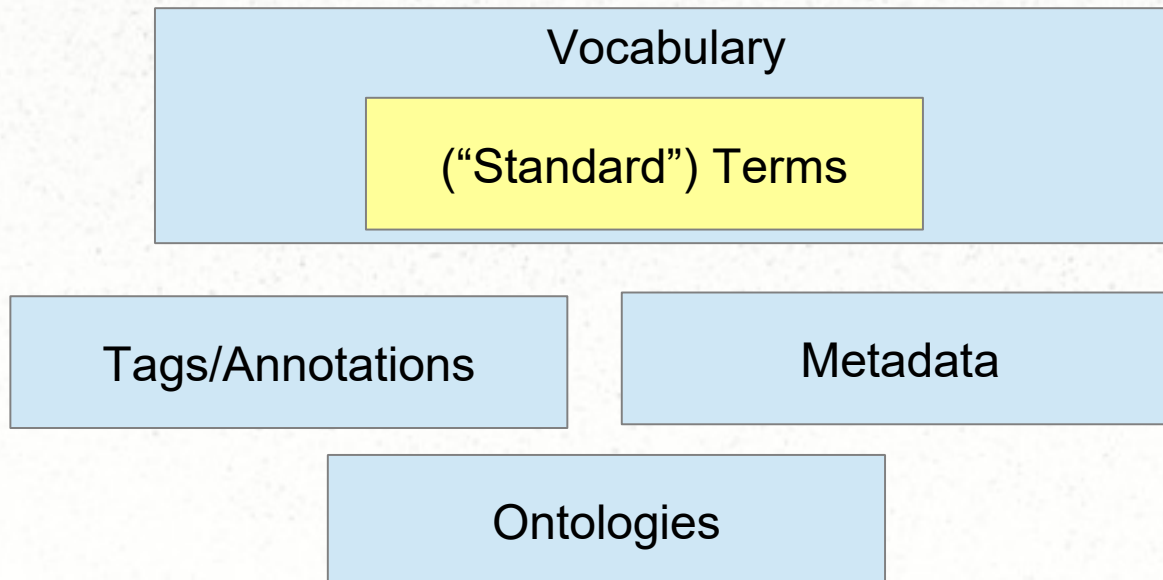
Gary Berg-Cross

DFT IG Co-Chair
Ontolog Forum and RDA US Advisory Committee

gbergcross@gmail.com

Vocabulary

("Standard") Terms

Tags/Annotations

Metadata

Ontologies

Outline

Vocabularies & the heterogeneity problem

Why tackle semantics?

Domain Group Interest & P8 BoF

BoF Follow up

- **Ontolog Forum Session**
- **DcVoCamp with Common Methods**
- **A Domain Vocabulary IG?**

We have vocabularies but where are we?

More **standardizing** of the egistration and management of domain vocabulary is needed - terms help as data/metadata documentation.

Controlled vocabulary used to index & retrieve data resources.

Examples from the CSDMS Standard Variable Names

- atmosphere carbon-dioxide partial pressure

- **atmosphere water liquid equivalent precipitation rate**

- bedrock surface time derivative of elevation

- channel bed manning coefficient

- channel cross-section wetted perimeter

- channel cross-section width to depth ratio

- earth axis tilt angle

- earth ellipsoid equatorial radius

- earth orbit eccentricity

Many things like Objects, roles and attributes are jammed into vocabulary

Things like Agriculture are complicated
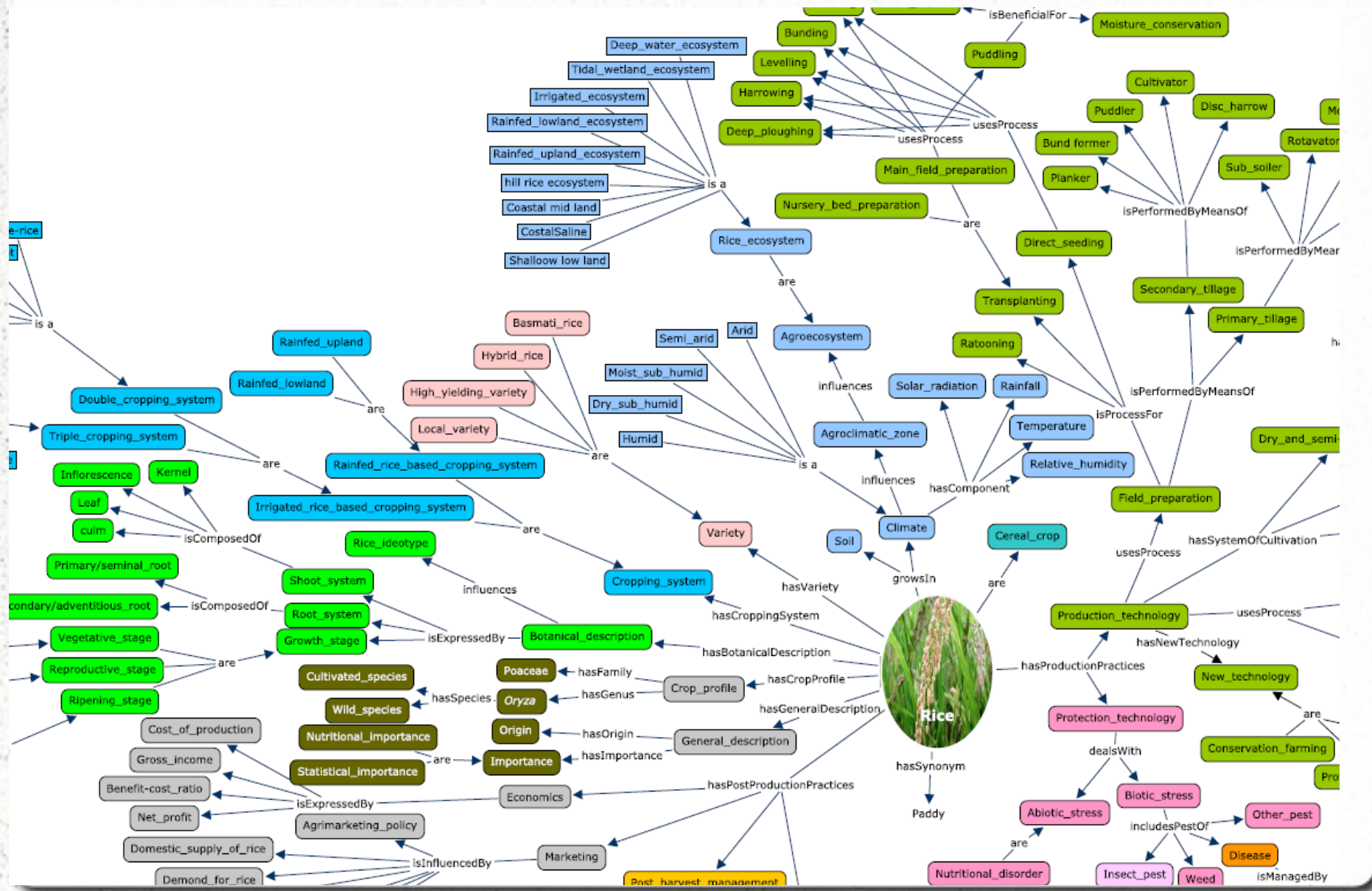
What are the Problems?

Legacy of early terminologies that were often **arbitrary** with little supporting conceptualizations or real standardization.

Domain vocabularies often have the same concept scope; but

- are represented with different terms,

- use different formats and formalisms, and

- are published and stored with alternative access methods.....

Naming schemes often suggest some implied semantics.
"Descriptions are more or less purposeful and theory-laden. Pharmacologists, for example, in their description of chemicals, emphasize the medical effects of chemicals, whereas "pure" chemists emphasis other things such as their structural properties."
Semantics and Knowledge Organization, Birger Hjørland

Linked Ocean Cloud: Mapping Challenges

Table 1. The number of Linked Data relationships between nodes of the Linked Ocean Data cloud as of 20th October 2015, used to generate Figures 2 and 3
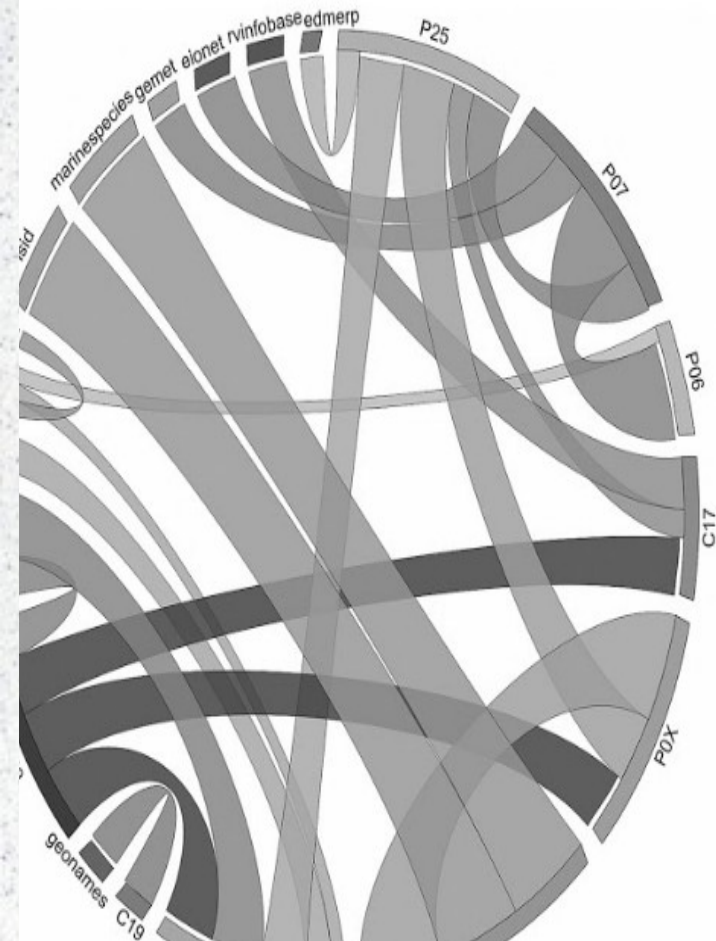
From Node	To Node	Number of Mappings	From Node	To Node	Number of Mappings
NVS-C19	geonames	126	BCODMO-Parameter	NVS-P03	4
NVS-L06	HeritageData	20	BCODMO-Parameter	NVS-P09	3
NVS-L06	MMISW-IOOS-Platform	26	BCODMO-Instrument	NVS-L05	93
NVS-L05	MMISW-TRDI-Glossary	2	BCODMO-Instrument	NVS-L22	33
NVS-L05	CAST	6	BCODMO-Instrument	NVS-C77	1
NVS-L22	MMISW-TRDI-Models	12	BCODMO-Platform	NVS-L06	168
NVS-C38	OrdSurv	205	R2R-Device	NVS-L05	90
NVS-P01	SISSVOC	58	R2R-Device	NVS-L22	24
NVS-P08	SISSVOC	1	R2R-Vessel	NVS-L06	1
NVS-P08	Programmes	10	R2R-Vessel	NVS-C19	42
NVS-P21	SISSVOC	3	R2R-Port	NVS-C38	258
NVS-P06	DBPEDIA	470	R2R-Holding	BCODMO-Deployment	362

General Heterogeneity Issue with Vocabularies

There are some cases with dozens and dozens of local standards for data making integration a challenge

Various types of standards are, for the most part, heterogeneous, meaning they:

-are mostly fragmented and disconnected, describing things like surface or groundwater which:

- lack a foundational grounding.

– use the same or similar terms but with differences in semantics.

(Illustrated in Simon's Key Note that explored the term "soil")

– are described using different formal (or non-formal) languages.

Bottom Line: *We need to systematize the already large body of domain definition work on terms and their meaning using a rationalized "consensus" knowledge of domain experts, especially as involved in RDA's efforts.* **Simple approaches to metadata don't do it.**

Domain Vocabulary BoF Session Agenda

Facilitating Domain Vocabulary Development & Harmonization

- Illustrated Conceptual Models & Tools - Gary Berg-Cross

· Domain Discussion - Each domain group provided a brief summary of their work such as:

· Relevant vocabularies, models & ontologies, standards and

· Community discussion regarding the state of:

- Vocabulary development, formalization, harmonization, infrastructure

- Common interests such as vocabulary development & services,

- Issues and best practice solutions in the domain vocabulary space

General Discussion & ideas in follow up work & plan for follow up virtual meetings and session at P9

Speakers from Specific Domains

Stuart Chalk - Chemistry Research Interest Group
IUPAC "color books" - references in chemistry - large curated set of good 'terms' - maybe it would be good to develop ontology that represents the knowledge in books

Bob Hanisch - Materials Science Registry WG
look across existing vocab, metadata schemas and come up with something the 'right size' for discovery/access in registry (DC not enough)

Mark Fox - quality of urban life interest group

Paul Uhlir - RDA/CODATA Legal Interoperability IG

Ana Kalbert- USGS & CSDMS interdisciplinary model community work

Steve Richard   Earth Science Physical Samples

Ilya Zaslavski – Global Water domain
Starting with use cases you can end up with too much information (30-40 K terms)

# *Example- Attempts at Standard Radiation-Related Names*

atmosphere_clouds_radiation~incoming~shortwave__absorbed_energy_flux
atmosphere_clouds_radiation~incoming~shortwave__absorptance
atmosphere_clouds_radiation~incoming~shortwave__reflectance
atmosphere_clouds_radiation~incoming~shortwave__reflected_energy_flux
atmosphere_clouds_radiation~incoming~shortwave__transmittance
atmosphere_clouds_radiation~incoming~shortwave__transmitted_energy_flux
atmosphere_clouds_radiation~outgoing~longwave__emittance
atmosphere_clouds_radiation~outgoing~longwave~downward__energy_flux
atmosphere_clouds_radiation~outgoing~longwave~upward__energy_flux
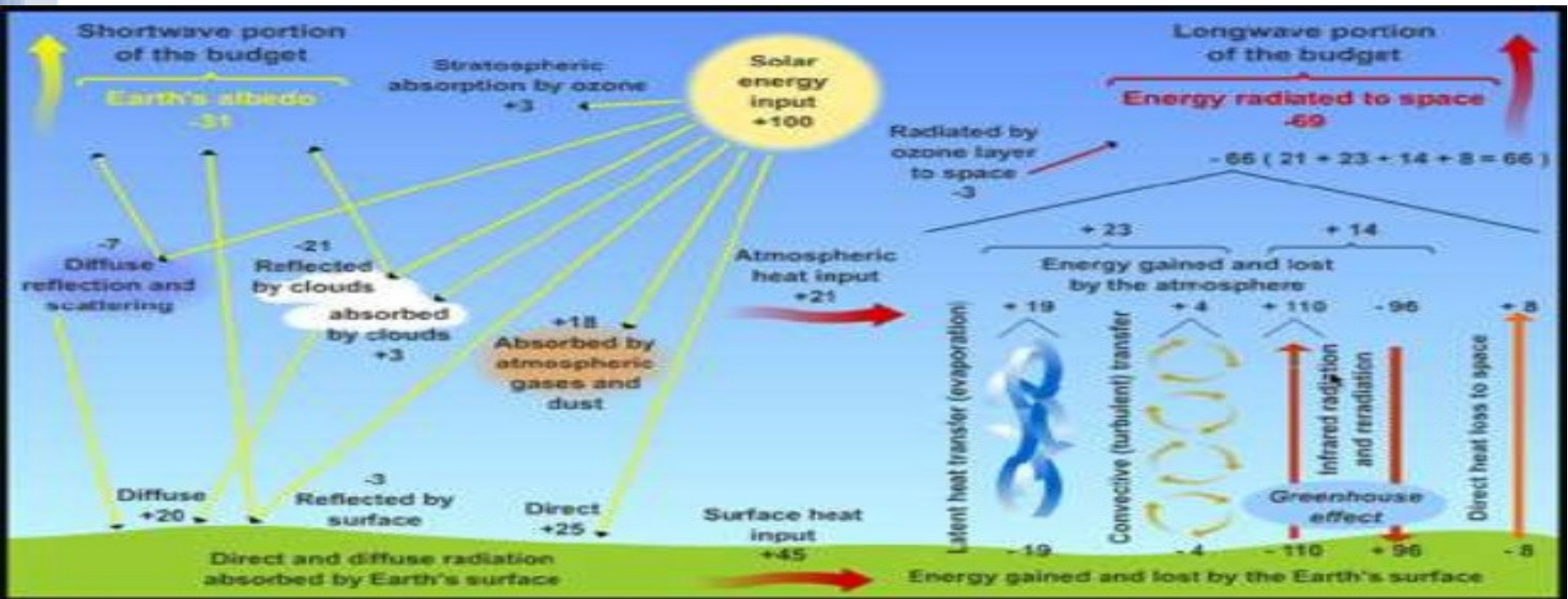
Semanticscience Integrated Ontology (SIO)
for research & knowledge discovery
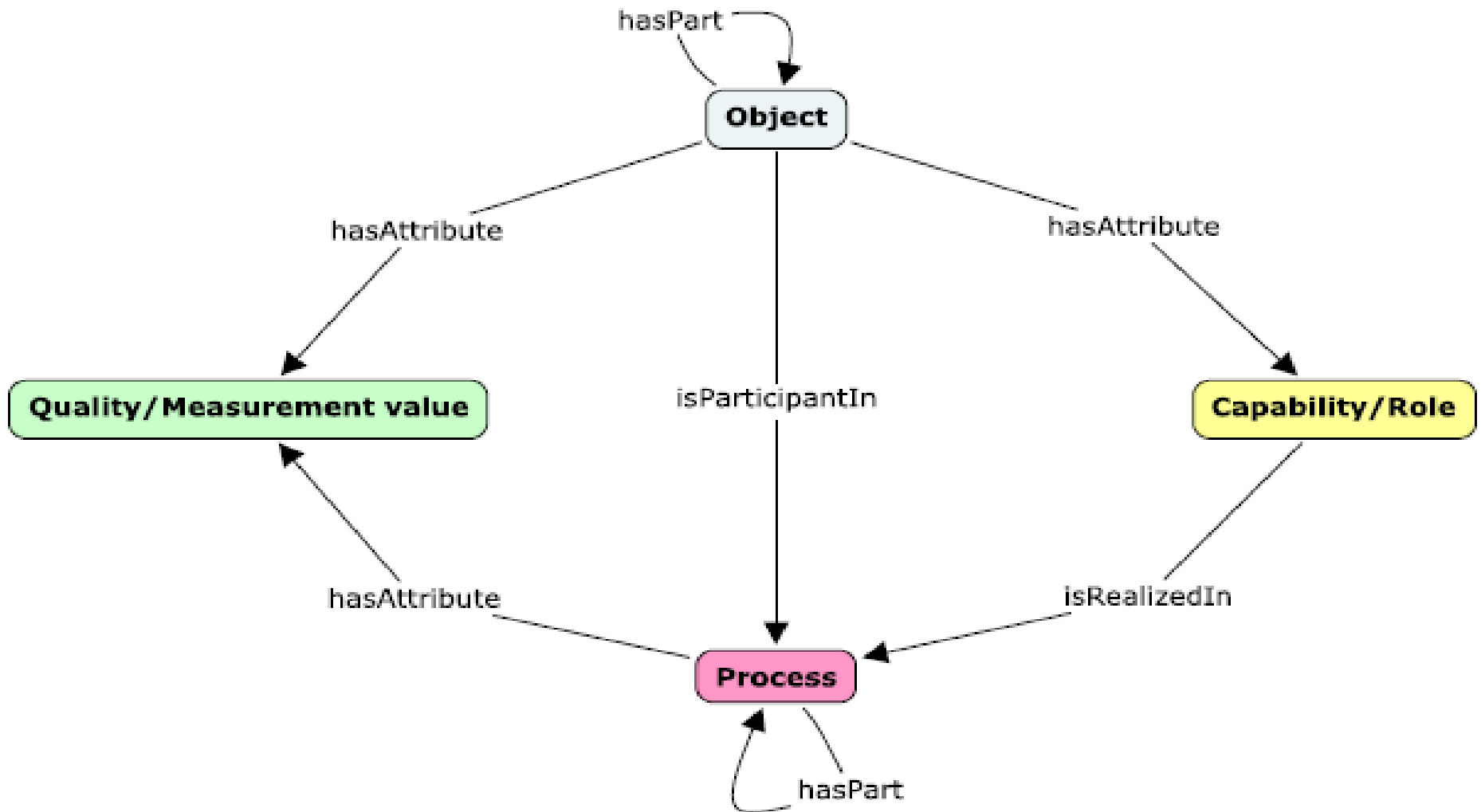
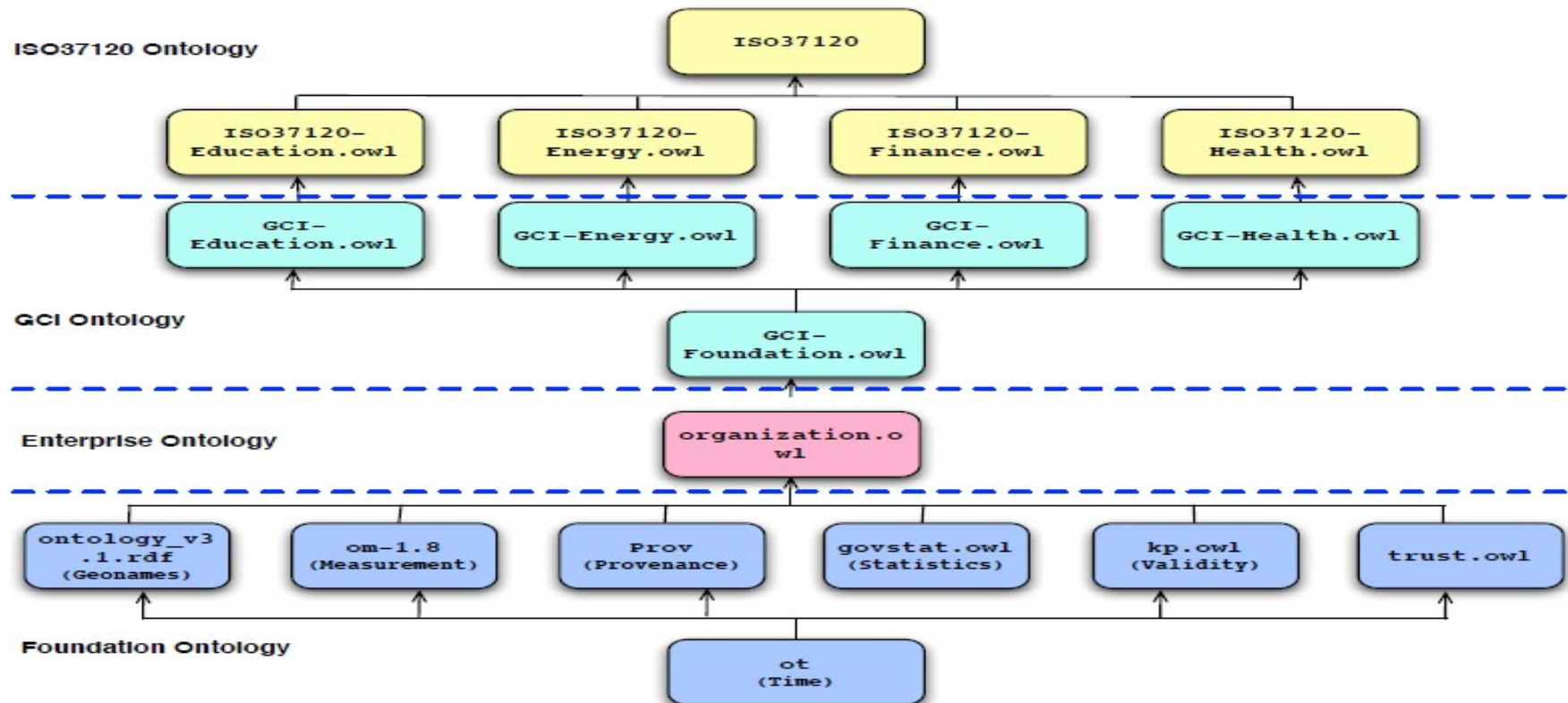*SIO View of CSDMS Model Radiation-Related Names*
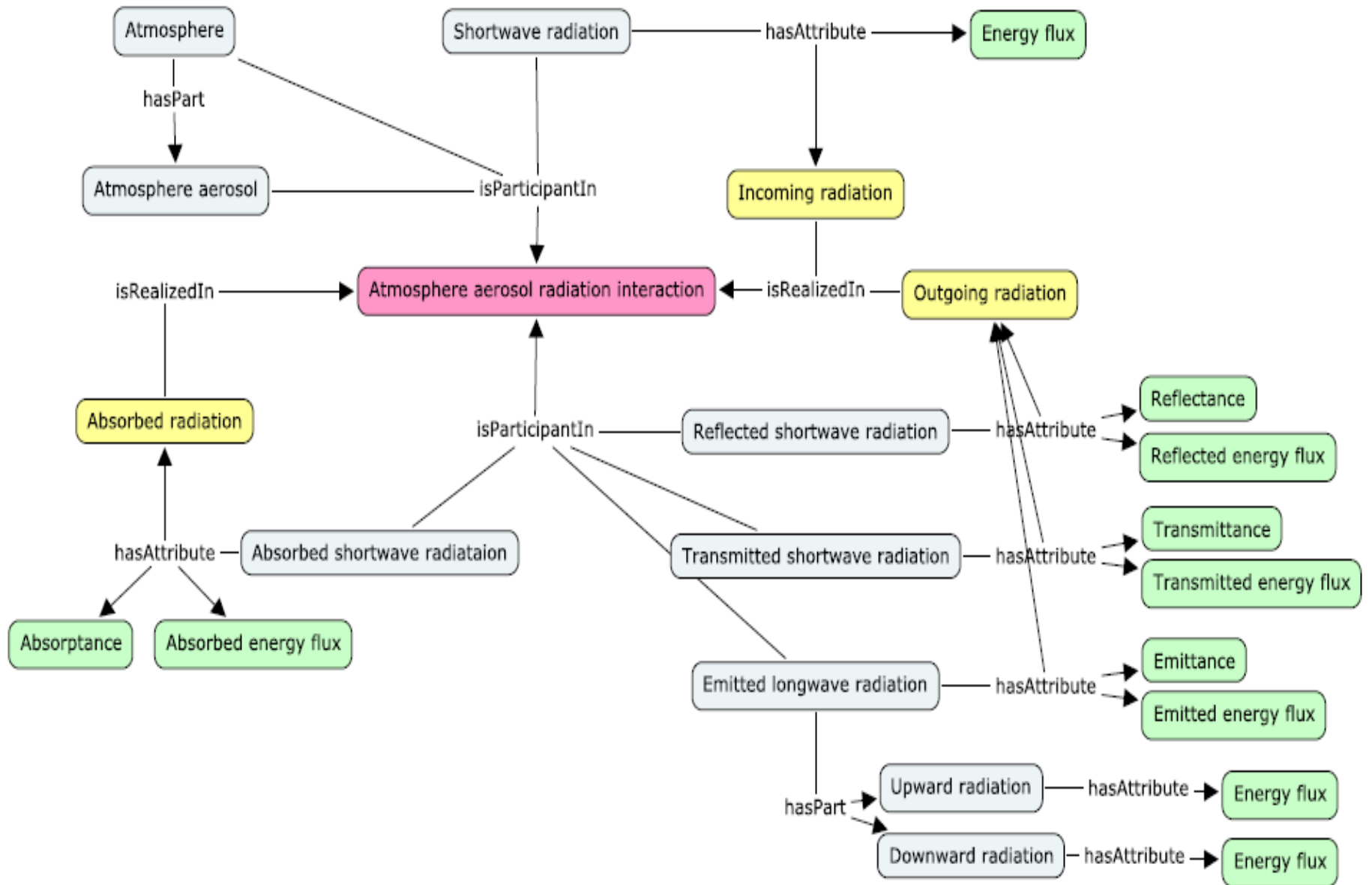
Example From our Bof

Quality of Urban Life IG Summary

Mark S. Fox
University of Toronto

GCI Ontologies Structure

Ontolog Forum

http://ontologforum.org/index.php/DomainVocabularies

Ontology Engineering Steps

Follow up of RDA Domain Vocabulary BoF

1. **Determine domain and scope**
2. Determine the Competency Questions
3. Consider reusing existing ontologies
4. Enumerate important terms
5. Identify classes and structure as a taxonomy
6. Define classes using properties
7. Define instances
8. Validate using competency questions

Using SPARQL.
SELECT ?event
WHERE {mixture, hasIngredient, ?p, ?p, hasCondition, ?hazardous}.

© M. S. Fox & M. Musen 2016

9

VoCamp *http://vocamp.org/wiki/GeoVoCampDC2016* *Topics*

Topography focused topic - basic terrain primitives, slope, slope length, shape, curvature (Lynn Usery & Gaurav Sinha)

Modeling Terminology *for* Chemical Safety: RDF vocabulary for Chemical Terminology (Leah McEwen)

A pattern to support Materials Research vocabularies (Kimberly Tryka, Alden Dima & Chandler Becker)
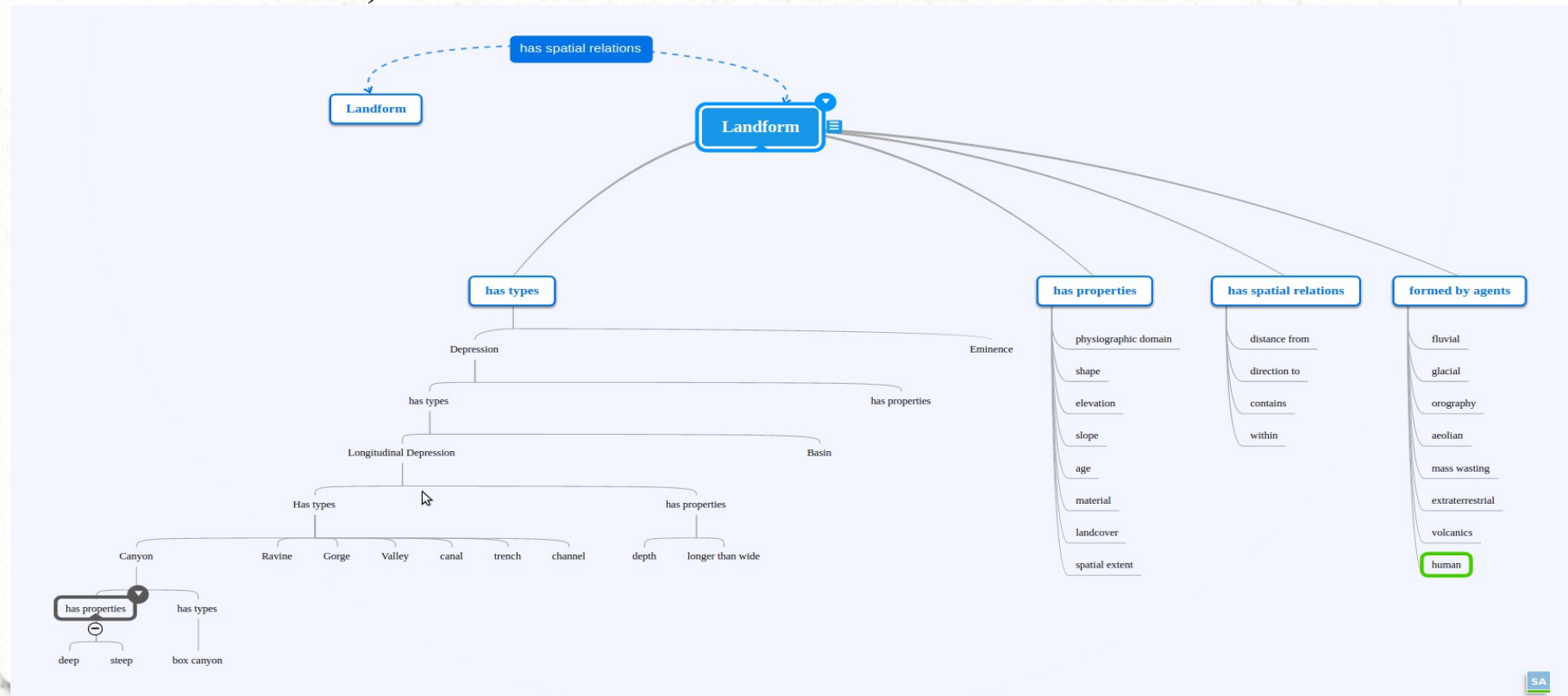
Light Ontological Engineering Vocamps

- Used to clarified agreement & reduce ambiguities/conflicts on domains (e.g. geospatial, chemical, materials & related phenomena)

Ontology Engineering

Problems, Vocabulary Components, Relation Identification & Clarification

Lightweight Methods: develop general, reusuable patterns that provide a common, modular framework to generate ontologies that are consistent and can support interoperability.

Conceptualization Phase
Agreements can then be formally represented in: Constrained, engineered models - Systematic organization & framing with visual expression.
•**Minimal encoding bias**: No assumptions about knowledge representation

Existing Voc & Ontologies

We like data-grounded work with interoperability in mind

Formalization

Work in Progress