

Methodological Notes: DC-Vocamp 2016

Gary Berg-Cross

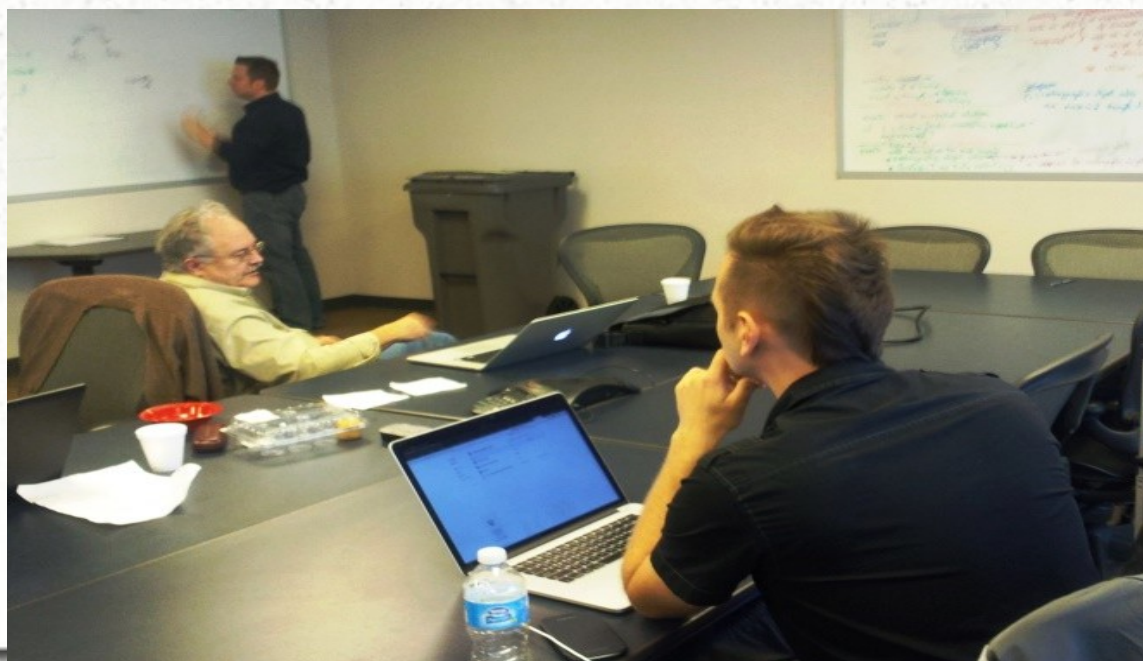
DCVoCamp Organizer

Ontolog Forum and RDA US Advisory Committee

gbergcross@gmail.com

<http://vocamp.org/wiki/GeoVoCampDC2016>

Tweet: #DCvocamp16



Outline

How to start

Teams

Ontology Engineering Steps

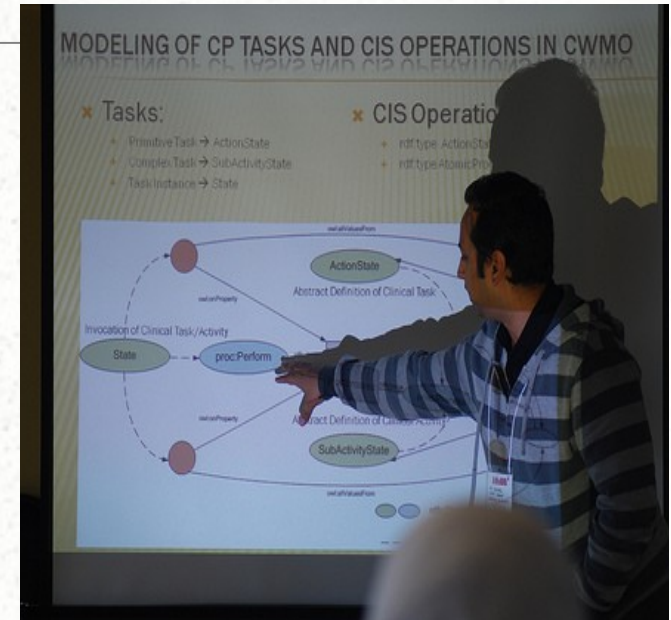
Lightweight methods

Problems, Competencies, Vocabularies,

Relation Identification & Clarification

Conceptualization, Patterns,

Reference Ontologies & Formalization



Interdisciplinary

Minimizing weak theory
vs
precise commitment

Anticipating various uses
vs
maximally general
ontology

Domain/Data Expert

Trade-offs

KE/Ontologist

Note taker

Tools...

Facilitator

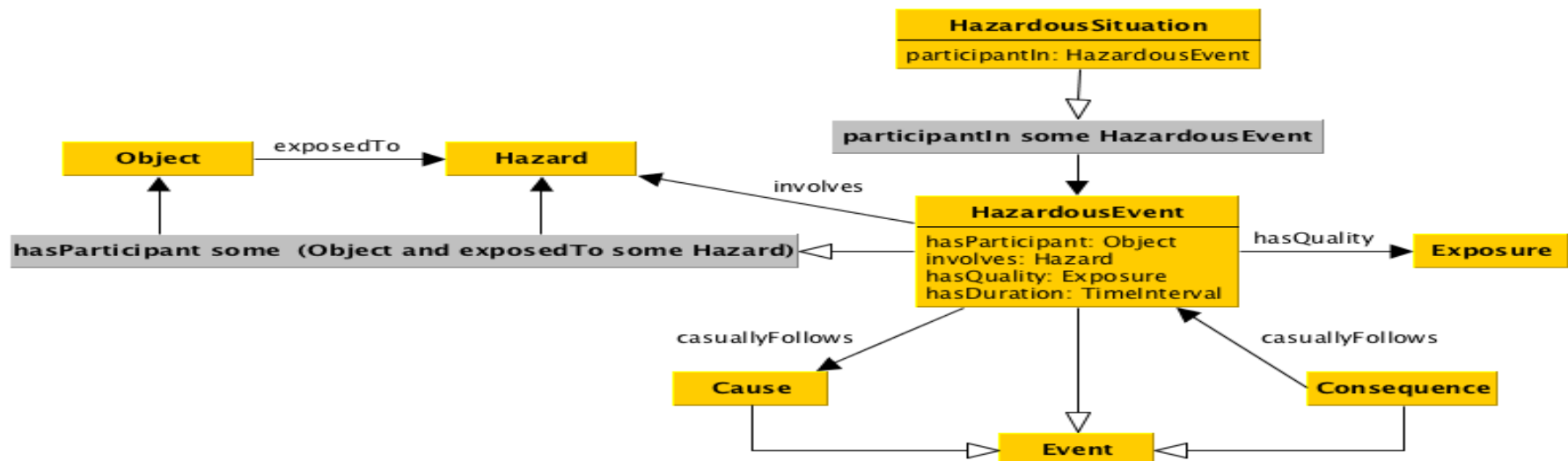
dialogue
system
ontology
based
information
application
architecture
view
natural
present
speech
evaluation
communication
approach
proceedings
search

Tools – drawing first?

It is often useful to start with hand/board drawings to accommodate conversational flow.

PowerPoint or Yed graphics can be used to tidy things up for presentation.

For some it is a modeling tool like CMAP with support for model constructs and automatic translation into OWL/TTL etc.





Ontology Engineering Steps

From 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}.

Lightweight Methods

Organize (Domain Scenarios for Scope) and Look for Low hanging fruit

For example, leverage initial vocabularies (e.g. Hazard Condition) and existing **conceptual models** to ensure that a semantics-driven framework is available for use in **early stages of work**

What are we talking about? What do you mean when you use these words..

Certain hazardous events, equipment, and conditions are associated with a hazardous situation... (are these the “setting?”)

We should **leverage** existing work but not slavishly

E.g. reference or include relevant data supporting vocabularies/ontologies,

Terminologies are **starting point as is real data**, on a path to the **concepts** and design behind what the terms mean to domain people and which are relatable to “data.”

The Chemical Safety/Hazard Problem

- We would like to have a framework with richer metadata using standard terms that are semantically expressed in RDF.
- Then we could query them.
- Focus on substance relationship to **outcomes and various conditions.**
- A **challenge** is that there are many terms and many alternative ways of creating/specifying a conceptual design.

Enumerating Chem Safety Concepts

often taxonomies that can't answer Qs

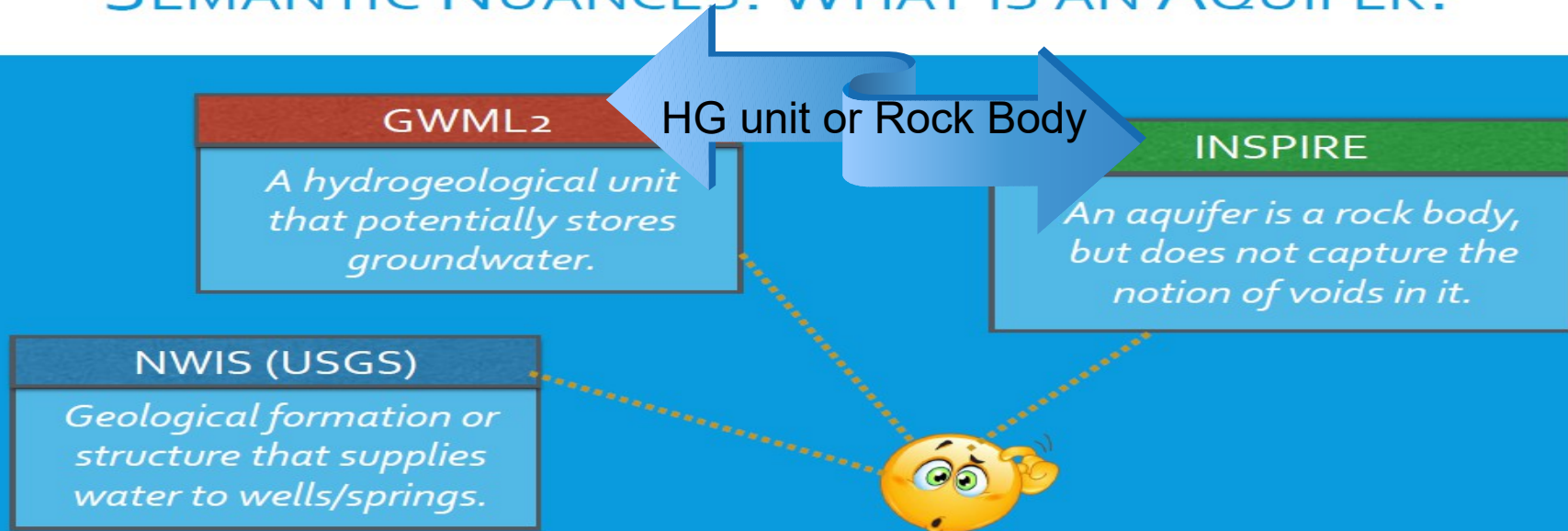
- a. Substances OR Substance gps (reducing reagents, combustibles)
- b. Process/operation (oxidize, ignite, activate, block?)
- c. Condition(s) (temperature, pressure, scale, etc.)
- d. Apparatus/equipment (glassware, tank, etc.)
- e. Concentration or Form/phase (of substances)
- g. Outcome(s) Events (responses/realization: explode, reacts violently)
- h. A disposition is a causal property that is linked to a realization, i.e. to a specific behavior **which the individual object that bears the disposition will show under certain circumstances or as response to a certain stimulus (trigger)**. E.g. Flammable. (How do we ground this?)

? Trigger....Probability?? What granularities????

Vocabularies can have conflicts

Clarity: *Definitions should be objective and complete*

SEMANTIC NUANCES: WHAT IS AN AQUIFER?



- These subtle differences are currently *not formalized*
 - typically only available in narrative form
 - prevents use of automated integration (e.g., ontology alignment) techniques
- How do we reconcile these semantic differences?

7

From “ DOMAINREFERENCEONTOLOGIESVS. DOMAINONTOLOGIES ,Torsten Hahmann, UMaine

Hazard Classification

For each chemical, the chemical manufacturer or importer shall determine the hazard classes, and where appropriate, the category of each class that apply to the chemical being classified...Chemical manufacturers, importers or employers classifying chemicals shall identify and consider the full range of available scientific literature and other evidence concerning the potential hazards.

Appendix A to §1910.1200 for classification of health hazards, and
Appendix B to §1910.1200 for the classification of physical hazards.

MSDS HyperGlossary is built from MSDS terms and is a professional safety community accepted and well used tool. It also has many cross-links (i.e. relationships) among the terms.

From <http://www.ilpi.com/msds/ref/hazardclassification.html>

Hazard Topic Competency Questions:

What Qs can the K-rep answer or tasks support?

What substance was a (person or organization or equipment etc.) exposed to?

What conditions are hazardous for this (X) setting/combination of substances? What conditions are standard? Necessary conditions.

What concentrations are hazardous for this (X) combination of substances?

What forms of (X) combination of substances are hazardous?

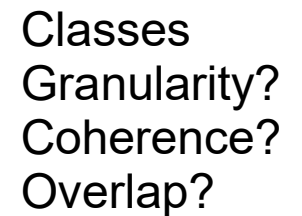
What is the cause of a hazardous event? What trigger types exist?

What are the outcomes of a hazardous event?

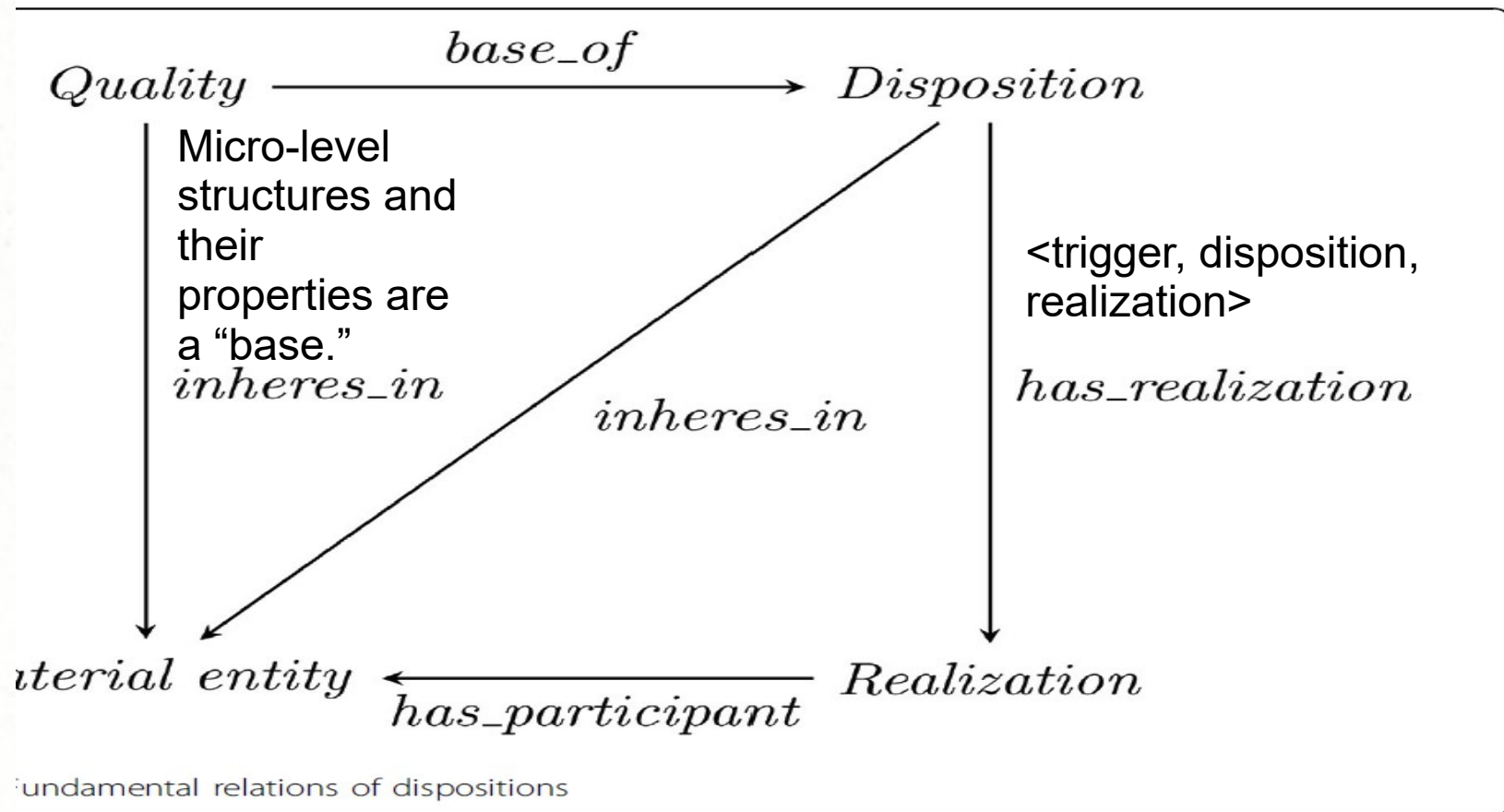
What types of processes are involved in a hazardous event?

Provides Scope
Useful to organize in a scenario

Exposure: The subjection of an object to a particular amount (form?) of a hazard.

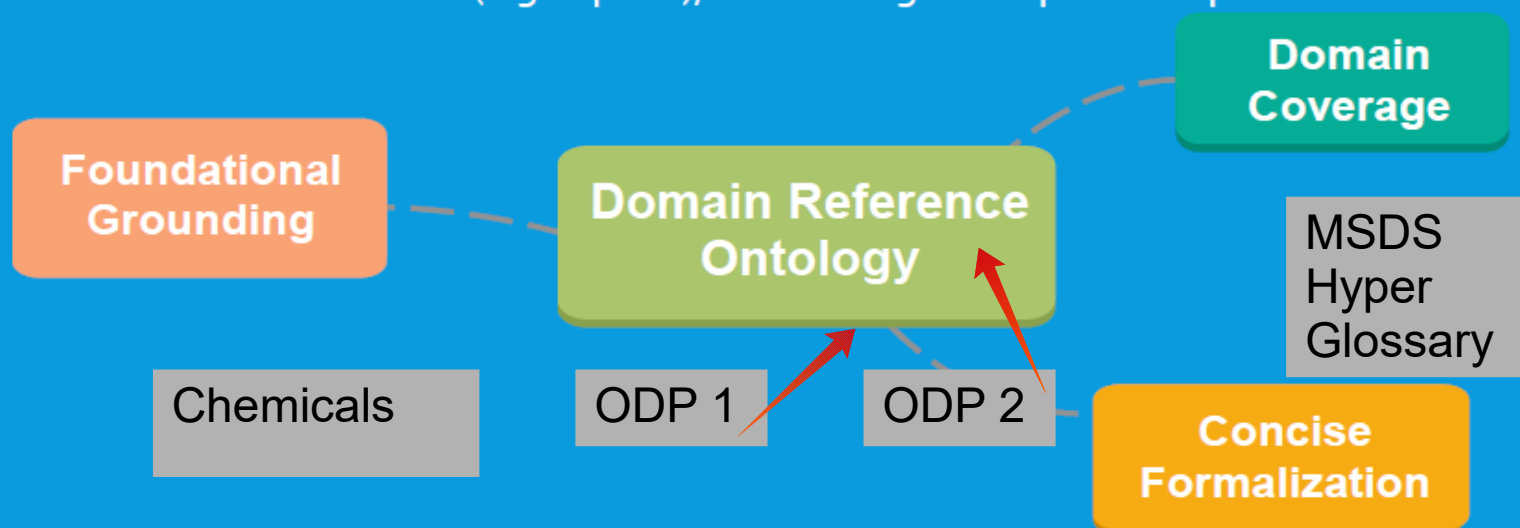


Leveraging Existing Work? A Disposition Pattern: Minimal ontological commitment to say no more than necessary



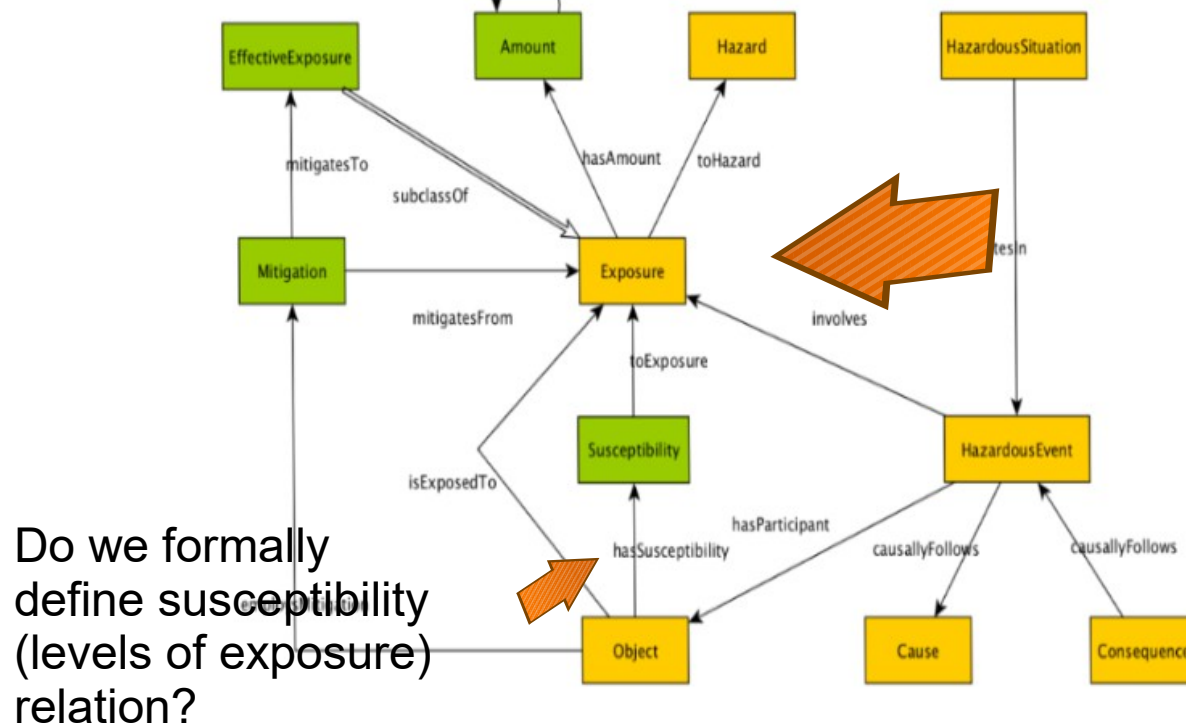
WHAT DOES A DOMAIN REFERENCE ONTOLOGY LOOK LIKE?

- Exhibits many characteristics of foundational ontologies: *foundational for their domain*
 1. Foundationally grounded
 2. Broad coverage on the highest level in the domain: focuses on the **key concepts and relations** in the domain; but does not aim to capture the domain comprehensively
 3. Specified in a highly expressive and fully machine-interpretable ontology language
 - Provides “neutral” language to express semantic differences; Purpose is not to directly define the scientific terms (e.g. aquifer), but ontological helper concepts and relations



Formalization (constraining the conceptual/data model)

HazardousEvent(he) exposure(e)^hasP participant(he; o)^isExposedTo(o; e)



Where do event or exposure subtypes end and instances start?

A hazard must involve at least 1 substance, one condition and 1 process. (?)

Cheatham et al. *unpubli*

Discussed in A Modification to the Hazardous Situation ODP to Support Risk Assessment and Mitigation, Michelle Cheatham, Holly Ferguson, Charles Vardeman II, and Cogan Shimizu