

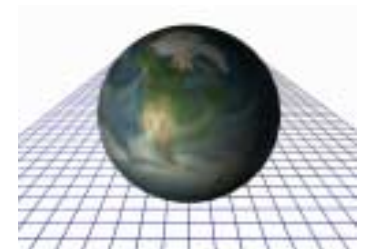
Ontologies and The Earth System Grid

Line Pouchard
(ORNL)

PI's: Ian Foster (ANL); Don Middleton (NCAR); and Dean Williams (LLNL)

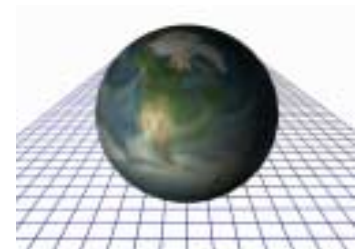
<http://www.earthsystemgrid.org>

The NIEeS Workshop
Cambridge, UK



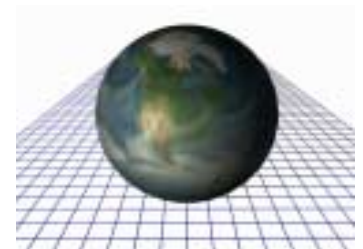
Overview: three threads

- Basic ideas about classification systems and ontologies
- Some practical or not so practical examples
- The approach in ESG

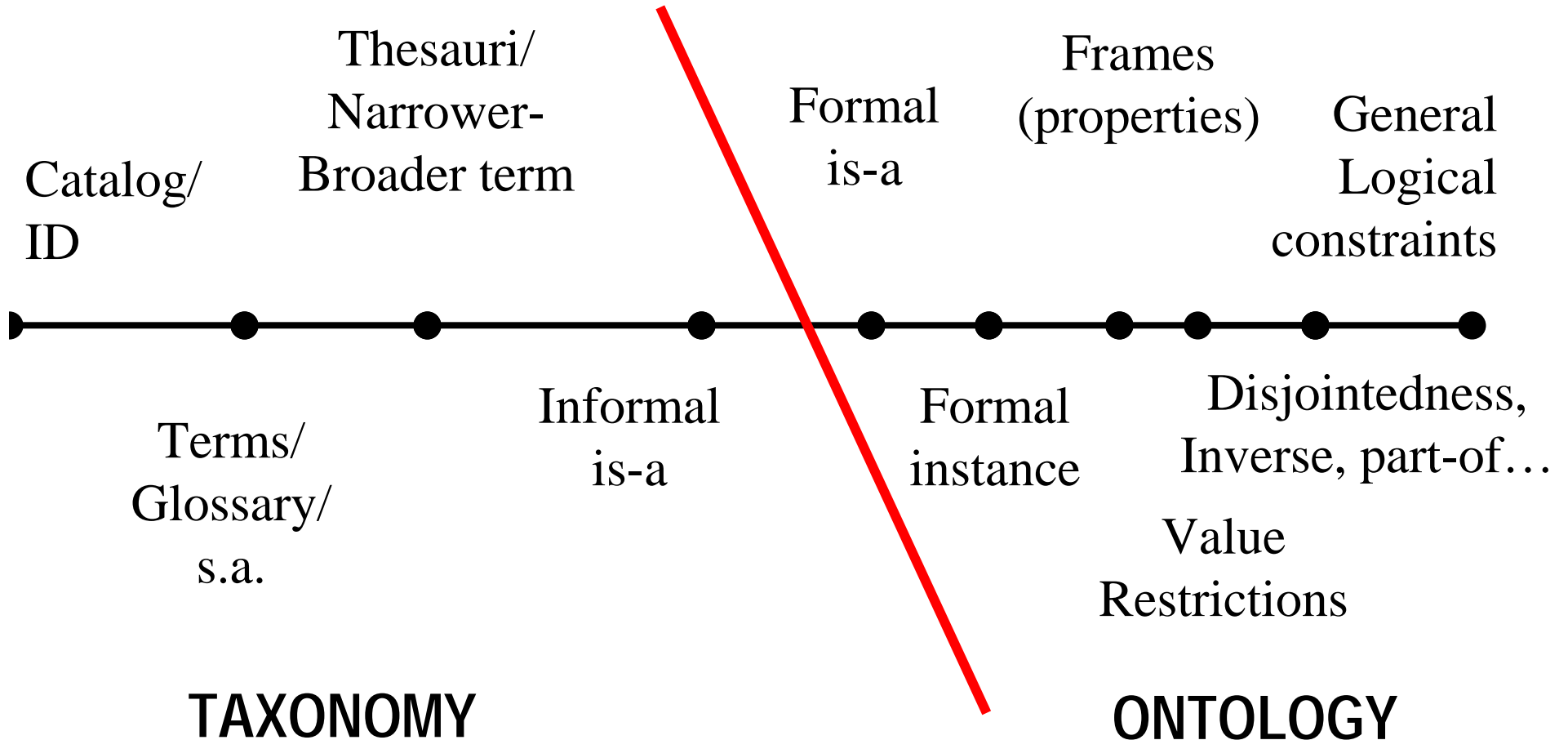


What are ontologies?

- Domain ontologies: the basic terms and relationships comprising the vocabulary of a domain area.
- Task ontologies: distinguish the operations done on classes from the class definitions.
- A set of definitions for these terms and axioms constraining the use of terms.
- A classification of these terms
 - Allow hierarchical classification systems, generalization and inheritance, aggregation, and a greater variety of structural relations than taxonomies and controlled vocabularies.
- The rules for combining terms and relations.



A continuum of classifications

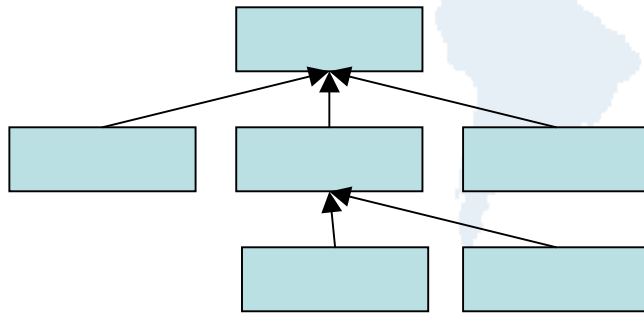


Based on McGuinness, 2001:

http://www.daml.org/2001/06/swday-ontologies/Ontologies-talk-060401_files/frame.htm

Differences in classification systems

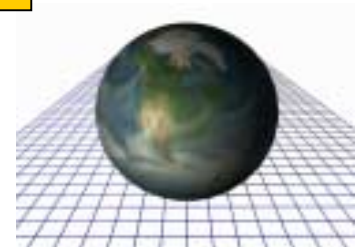
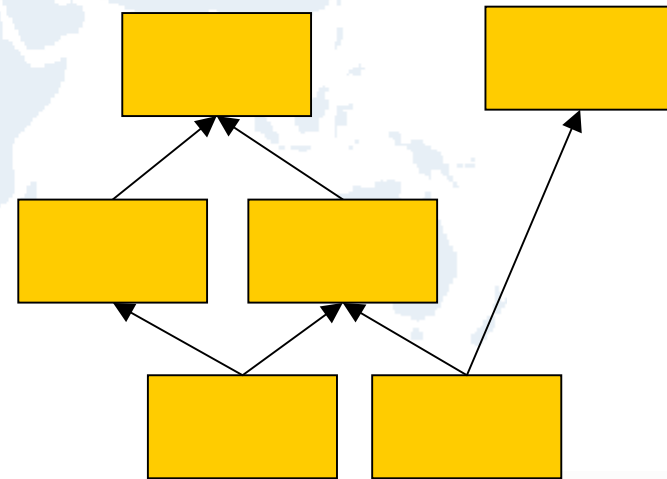
- Controlled vocabularies
 - Standard Names
- Taxonomies
- Data dictionaries
- Object-oriented systems



- Hierarchical structures

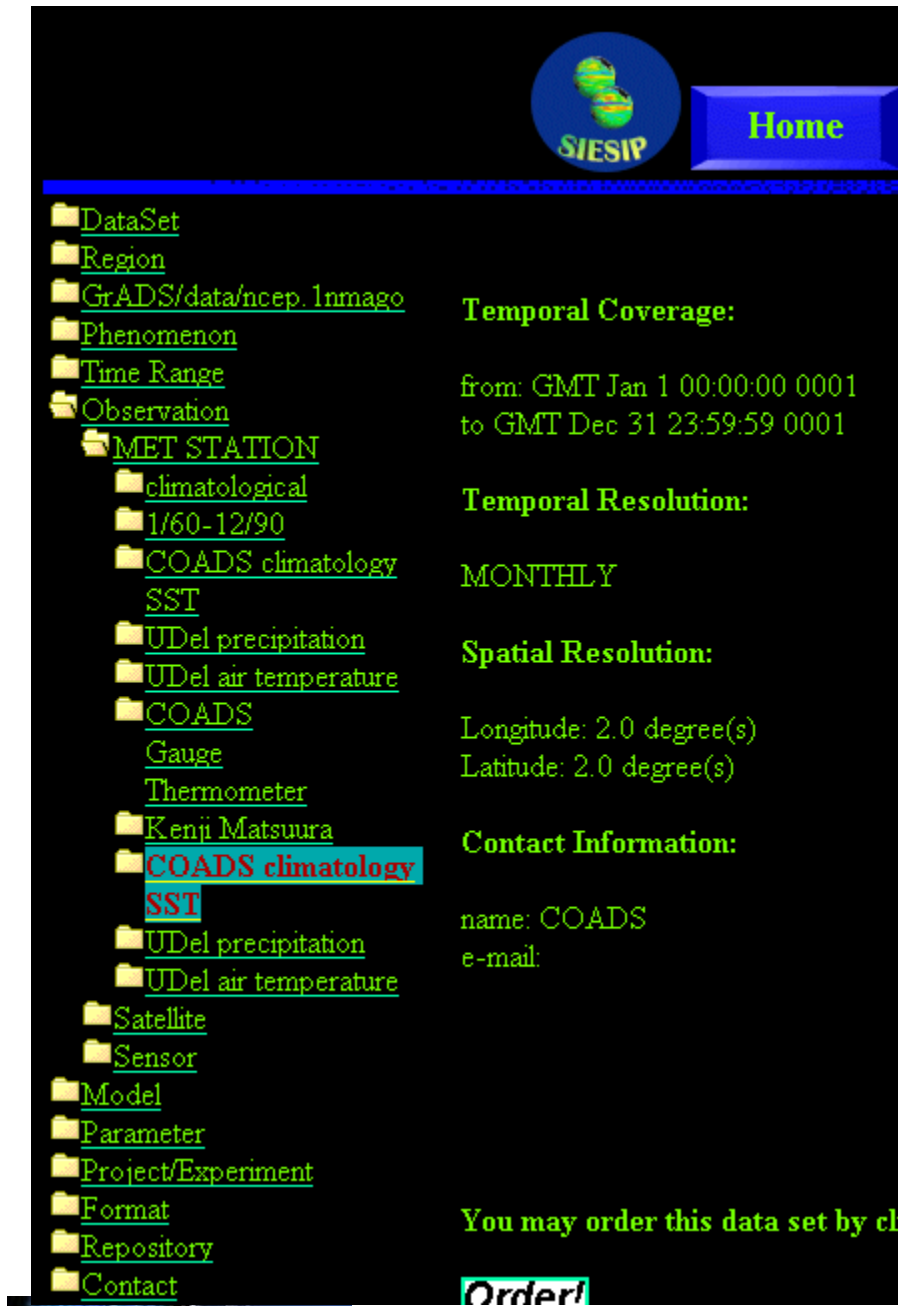


- Ontologies



Drawbacks of a Tree Hierarchy


<http://www.siesip.gmu.edu/>



The screenshot shows the SIESIP website interface. At the top, there is a logo with a globe and the text 'SIESIP' and a blue button labeled 'Home'. Below this is a navigation menu with a tree structure. The selected path is: DataSet > Region > GrADS/data/ncep.1nmago > Phenomenon > Time Range > Observation > MET STATION > climatological > 1/60-12/90 > COADS climatology > SST. To the right of the menu, the following information is displayed:

- Temporal Coverage:** from: GMT Jan 1 00:00:00 0001 to GMT Dec 31 23:59:59 0001
- Temporal Resolution:** MONTHLY
- Spatial Resolution:** Longitude: 2.0 degree(s), Latitude: 2.0 degree(s)
- Contact Information:** name: COADS, e-mail:

At the bottom of the page, there is a text box that says "You may order this data set by cli" and a button labeled "Order!".



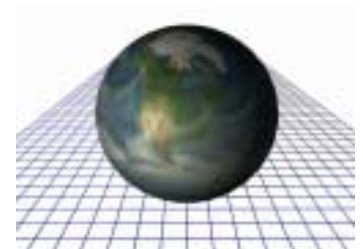
The screenshot shows the SIESIP website interface with a different tree hierarchy. The selected path is: DataSet > Region > GrADS/data/ncep.1nmago > Phenomenon > Time Range > Observation > Model > Parameter > Project/Experiment > Format > GRIB > DODS > climatological > COADS climatology > SST. To the right of the menu, the following information is displayed:

- Temporal Coverage:** from: GMT Jan 1 00:00:00 0001 to GMT Dec 31 23:59:59 0001
- Temporal Resolution:** MONTHLY
- Spatial Resolution:** Longitude: 2.0 degree(s), Latitude: 2.0 degree(s)
- Contact Information:** name: COADS, e-mail:

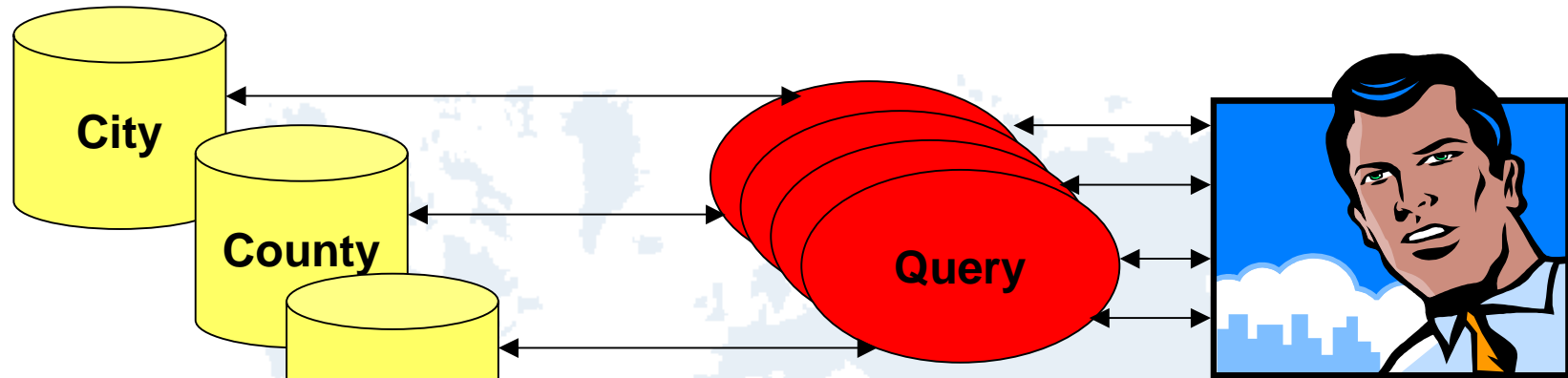
At the bottom of the page, there is a text box that says "You may order this data set by cli" and a button labeled "Order!".

How can we use ontologies?

- Database schemas, knowledge base schemas
- Declarative knowledge about XML elements and attributes
- Visualize, browse, query, edit metadata
- Formulate complex queries
- Make domain assumptions explicit
- Merge several ontologies to form a domain
- Can reference and extend one another
- A single 'master' ontology is not practical (same problems as a standard)
- A bottom-up approach



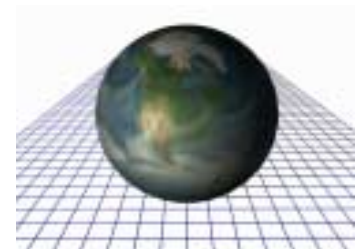
Example Scenario: Disaster Management



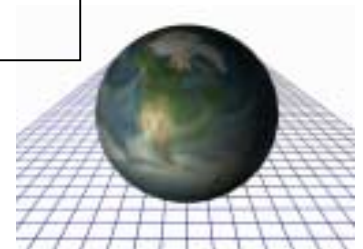
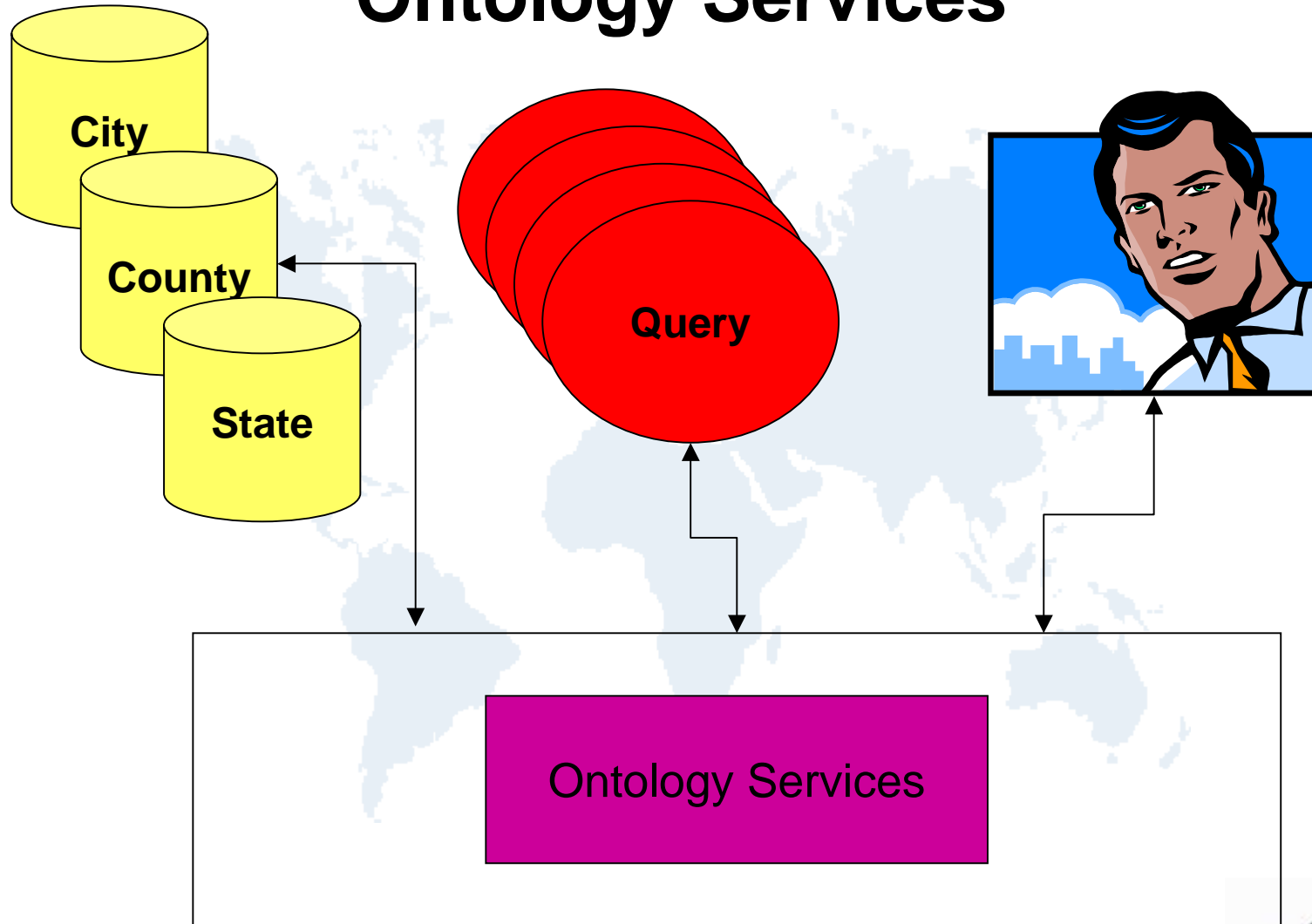
- Personnel
 - Rescue units
 - Firefighters
 - Rescue teams
 - Cost guards
 - Volunteers
- DB schemas
- Rescue vehicle
 - Fire engine
 - Fire truck
 - Truck
 - Helicopter
 - Cost guard vessel
 - Police Cruiser
 - Red cross vehicles
 - Personal cars

QUESTIONS

- How many personnel are available for rescue?
- How much equipment?
- Who has what?
- Where is it?



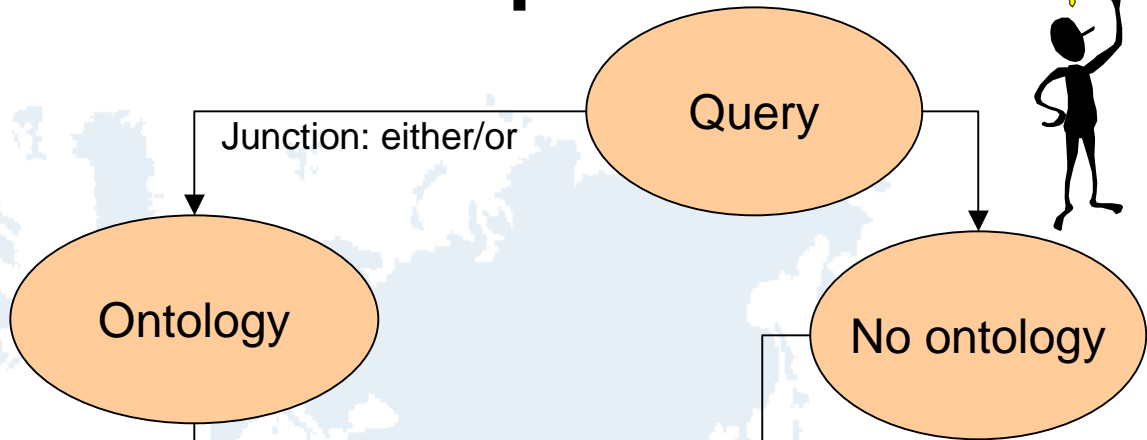
Ontology Services



Example scenario: atmospheric sciences

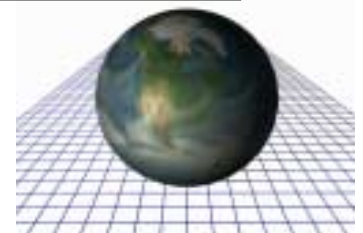
User wants to send a query

Temperature is noted as T



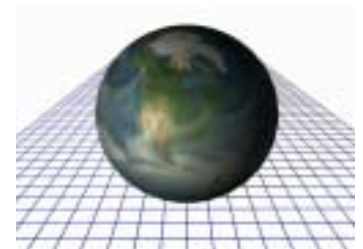
- No pre-required knowledge of the format
- No pre-required knowledge of the variable names
- A query on the variable definition
- Acquire knowledge about
 - sub- domains
 - relationships between variables:

- Knowledge of the format: NcML
- Knowledge of the variable names in possible models
- A query on the exact variable name: no temperature, no temp, no Temperature_name, etc... if T
- At least one query per format
- No knowledge of the relationship between two variables



Communication based on content

- Provide a common frame of reference and some common vocabularies in a domain and sub-domains
- No imposed data model and structure from top-down
- Relationships between terms:
 - A *CoordinateVariable* “*is_a*” *Variable*
 - A *GeneralCoordinateVariable* “*is_bound_by*” A *BoundaryVariable*
 - A *Dimension* “*is_part_of*” a *CoordinateSystem*
- Organize large amounts of metadata on a topic
- System inter-operability based on semantics
- Agent-based systems

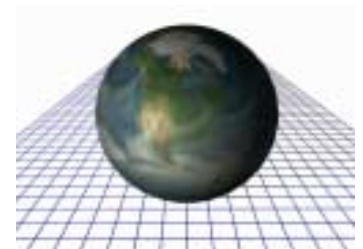


Challenges, Issues and Tools

- Languages
 - KIF: expressive, impractical
 - DAML+OIL, OWL: gains acceptance
 - RDFS, RDF: enriched XML
- Graphical representations
 - UML
 - Directed Acyclic Graphs
- Interfaces: APIs and GUIs
- Import/Export Capabilities
- Editors
 - single and group
- Browsers
- Query tools
- Inference Engines
- Repositories
- Tools for ontology re-use
- Tools for aggregation

Team-based iterative development

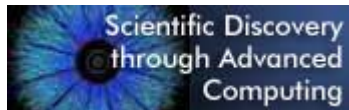
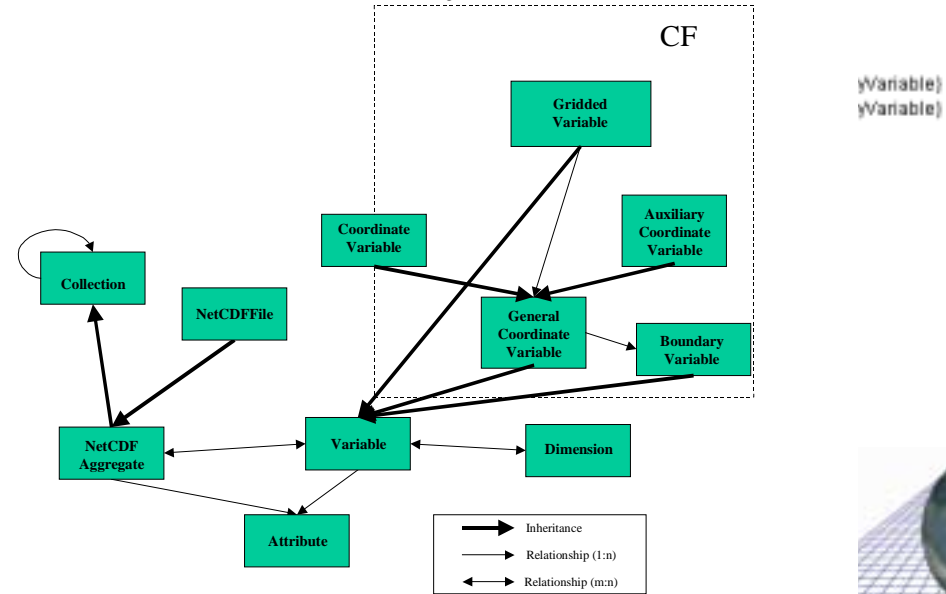
***Domain experts, Users, Ontology experts,
Developers***



An Ontology Tool: Protégé 2000

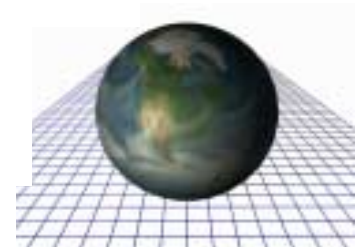
(Protégé-2000: Copyright © 2002 Stanford Medical Informatics)

CF / NetCDF Object Model



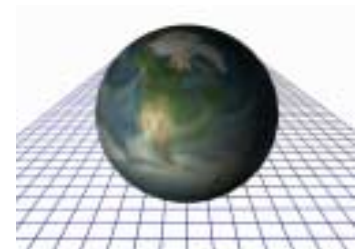
September 10, 2002

Earth System Grid - 1000000



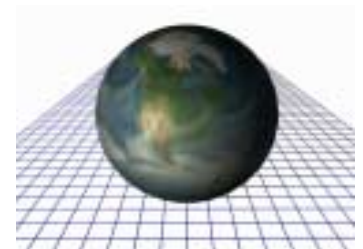
Possible uses of Protégé for ESG

- Extensible with plug-ins
 - A database backend plug-in
 - A Protégé server plug-in
- As front-end to a metadata repository
- or itself a metadata repository
 - Querying metadata, returns metadata instances, returns relationships, expresses overlaps
- Weaknesses: query construction, decoupling DB and Protégé
- Grid performance and scalability levels are untested
- Integration with data-intensive, distributed systems like ESG



Ontology Languages and Representation

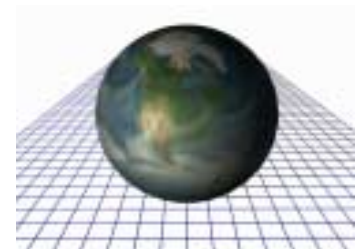
- KIF example:
 - Process Specification Language (PSL)
www.mel.nist.gov/psl
- IDL
- Directed Acyclic Graph example:
 - Gene Ontology
www.geneontology.org
- RDF, RDF-S, DAML+OIL, OWL
- others



PSL xor activity KIF representation

An activity ?a is an xor activity if whenever ?a occurs, there exists a unique subactivity which occurs. Nondeterminism arises from the uncertainty over which activity occurs.

```
(defrelation xor (?a) :=
(forall (?occ)
  (
    (occurrence ?occ ?a)
    (exists (?a1 ?occl)
      (and (subactivity ?a1 ?a)
            (occurrence ?occl ?a1)
            (subactivity_occurrence ?occl ?occ))))))
(forall (?a2 ?occ2)
  (= (and (subactivity ?a2 ?a)
          (occurrence ?occ2 ?a2)
          (not (subactivity_occurrence ?occ2 ?occl))))))
```

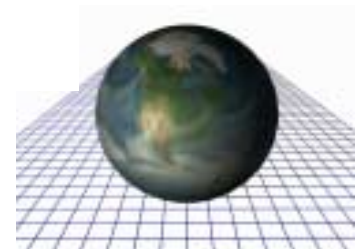


A Simple Manufacturing Ontology IDL Representation

```
From: Ontology-librarian@KSL.Stanford.EDU
Sent: Thursday, October 21, 1999 12:39 PM
To: pouchardlc@ornl.gov
Subject: Translation of SIMPLE-MANUFACTURING into IDL

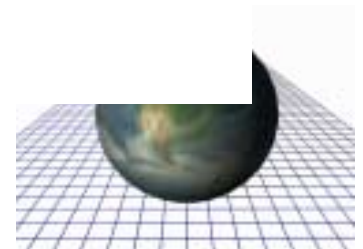
...
// Translation of Simple-Manufacturing
// on Thursday, 21 October 1999
// into target language Idl
|...
|/// Translating Utensil as if it were an IDL interface
|///
interface Utensil : Object
/* used in making cakes. includes appliances. re-usable. */

    // Additional information about Utensil:
    // Instance-Of: Class, Primitive
    // Disjoint-Decomposition: (Setof Oven Eggbreaker Cakeform
Knife Timer Sieve Potstove Bowlspoon Tablespoon)
{
};
```

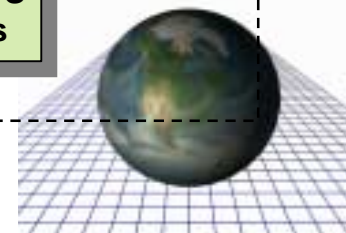
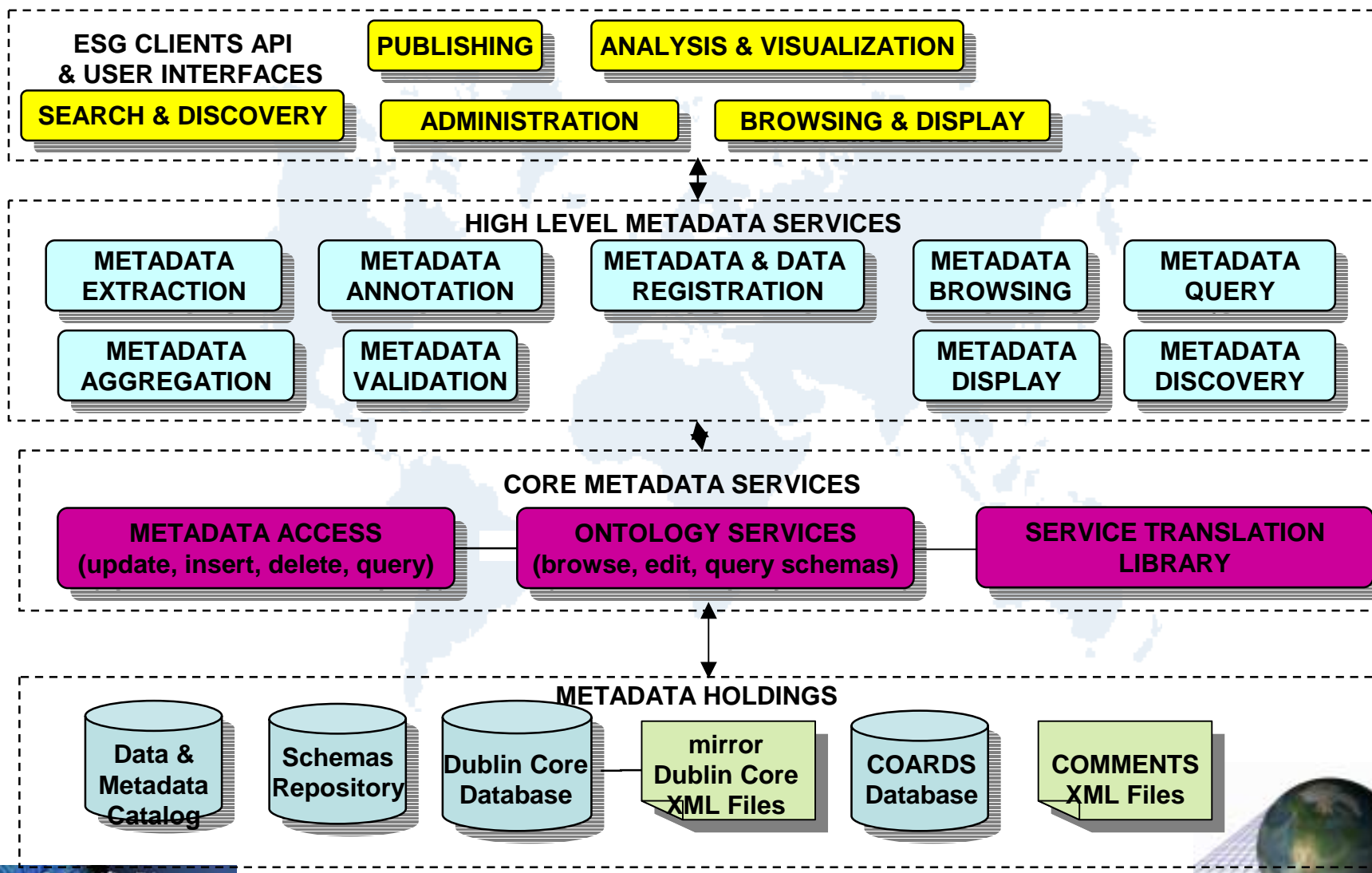


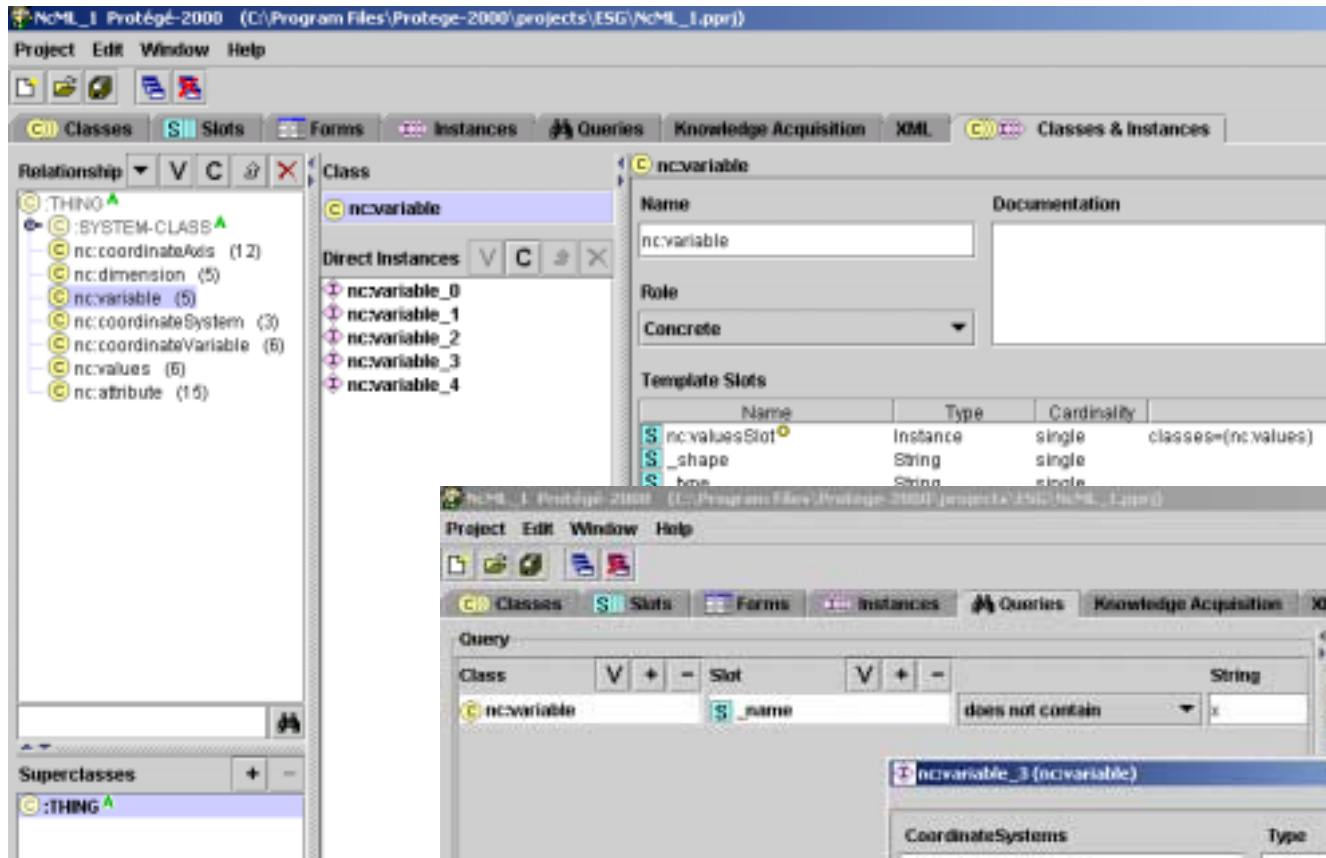
GO Process Ontology DAG representation

```
!autogenerated-by:    DAG-Edit version 1.206
!saved-by:            mall
!date:                Thu Oct 11 18:49:16 GMT+01:00 2001
!version: $Revision: 2.107 $
!note:                file automatically generated by GO-Editor
$Gene_ontology ; GO:0003673
<biological_process ; GO:0008150
  %behavior ; GO:0007610
  ....
  %reproductive behavior ; GO:0019098
    %mating behavior ; GO:0007617
      %mating ; GO:0007618
        %copulation ; GO:0007620
          %courtship behavior ; GO:0007619
            %female courtship behavior ; GO:0008050
              %male courtship behavior ; GO:0008049
                %male courtship behavior (sensu Drosophila) ; GO:0016542
                  <male courtship behavior (sensu Drosophila)\, licking ; GO:0016546
                    <male courtship behavior (sensu Drosophila)\, orientation ; GO:0016543
                      <male courtship behavior (sensu Drosophila)\, tapping ; GO:0016544
                        <male courtship behavior (sensu Drosophila)\, wing vibration ; GO:0016545
              %repression of female receptivity ; GO:0007621
            %oviposition ; GO:0018991
          %rhythmic behavior ; GO:0007622
```



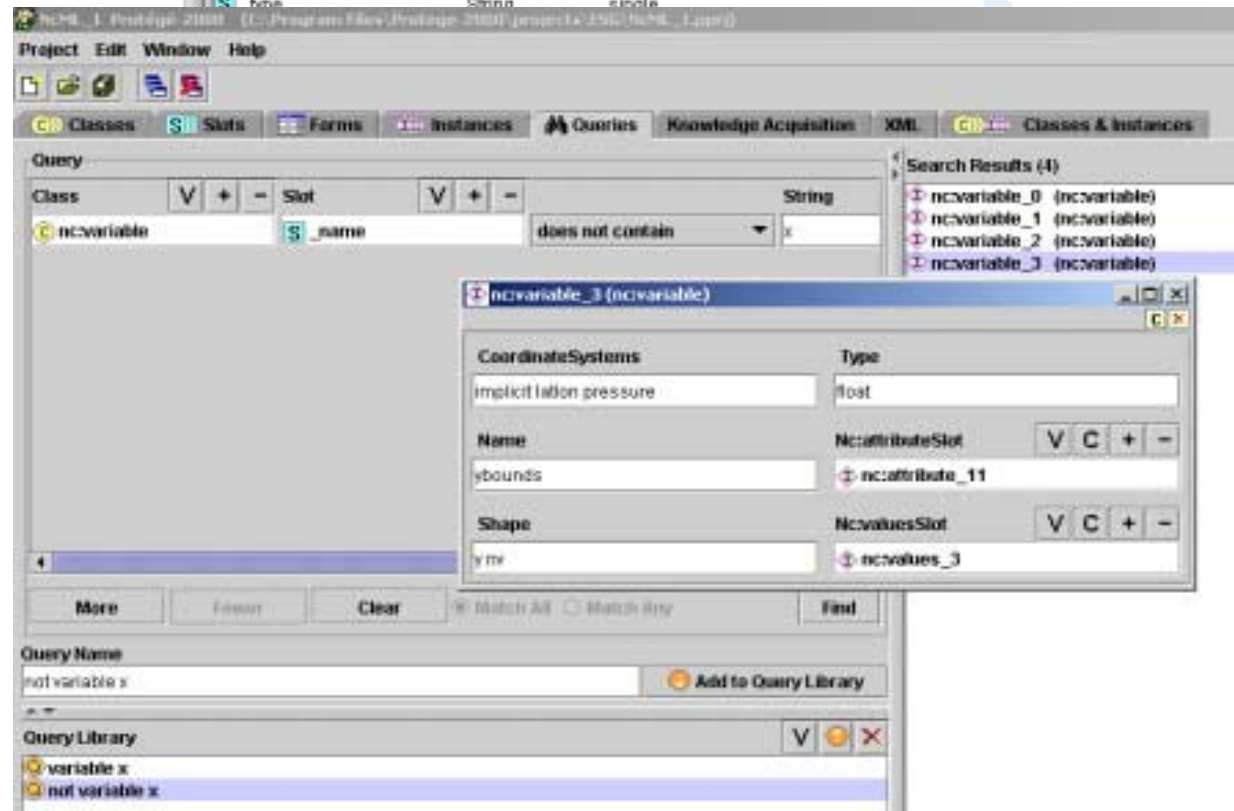
Ontologies in ESG Metadata Services





NcML in Protégé

no explicit relations between elements

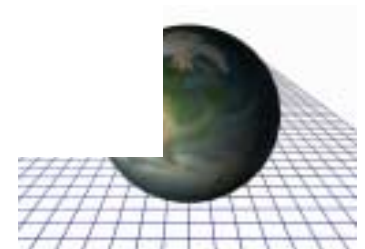


(Protégé-2000:
Copyright © 2002
Stanford Medical
Informatics)



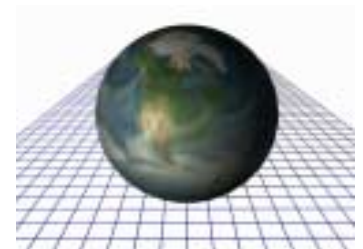
September 10, 2002

Earth System Grid – Pouchard



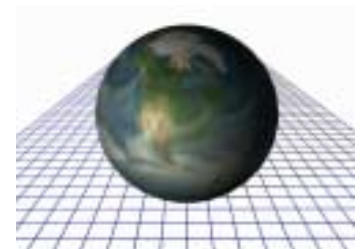
The Semantic Grid

- www.semanticgrid.org
- A vision: incorporating some semantics in Grid infrastructure.
- A set of technologies based on the Semantic Web
 - A collection of tools.
 - Representation languages.
 - Uses DAML+OIL/OWL
 - OWL is a W3C standard candidate.
- Based on services and components.



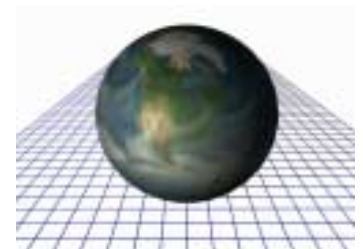
DAML + OIL language features

- **Equivalences**
 - `equivalentTo`, `sameClassAs`, `samePropertyAs`, `sameIndividualAs`
- **Booleans**
 - `unionOf`, `disjointUnionOf`, `intersectionOf`, `complementOf`, `Disjointwith`, `differentIndividualFrom`, `one of`
- **Restrictions**
 - `toClass`, `hasValue`, `hasClass`, `onProperty`
- **Cardinality**
 - `minCardinality`, `maxCardinality`, `hasClassQ`, `minCardinalityQ`, `maxCardinalityQ`, `cardinalityQ`
- **Class properties**
 - `inverseOf`, `UniqueProperty`, `TransitiveProperty`, `unambiguous Property`



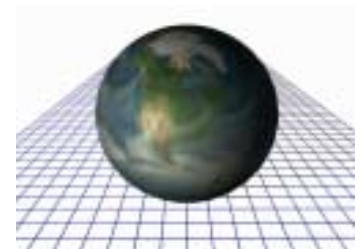
RDF and RDF-S specification

- Used to create a collection of assertions – specified as triples
 - ‘Cat’ ‘is an instance of’ ‘Pet’
 - ‘Josephine’ ‘is an instance of’ ‘Cat’
 - ‘Cat’ ‘has the property’ ‘purrs’
- RDF tools can make inferences from these assertions
 - Josephine purrs
 - Josephine has other properties associated with pets
- Class, Property, subClassOf, subPropertyOf, range and domain [type of a triple] rdfs:range, rdfs:domain



DAML+OIL, OWL

- The DAML ontology language is a collection of RDF Schema triples (i.e. a valid RDF ontology) using several agreed upon standard terms
- Consequently, it inherits all features of RDF and RDF-S
- But also adds new expressivity and ability to represent further common sense knowledge
- www.daml.org
- OWL is a W3C standard candidate
- <http://www.daml.org/2002/06/webont/owl-ref-proposed#1>



Summary

- Where is ESG?
 - Understand the need, solutions in other domains, and the tools available
 - Some standardization of data formats exist
- Path?
 - Add another component to the core metadata services
 - Some enriched form of XML metadata
 - Re-use partial metadata schemas such as pedigree from other Scidac projects
 - Collaborate with other projects such as the NIEeS project

