# *Ontology Design Patterns*:

## Bridging the Gap Between Local Semantic Use Cases and Large-Scale, Long-Term Data Integration


Adam Shepherd

EGU 2015-8413

Co-authors: R. Arko, A. Krisnadhi, P. Hitzler, K.Janowicz, C. Chandler, T. Narock, M. Cheatham, M. Schildhauer, M. Jones, L. Raymond, A. Mickle. T. Finin, D. Fils, S. Carbotte, K. Lehnert

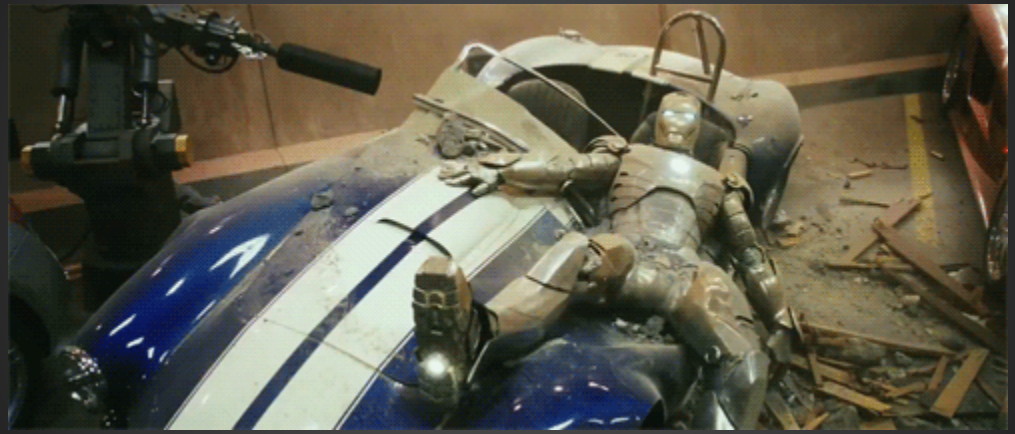
Biological and Chemical Oceanography Data Management Office
Woods Hole Oceanographic Institution


www.geolink.org
schema.geolink.org

BCO-DMO
Biological & Chemical Oceanography Data Management Office

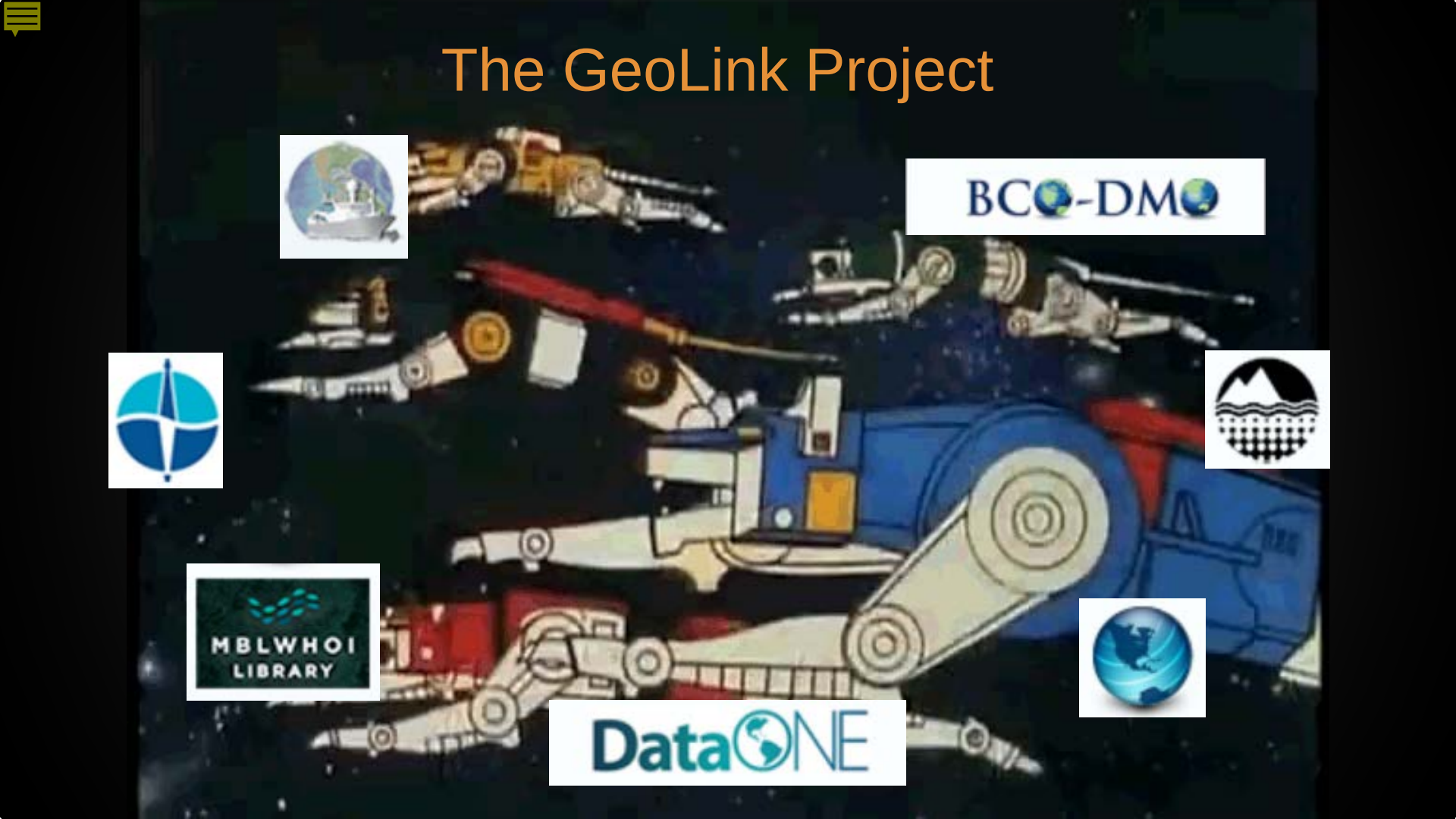

“Come and find data, **my community**…”

# The Grand Challenges of Science

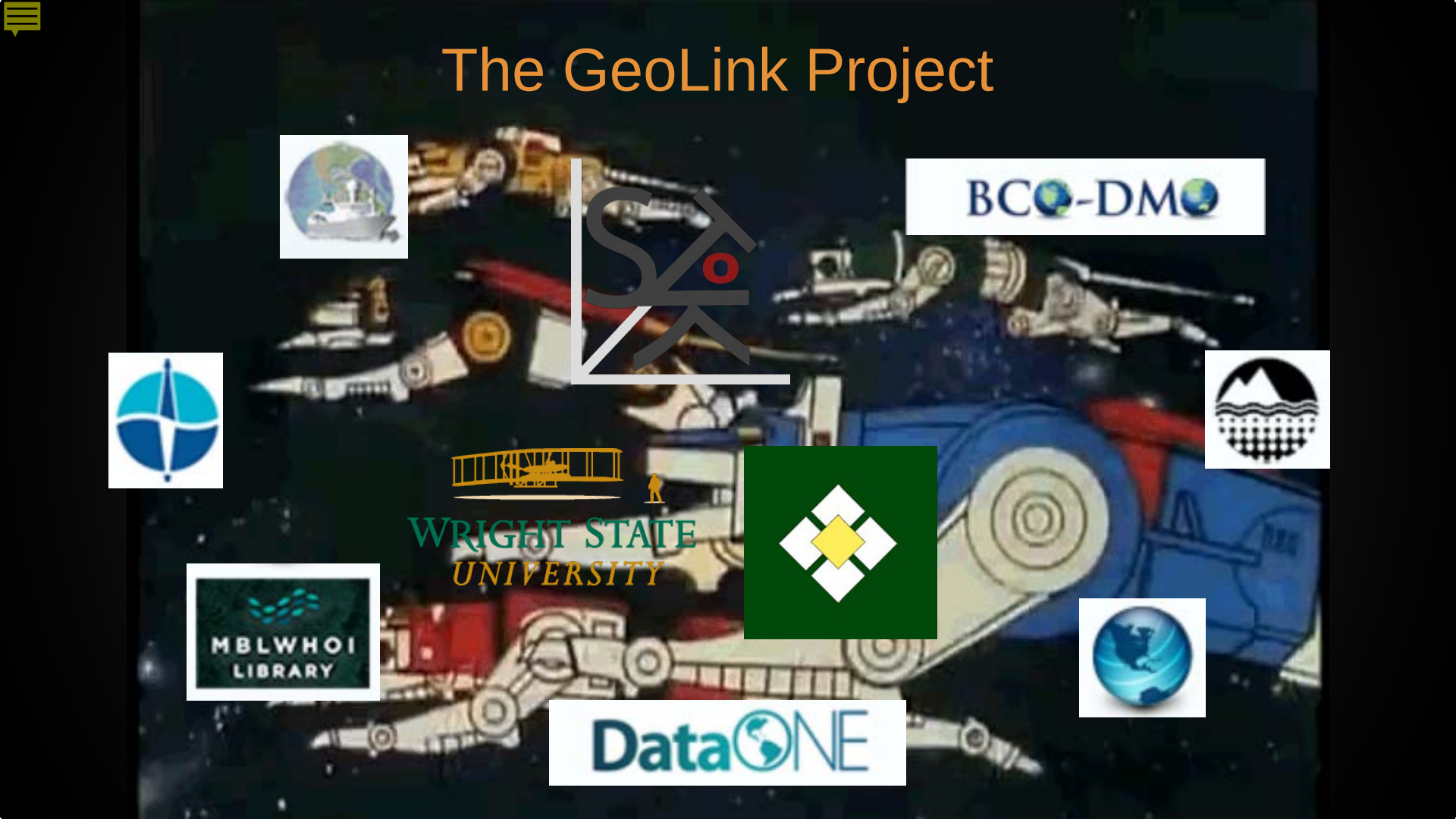


# The GeoLink Project

# The GeoLink Project



# The GeoLink Project



# Perspectives are going to differ...

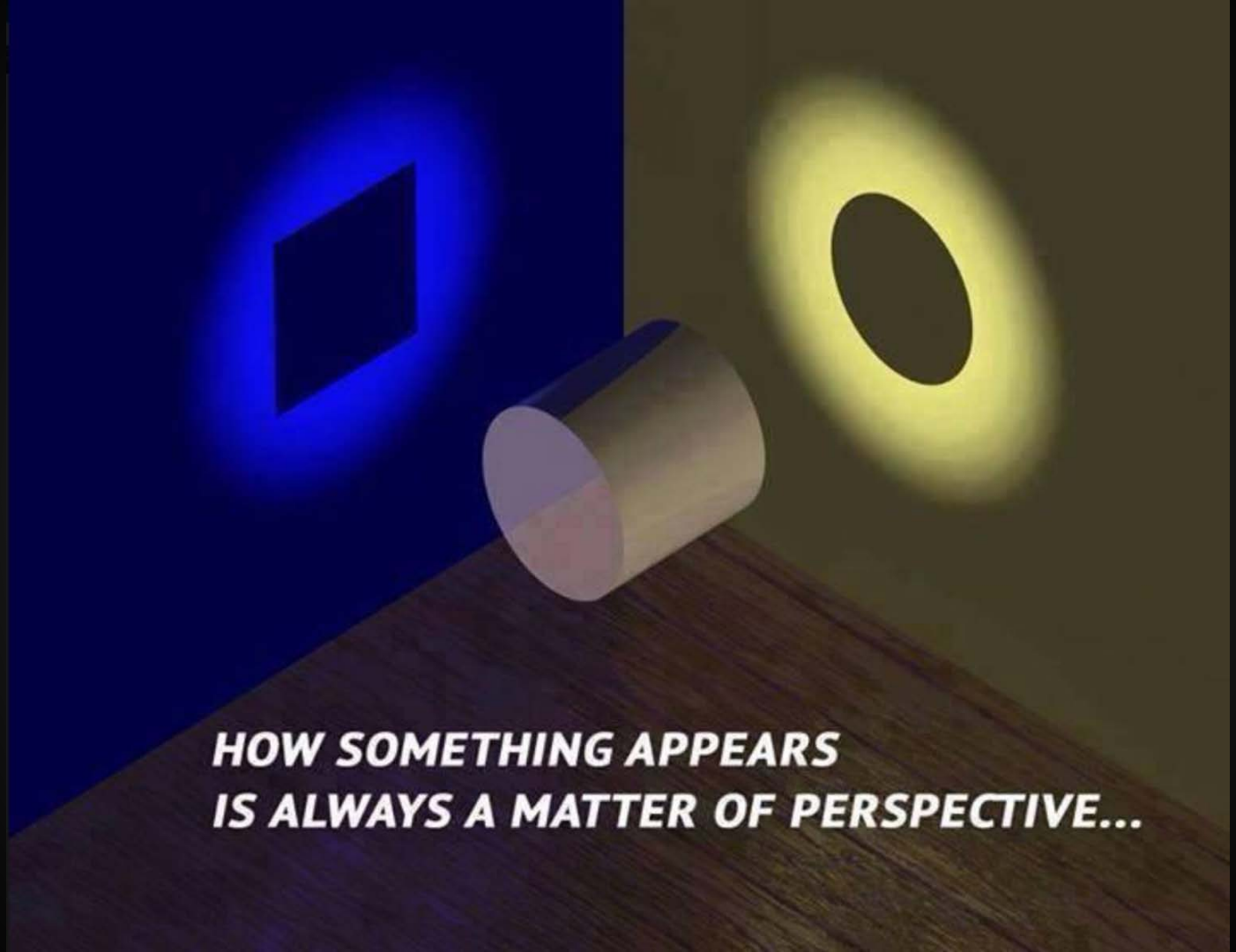
Marine Geology

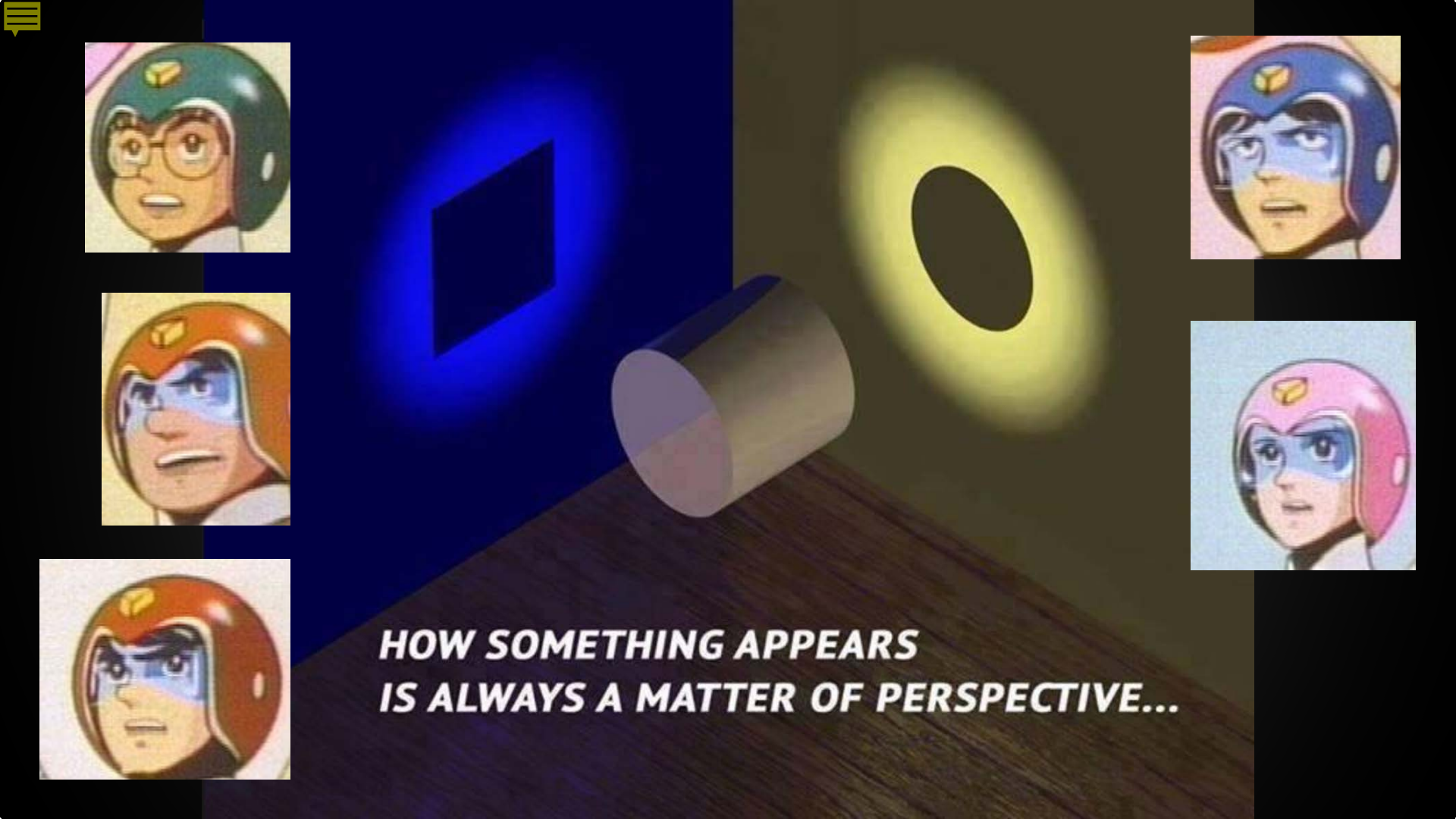
Marine Ecosystems

Publications

Biogeochemistry

Paleoclimatology


HOW SOMETHING APPEARS
IS ALWAYS A MATTER OF PERSPECTIVE...


HOW SOMETHING APPEARS
IS ALWAYS A MATTER OF PERSPECTIVE...

For large scale integration,

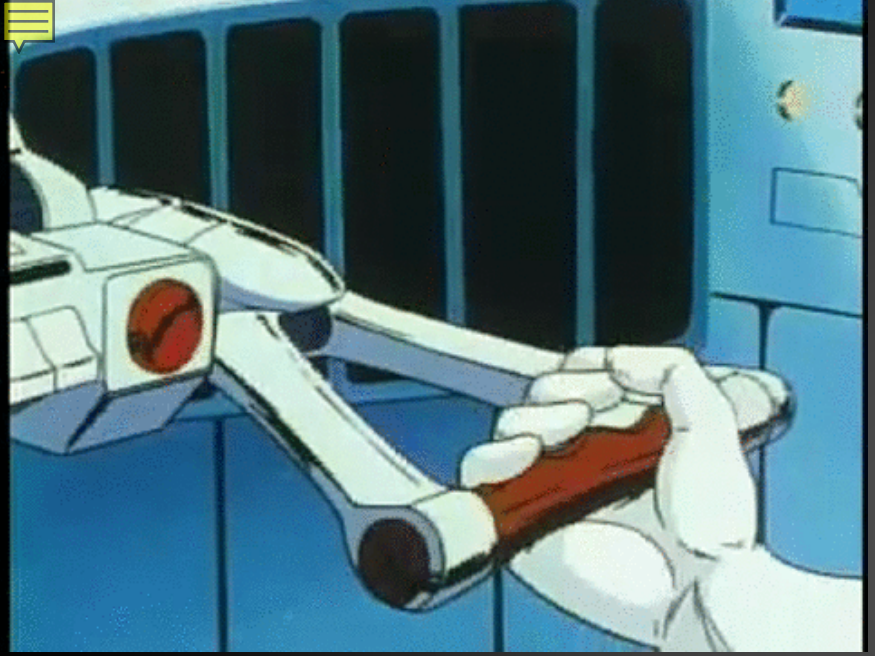
ontological commitments *can be*

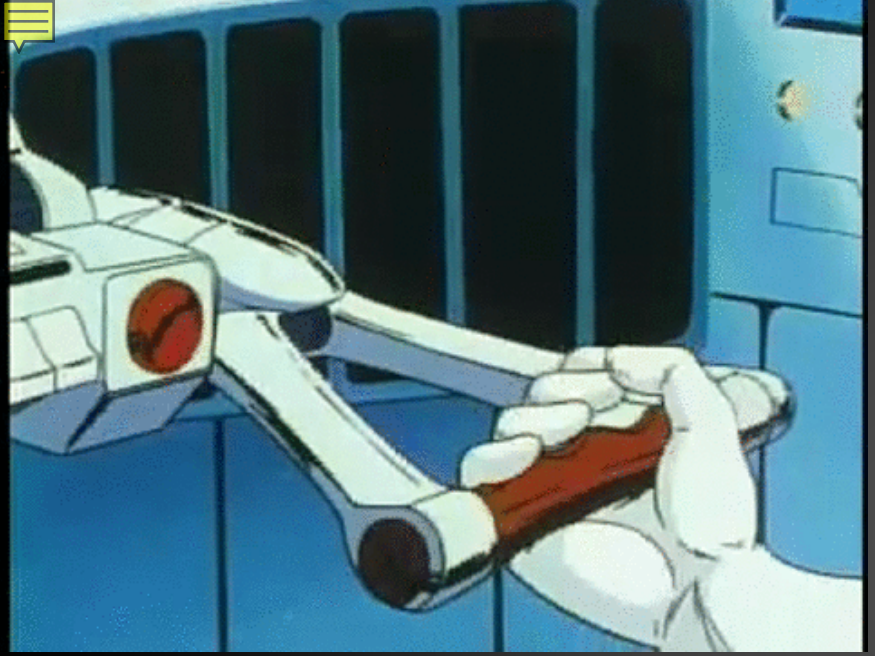
too restrictive...

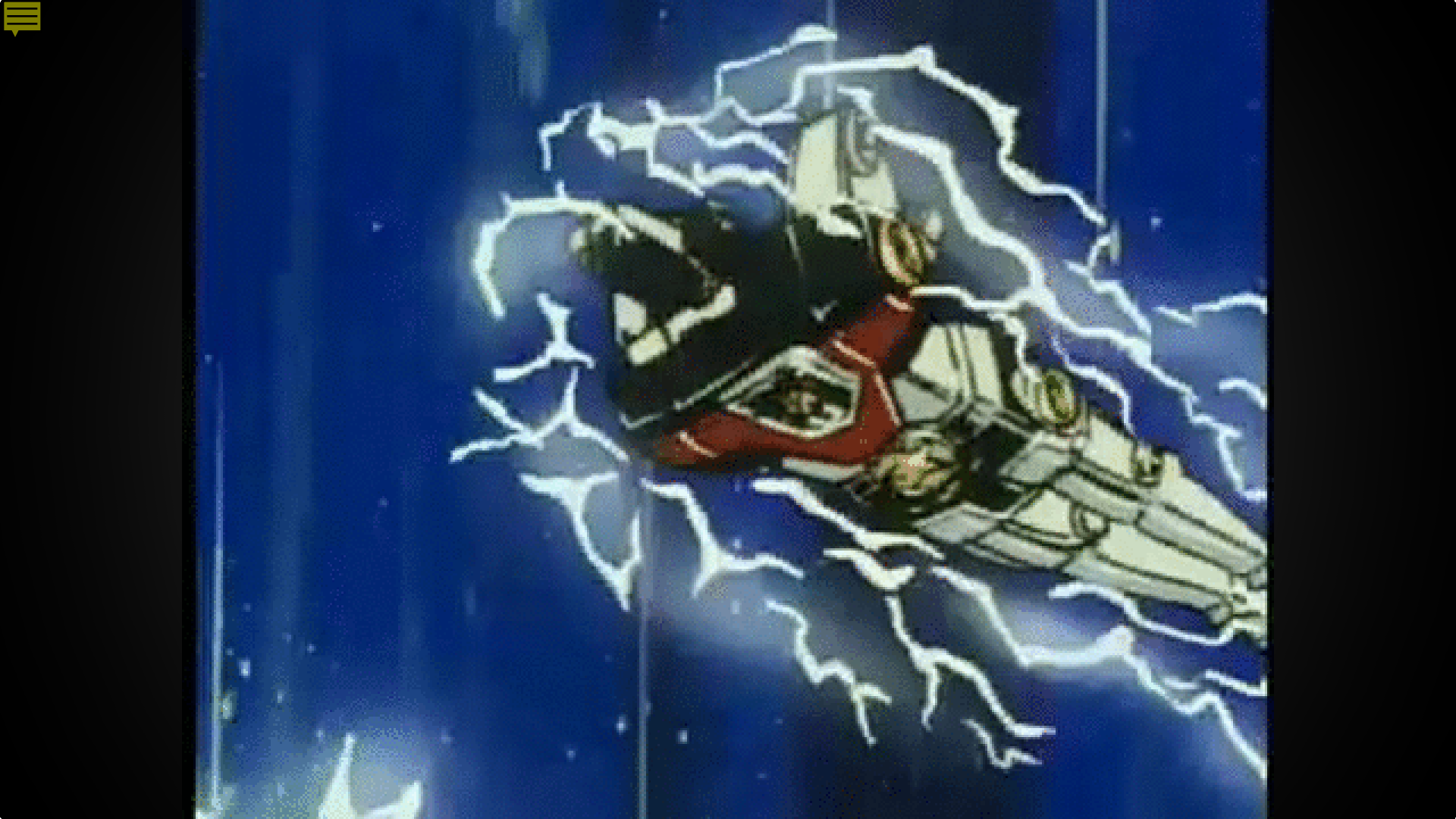
For large scale integration,

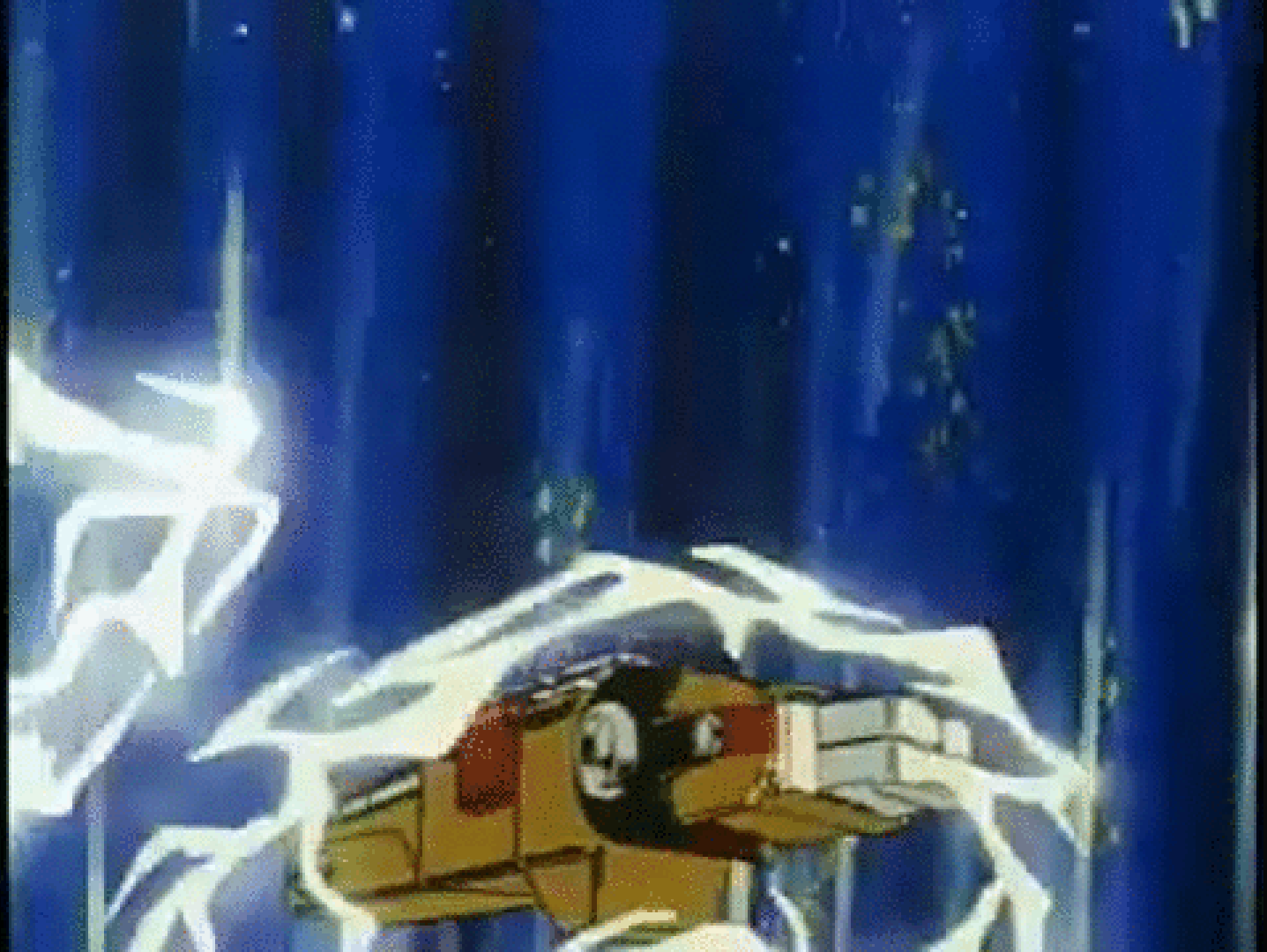
ontological commitments *can be*

not restrictive enough...









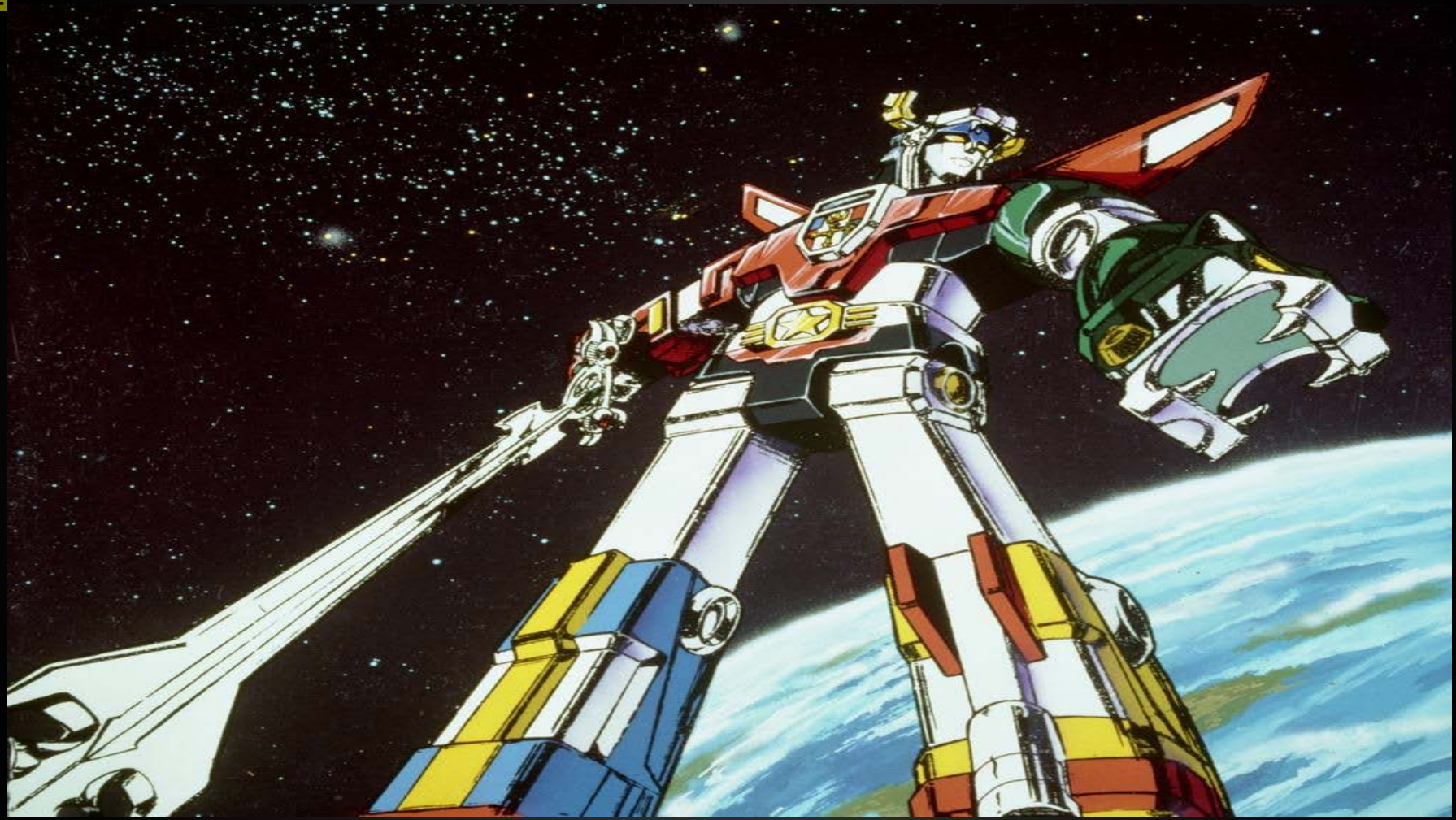
Paleoclimatology

Marine Geology


Publications

Marine Ecosystems

Biogeochemistry



# *Ontology Design Patterns*:

**Bridging the Gap Between Local Semantic Use Cases and Large-Scale, Long-Term Data Integration**

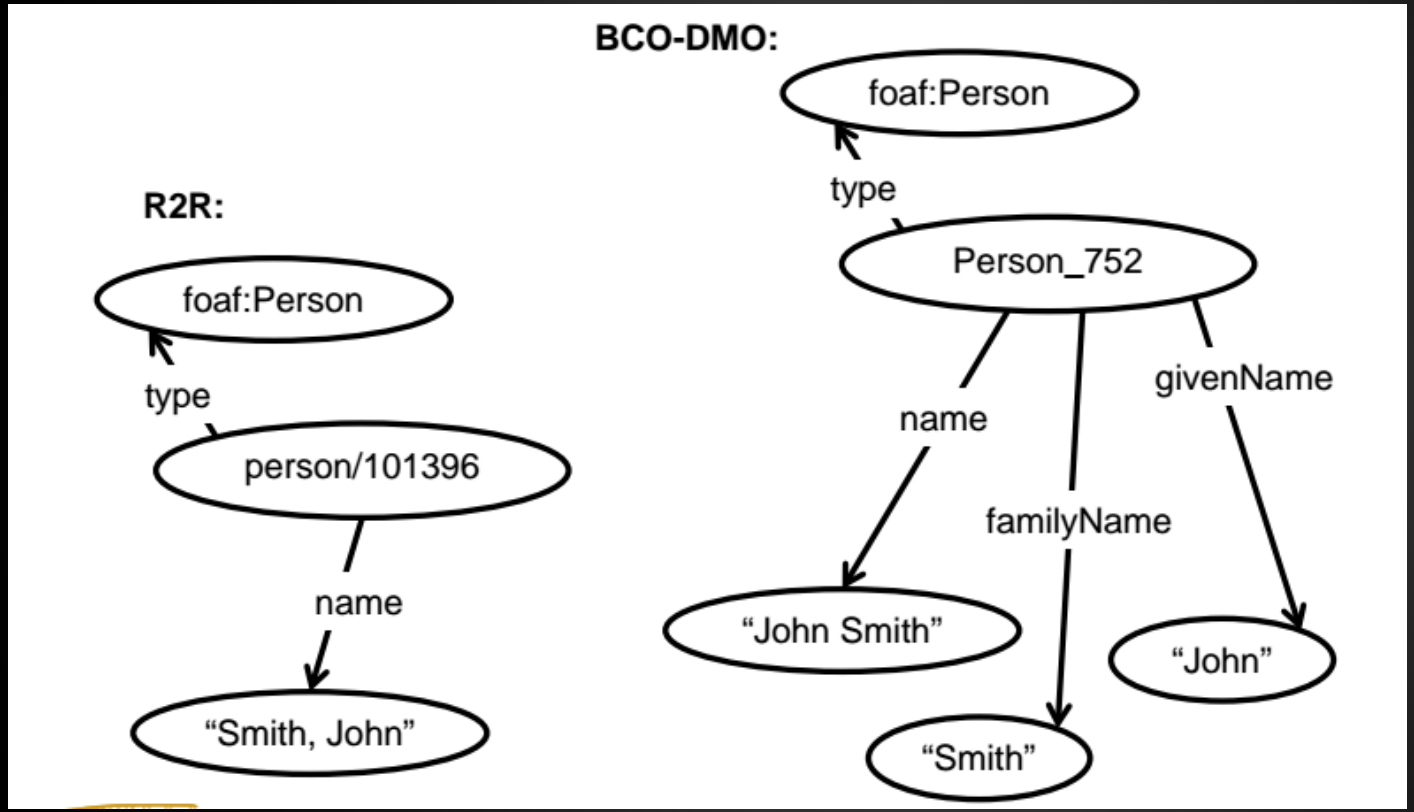
Adam Shepherd

EGU 2015-8413

www.geolink.org
schema.geolink.org

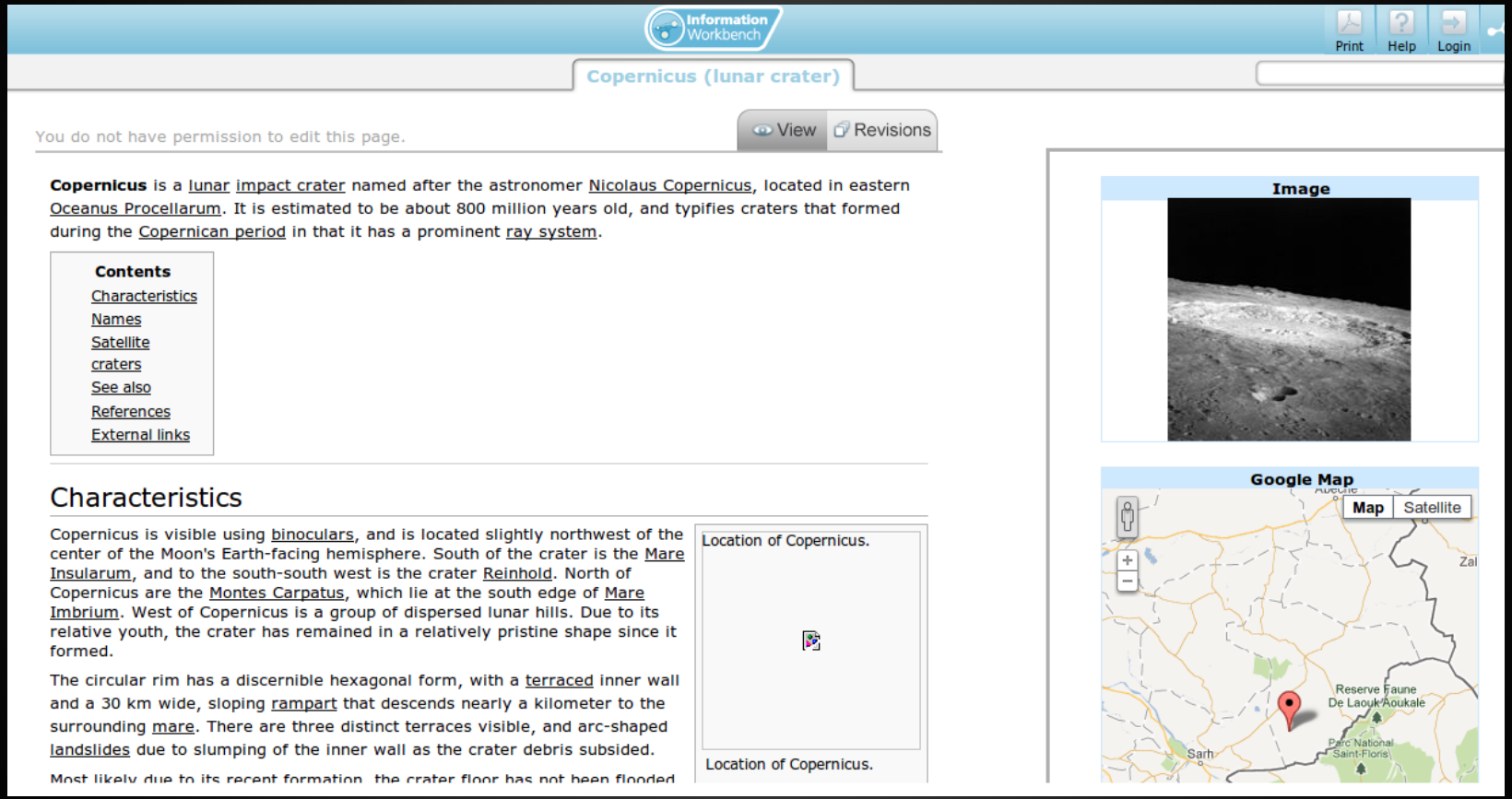

BCO-DMO
Biological & Chemical Oceanography Data Management Office

# Linked Data is Not Enough: *Problematic Alignment*



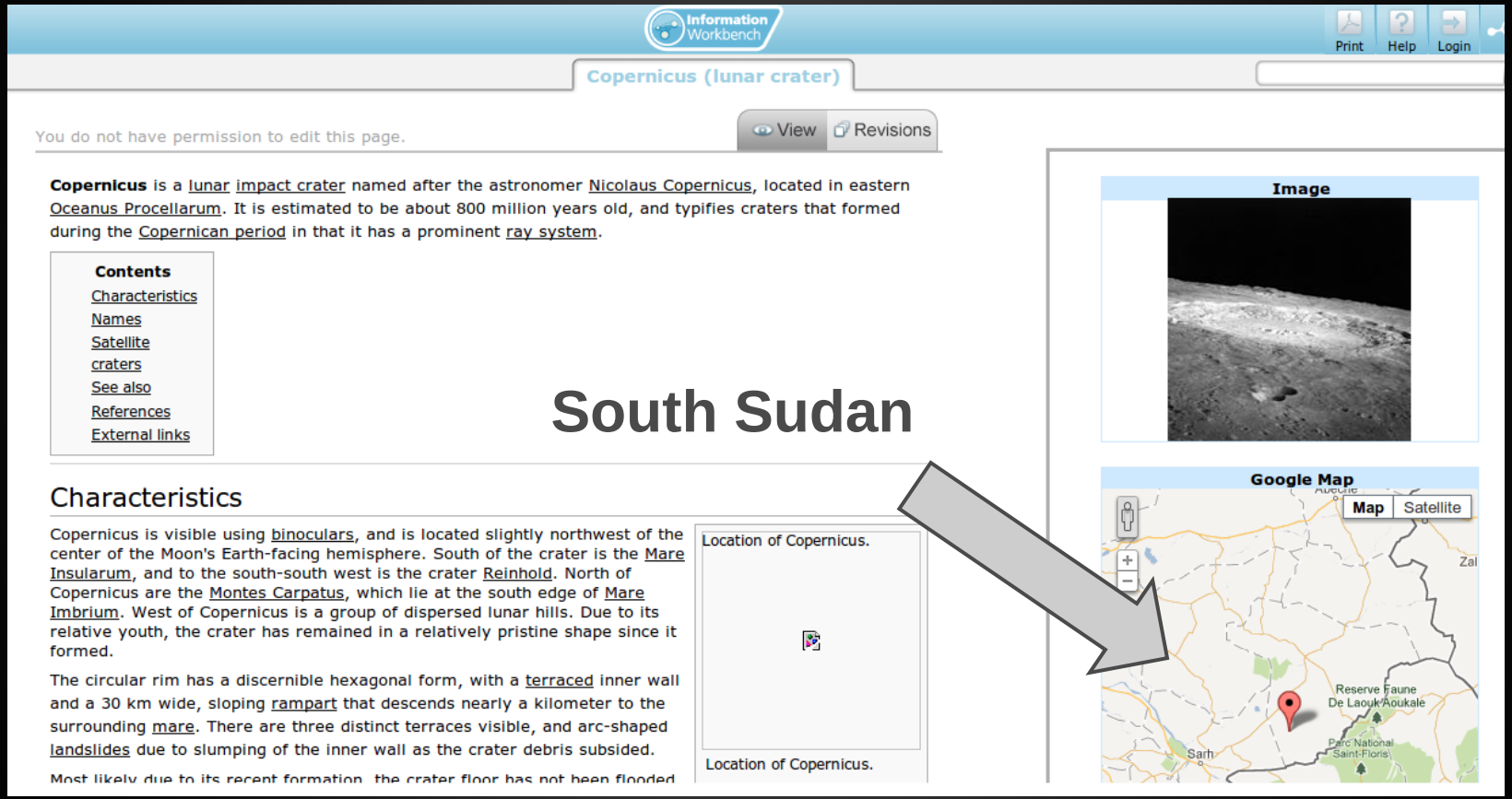

Credit: P. Hitzler (2014)

# Linked Data is Not Enough: *Problematic Integration*




Credit: K. Janowicz (2012)

# Linked Data is Not Enough: *Problematic Integration*




Problem: Linked Data missing coordinate reference system

Credit: K. Janowicz (2012)

# Variance

## An Ontology of Potholes

**Potholes** are defined as **cracks**

- of more than **30mm depth** *[North East Somerset, UK]*
- with a width of a **'large dinner plate'** (300mm) and the depth of a **'golf ball'** (40mm) *[Gloucestershire, UK]*
- with a width of a **'dinner plate'** (200mm) and a minimum depth of a **'fist'** (40mm) *[Worcestershire, UK]*
- if their depth is **'a pound coin and a 1p coin side by side'** *[Coventry]*

$\Longrightarrow$

Due to a severe **winter** (T) millions of potholes need to be repaired by the **local councils** (S) that are **legally responsible** for the **roads maintenance** (A) within their administrative **boundaries** (S).

Credit: K. Janowicz (2011)

# Variance


Credit: K. Janowicz (2011)

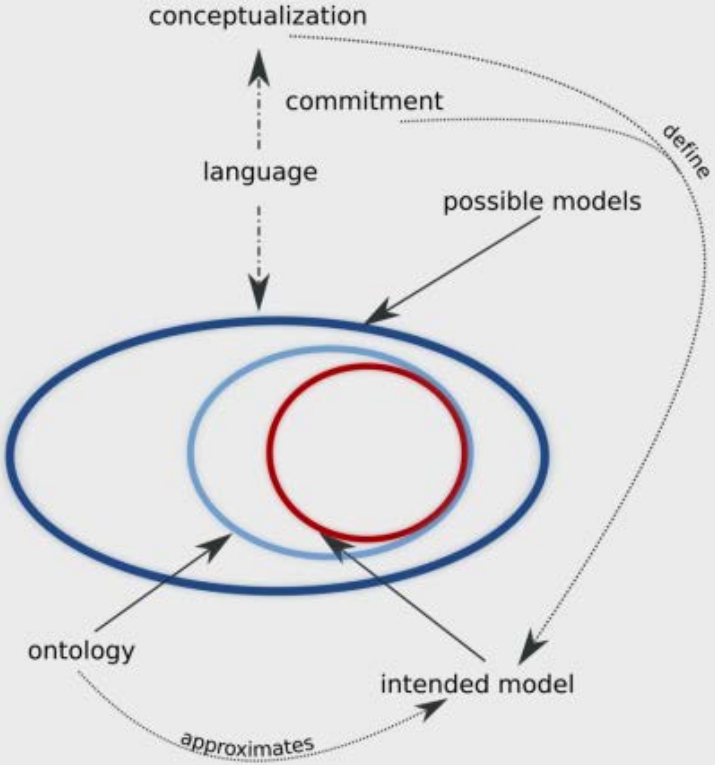
# Reasoning w. Data: Creates Pothole instances of...

major roadway collapse

new bridge construction

Credit: K. Janowicz (2011)

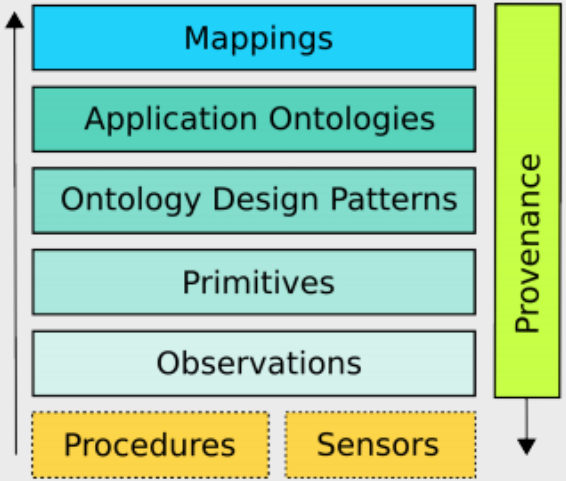
# ONTOLOGIES DO NOT FIX MEANING...




**... they restrict interpretations.**

Credit: K. Janowicz (2014)

# Observation-Driven Ontology Engineering




- Local and crisp **microtheories** instead of global ontologies
- Mine ontological primitives out of real **observation data**
- Assist domain experts in becoming knowledge engineers by developing reusable **patterns**
- **Defer** the introduction of classes that are heavy on ontological **commitments** (e.g., forest)
- Be driven by publishing, **discovery**, **reuse**, and integration needs.

- Ontologies should be about **communication** and not about replacing **numerical models**; we should **not** try to develop an **universal ontology** for rivers, mountains, forests, and so forth

Credit: K. Janowicz (2014)

# Example Use Case

**Ocean ecosystems: “What’s happening in the subpolar North Atlantic (surface to 1000 m), where are the zooplankton, and when do they come out of diapause?”**

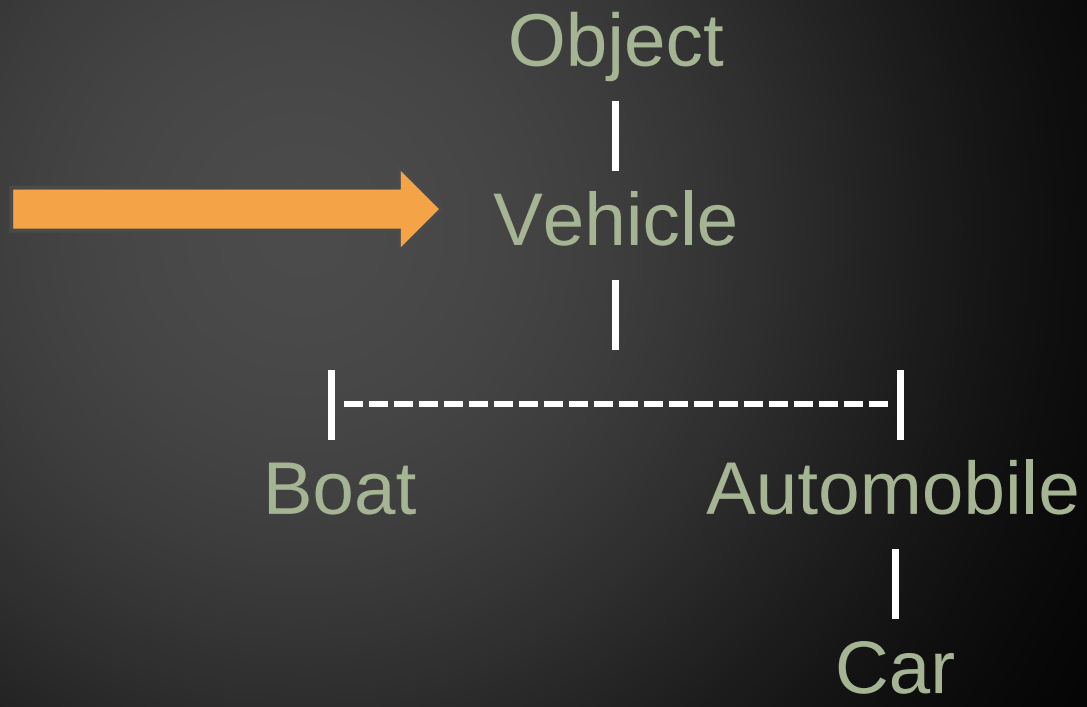
Requires primary production, species composition, abundance and size distribution, nutrients, and basic hydrographic data.

Credit: R. Arko (2015)

# Arriving at a Pattern Concept

Repo A: CAR

Repo B: BOAT




*by* Lowest Common Subsumer

# GeoLink concepts

- field expeditions
- sensor + sample datasets
- laboratory analyses
- journal publications
- conference presentations
- theses/reports
- funding awards

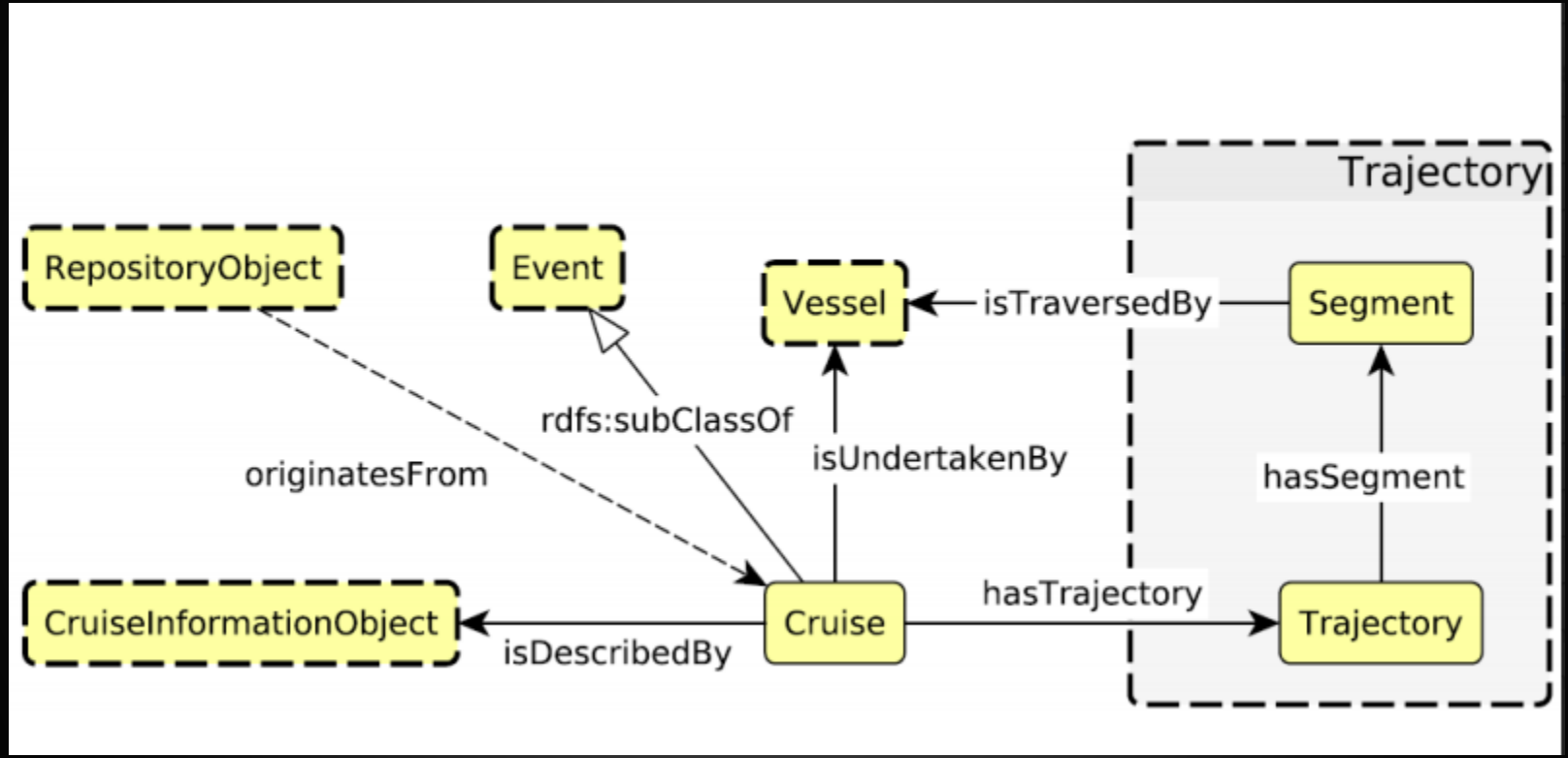
*To see these concepts as ontology design patterns, see*:

http://schema.geolink.org

**Diagrams**: https://github.com/ec-geolink/design/tree/master/patterns/diagrams

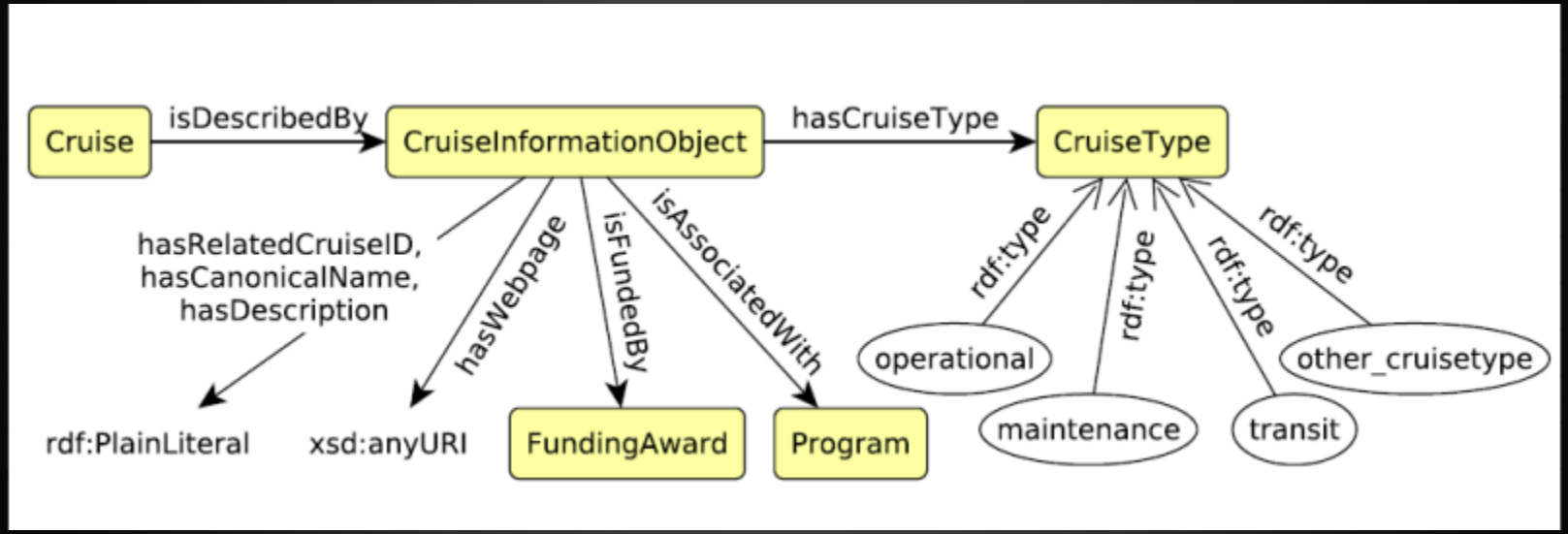
Credit: R. Arko (2015)

# GeoLink: Cruise Pattern



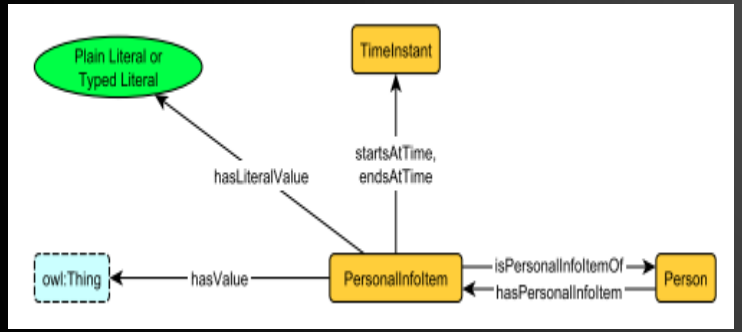

Credit: P. Hitzler

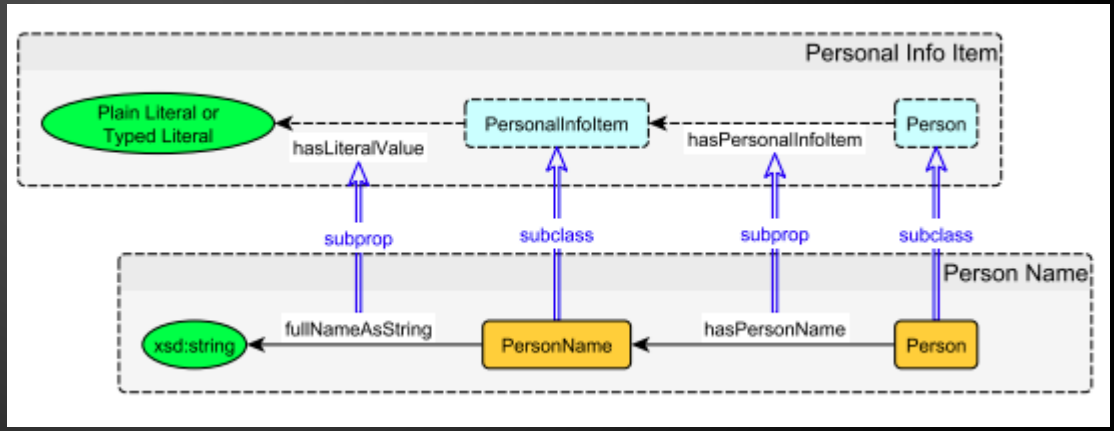
# Matured Ontology Engineering:
## The InformationObject pattern




Credit: P. Hitzler

# GeoLink: The Person Integration Problem


Credit: P. Hitzler

# How do we implement?

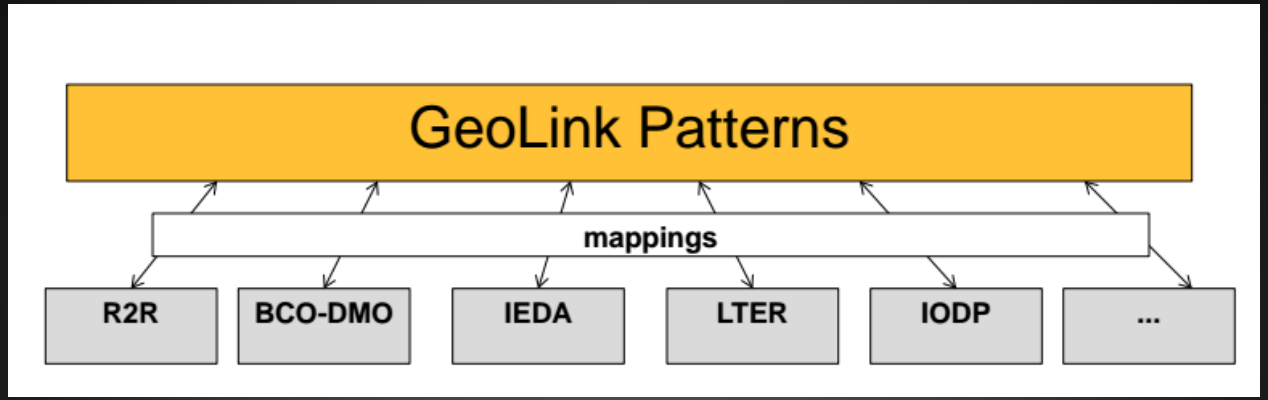
*Our local community still wants the lion - a more specific representation.*

Issues:

- Complexity of Patterns
  - comprehension by new partners
- Local adoption of pattern schema
  - impact on existing software, performance

*Broader discovery requires something greater.*

# Ontology Design Pattern: Views



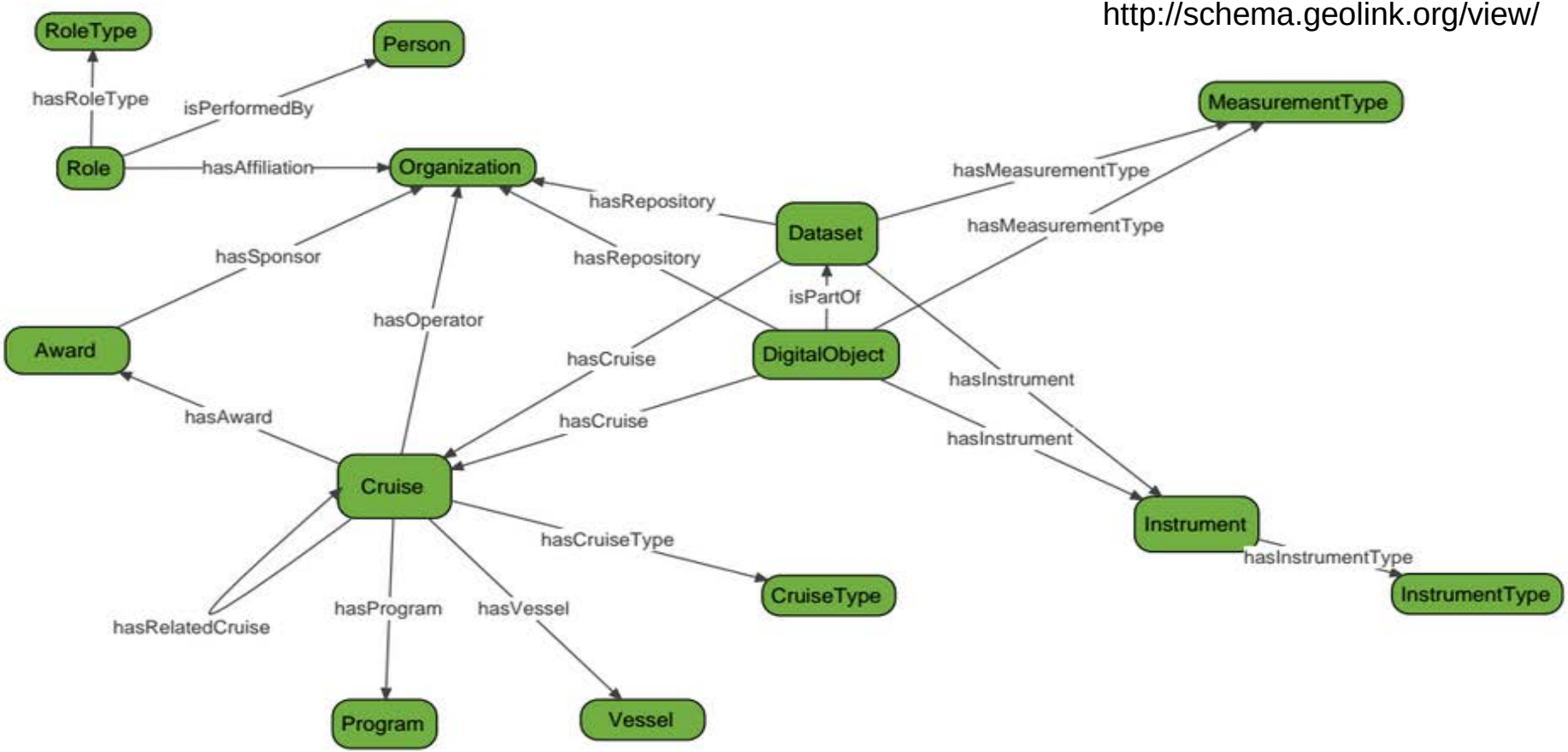

Data partners transform existing data to the patterns through a “View” construct.

*Example constructs:*

- SPARQL ‘CONSTRUCT’ statements
- RDF data processing script

# GeoLink: Aligned “View” Ontology

http://schema.geolink.org/view/



# BCO-DMO: Transforming our Local Data

*Local data can mean our linked data or data in a database, etc.*

***What matters is***

*1)* preserving the *existing* local linked data
*2)* publishing GeoLink pattern *compatible* data

http://geolink.bco-dmo.org/id/dataset/2415

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:glview="http://schema.geolink.org/
  <rdf:Description rdf:about="http://geolink.bco-dmo.org/id/dataset/2415">
    <rdf:type rdf:resource="http://schema.geolink.org/dev/view#Dataset"/>
    <glview:hasRepository rdf:resource="http://geolink.bco-dmo.org/id/organization/191"/>
    <bcoview:datasetLandingPage rdf:datatype="xsd:anyURI">http://www.bco-dmo.org/dataset/2415/data</bcov
    <glview:title xml:lang="en-US">alongtracksj</glview:title>
    <glview:description rdf:datatype="http://www.w3.org/1999/02/22-rdf-syntax-ns#HTML">Seward Johnson al
     Cruises:  9505 and 9507&lt;br />Shipboard meteorology and sea surface measurements&lt;br />along th
    <glview:hasCruise rdf:resource="http://geolink.bco-dmo.org/id/deployment/57484"/>
    <glview:hasCruise rdf:resource="http://geolink.bco-dmo.org/id/deployment/57486"/>
    <glview:hasParticipant rdf:resource="http://geolink.bco-dmo.org/id/role/456971"/>
    <glview:hasParticipant rdf:resource="http://geolink.bco-dmo.org/id/role/456970"/>
    <glview:hasInstrument rdf:resource="http://geolink.bco-dmo.org/id/dataset-instrument/4232"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/1062"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/969"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/789"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/1066"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/730"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/731"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/997"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/1001"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/955"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/836"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/1060"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/1061"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/924"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/940"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/1044"/>
    <glview:hasMeasurementType rdf:resource="http://geolink.bco-dmo.org/id/measurement-type/1052"/>
    <bcoview:isFundedBy rdf:resource="http://geolink.bco-dmo.org/id/award/54610"/>
    <bcoview:isFundedBy rdf:resource="http://geolink.bco-dmo.org/id/award/54626"/>
  </rdf:Description>
```

# Ontology Design Patterns (ODPs)

1. **Modular**

   Use the modules you have data for

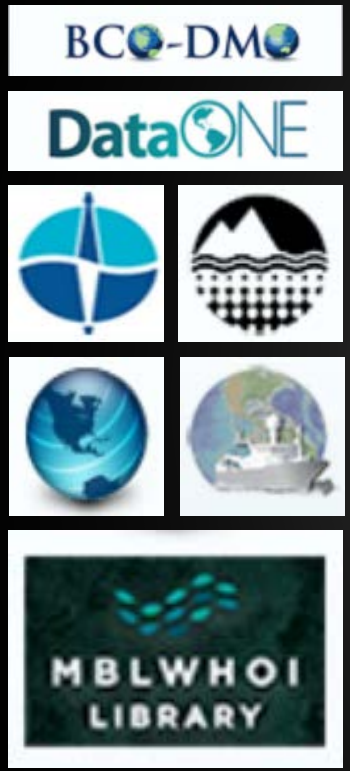
2. **Extendable**

   When additions or replacements are needed, existing data isn't affected

3. **Not disruptive**

   Existing Linked Data can remain unchanged

# GeoLink: Repositories




Biological and Chemical Oceanography Data Management Office (BCO-DMO)
Data Observation Network for Earth (DataONE)
Interdisciplinary Earth Data Alliance (IEDA)
International Ocean Discovery Program (IODP)
Long Term Ecological Research (LTER) Network
Marine Biological Laboratory/Woods Hole Oceanographic Institution (MBLWHOI) Library
Rolling Deck to Repository (R2R) Program

*Additional Collaborators*

Advanced Cooperative Arctic Data and Information Service (ACADIS)
CLIVAR and Carbon Hydrographic Data Office (CCHDO)
Federation of Earth Science Information Partners (ESIP)
Index to Marine and Lacustrine Geological Samples (IMLGS)
U.S. Geological Survey - National Geochemical Database (NGDB)

# GeoLink Schema

http://schema.geolink.org/

# GeoLink on github.com

https://github.com/ec-geolink/design

# References


Arko, R. et al (2015). GeoLink: Semantics and Linked Data for the Geosciences. Presented at EarthCube Tech Meeting 2015.

Hitzler, P. et al (2015). Ontology Design Patterns in OceanLink and GeoLink. Presented at GeoVoCamp, Santa Barbara, CA.

Hitzler, P. et al (2014). Ontology Design Patterns for Large-Scale Data Interchange and Discovery. Presented at EKAW 2014.

Janowicz, K. et al (2014). GeoLink — Ontology (Design Pattern) Engineering. Presented at GeoLink Partners Meeting, Santa Barbara, CA

Janowicz, K. (2012). Place and Location on the Web of Linked Data. http://stko.geog.ucsb.edu/location_linked_data

Janowicz, K. et al (2011). Geospatial Semantics - Problems and Projects. Presented at GeoVISTA Center, Pennsylvania State University.