Semantic Web:
Core Concepts and Mechanisms

MMI ORR – Ontology Registry and Repository

Carlos A. Rueda
Monterey Bay Aquarium Research Institute
Moss Landing, CA

ESIP 2016 Summer meeting

What’s all this about?

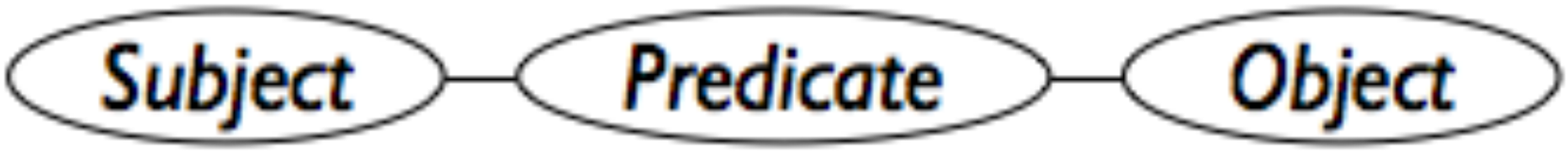
- It’s all about formally capturing knowledge about the world
  so computers can be more useful

  so we can tackle pressing problems more effectively and efficiently

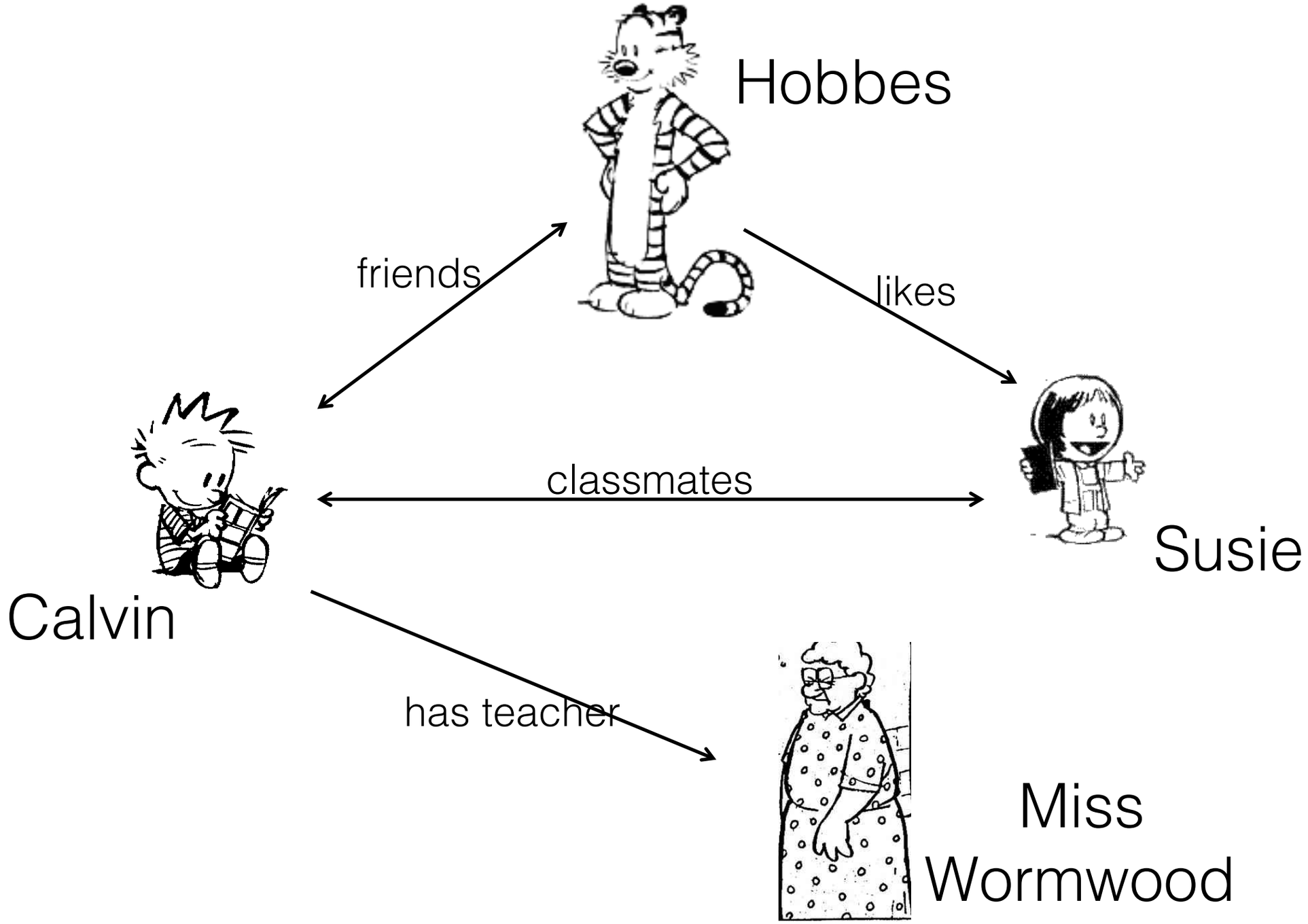
Capturing knowledge

- Knowledge expressed as *statements*

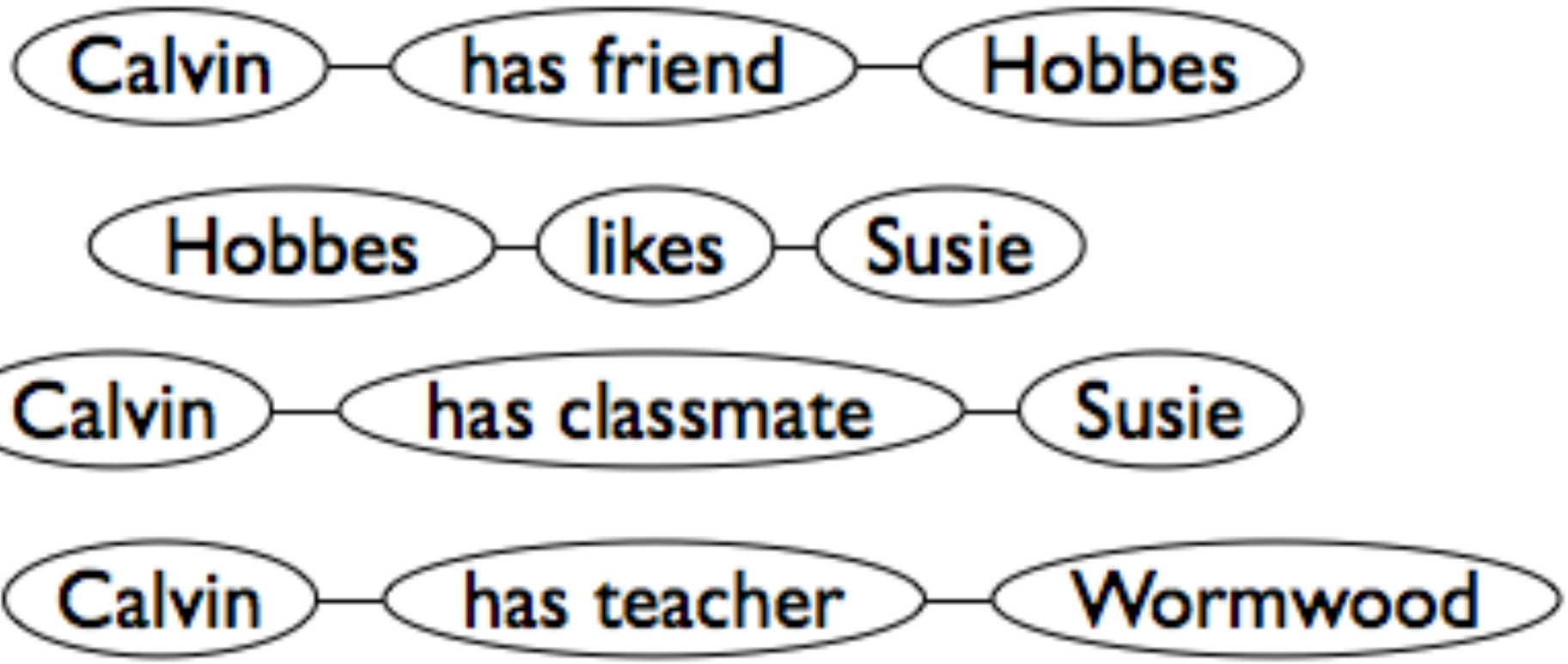
  Statements modeled as *triples* of the form:



Some knowledge

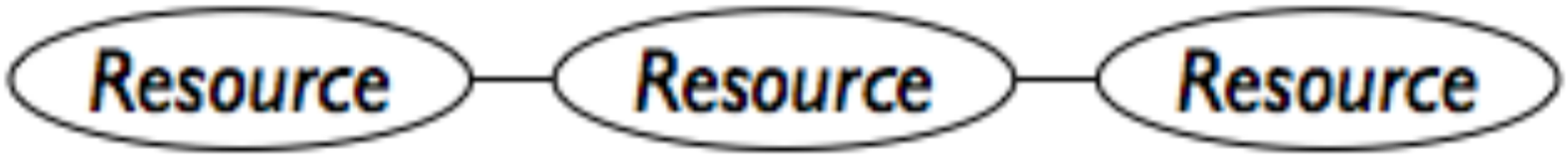


Capturing semantics with triples



RDF: Resource Description Framework

- W3C standard to express information about resources
- Anything can be a resource, including physical things, documents, abstract concepts, numbers and strings
- The triple components denote resources



RDF: Resource Description Framework

- Designed to support the Semantic Web
- In much the same way that HTML supports the Web
- RDF itself does not provide the machinery of inference
- AAA: “Anyone can say anything about anything”
- RDF-based applications must find ways to deal with conflicting sources of information

https://www.w3.org/TR/2002/WD-rdf-concepts-20020829/#xtocid48014

Resources

- Resources are denoted to by IRIs and literals
- IRI = Internationalized Resource Identifier
  - To identify resources, and to link to them
- Literals denote values according to known datatypes (numbers, strings, dates, ..)

IRIs or URIs?

- URIs used in RDF 1.0

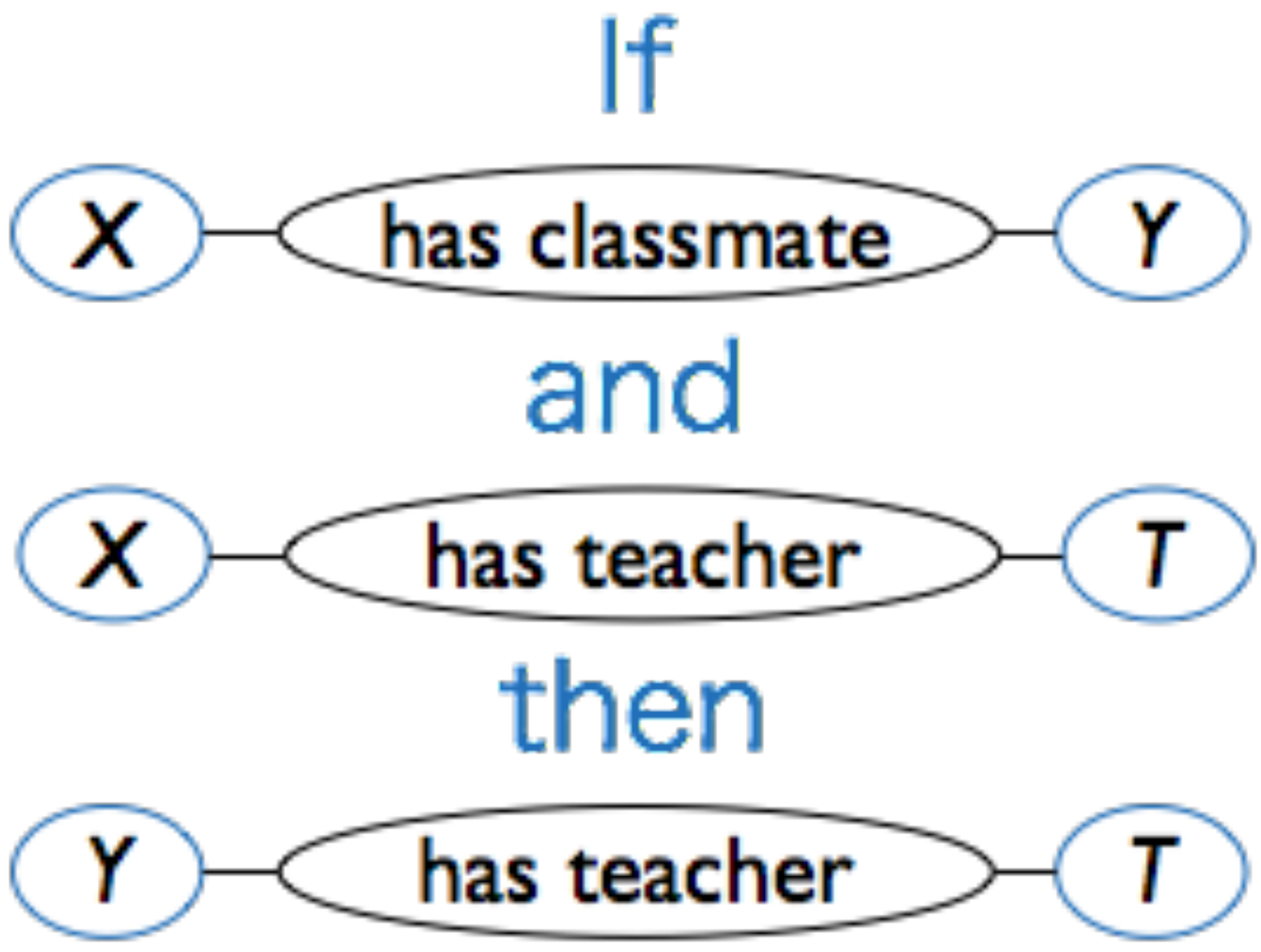
- IRIs now used in RDF 1.1
  IRI: Generalization of URI allowing non-ASCII characters to be used in the IRI character string

- Every URI is an IRI

- URIs still prevalent, with mapping needed from IRIs to URIs when retrieval over the HTTP protocol

Rules for inference

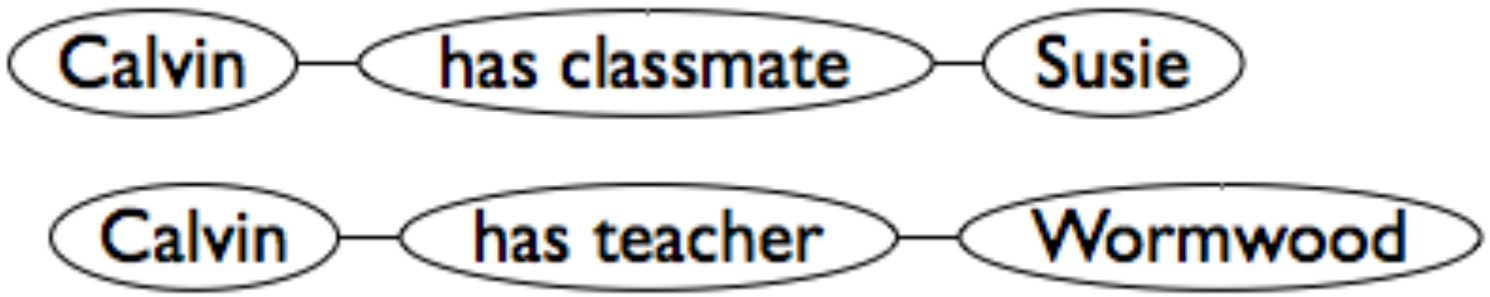
- Example



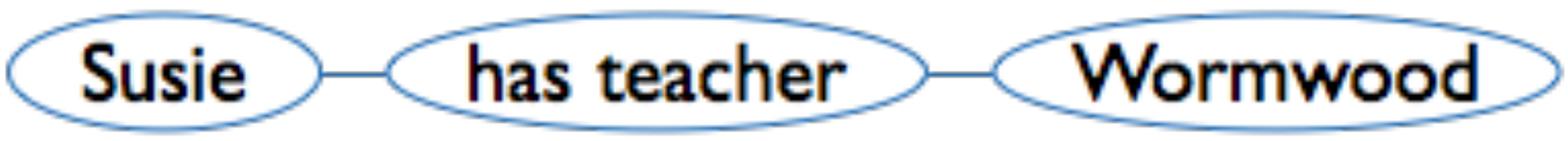
RIF: Rule Interchange Format (W3C)

Inference

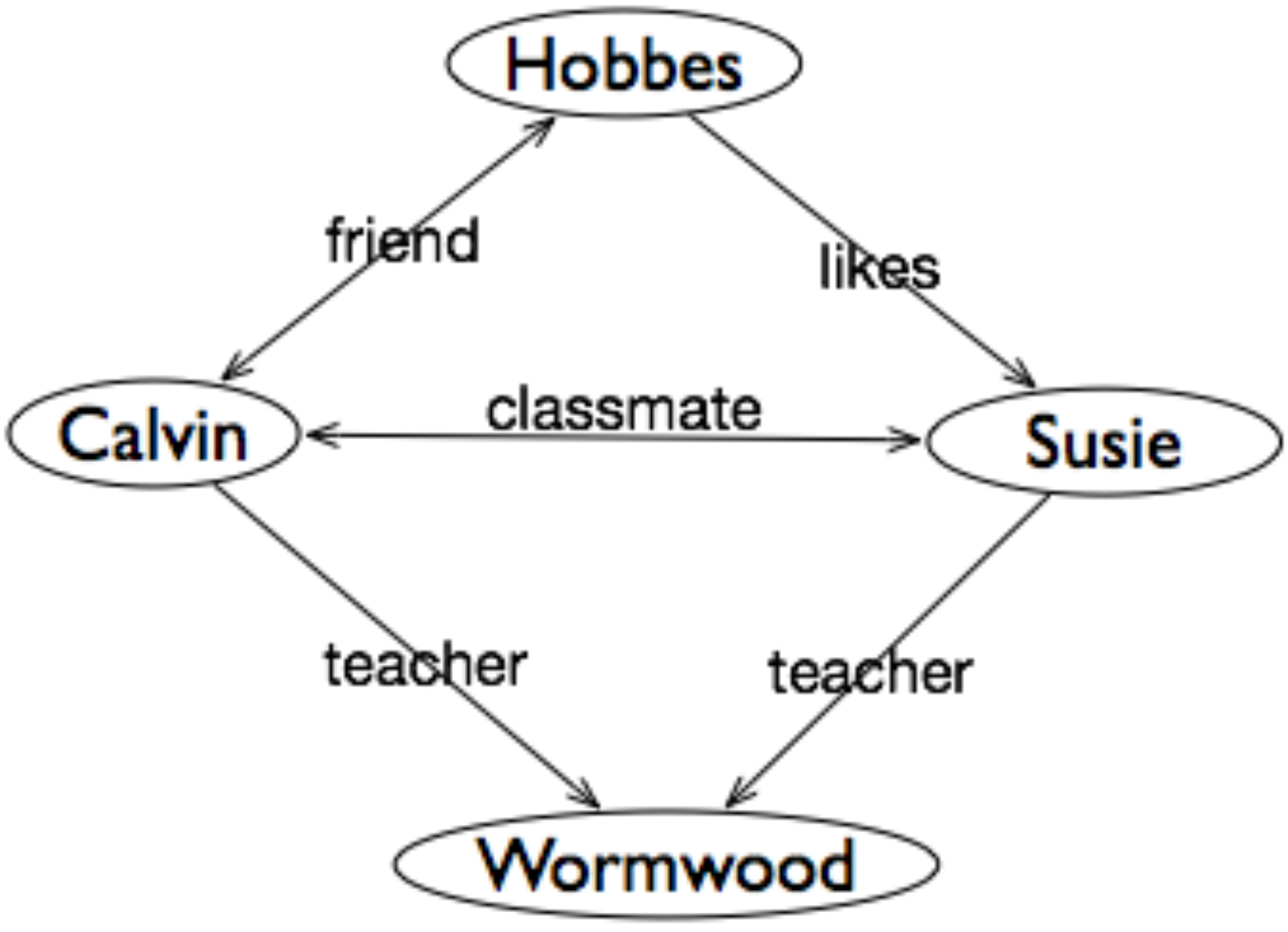
- So, given these facts:



- one can infer the following:



Graph-based data model



Reification

Capturing RDF triple data

- Ontology Editors
  - Protégé / WebProtégé (Stanford)
  - TopBraid Composer (TopQuadrant)
- Libraries
  - Apache Jena; OWL API; RDFLib;

Vocabularies

- Referring to particular subjects, properties and objects in triples means we are dealing with vocabularies

- That is, naming things and using names introduced by others

  - “This ‘SST’ dataset was produced by organization ‘Acme’”

What about ontologies?

- Vocabularies <u>are</u> ontologies
- A way to think of a possible (loose) differentiation:
  - Tend to use “ontology” when the resources in your triples and the relationships among those resources are increasingly more elaborate in terms of intended semantics
- Let’s use “vocabulary” and “ontology” interchangeably here

Vocabularies

- Should be controlled vocabularies:
    - with names (and associated definitions/attributes) agreed by the community
    - to reduce discrepancies
    - to facilitate data discovery, reuse, and integration
    - to enable crosswalks/mappings
    - is short, to promote and facilitate interoperability

Controlled vocabulary example:

CF Standard names

- http://cfconventions.org/standard-names.html
- Precise description of 2,700+ physical quantities
  - name
  - description
  - canonical units

Name	Definition	Units
acoustic_signal_roundtrip_travel_time_in_sea_water	The quantity with standard name acoustic_signal_roundtrip_travel_time_in_sea_water is the time taken for an acoustic signal to propagate from the emitting instrument to a reflecting surface and back again to the instrument. In the case of an instrument based on the sea floor and measuring the roundtrip time to the sea surface, the data are commonly used as a measure of ocean heat content.	s
aerodynamic_particle_diameter	The diameter of a spherical particle with density 1000 kg m-3 having the same aerodynamic properties as the particles in question.	m
aerodynamic_resistance	The "aerodynamic_resistance" is the resistance to mixing through the boundary layer toward the surface by means of the dominant process, turbulent transport. Reference: Wesely, M. L., 1989, doi:10.1016/0004-6981(89)90153-4.	m-1 s
age_of_sea_ice	"Age of sea ice" means the length of time elapsed since the ice formed.	year
age_of_stratospheric_air	"Age of stratospheric air" means an estimate of the time since a parcel of stratospheric air was last in contact with the troposphere.	s
age_of_surface_snow	"Age of surface snow" means the length of time elapsed since the snow accumulated on the earth's surface. The surface called "surface" means the lower boundary of the atmosphere.	day
air_density		kg m-3
air_potential_temperature	Potential temperature is the temperature a parcel of air or sea water would have if moved adiabatically to sea level pressure.	K

Vocabularies to use in your vocabularies

- **RDF**: (*Resource Description Framework*)
  - type, Property, Statement, …
  - subject, predicate, object, …
- **RDFS**: (*RDF Schema*)
  - Resource, Class, subClassOf, subPropertyOf,…
  - comment, label, seeAlso, isDefinedBy, …

Vocabularies to use in your vocabularies

- **SKOS**: (*Simple Knowledge Organization System*)
  - definition, note, …
  - exactMatch, closeMatch, relatedMatch, …
- **OWL**: (*Web Ontology Language*)
  - Ontology, inverseOf, ReflexiveProperty , …
  - sameAs, versionInfo, …

Vocabularies to use in your vocabularies

- **DCT**: (*Dublin Core Terms*)
  - title, description, creator, contributor…
  - rights, license, …
- **OMV**: (Ontology Metadata Vocabulary)
  - name, description, hasCreator, keywords,…
  - sameAs, …

Does semantic interoperability need an overarching vocabulary?

- No! … and such a goal is overly unrealistic in general
- But it's fine to
  - Define what makes sense to your case
  - Map your names to names is other vocabularies as convenient/needed for interoperability
  - Propose additions to common vocabularies

Vocabularies: Summary

- Use standard vocabularies
  - in your data/metadata
  - in your own vocabularies, too!
- Participate in community vocabulary development activities

All of the above in practice:

ORR – Ontology Registry and Repository

ORR Origins

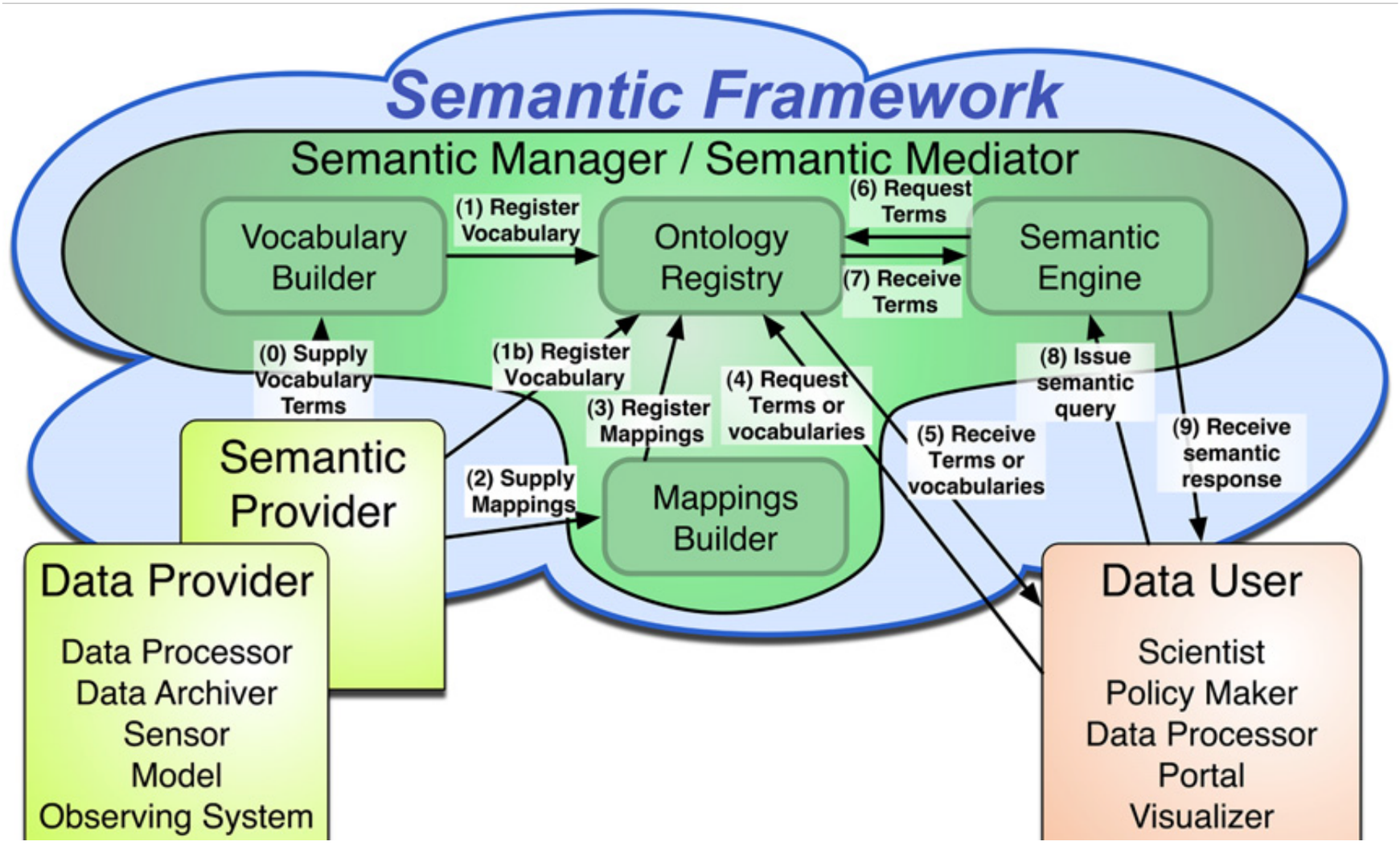
- *MMI* – Marine Metadata Interoperability project
  - https://marinemetadata.org/

> *"Promoting the exchange, integration and use of marine data through enhanced data publishing, discovery, documentation and accessibility."*

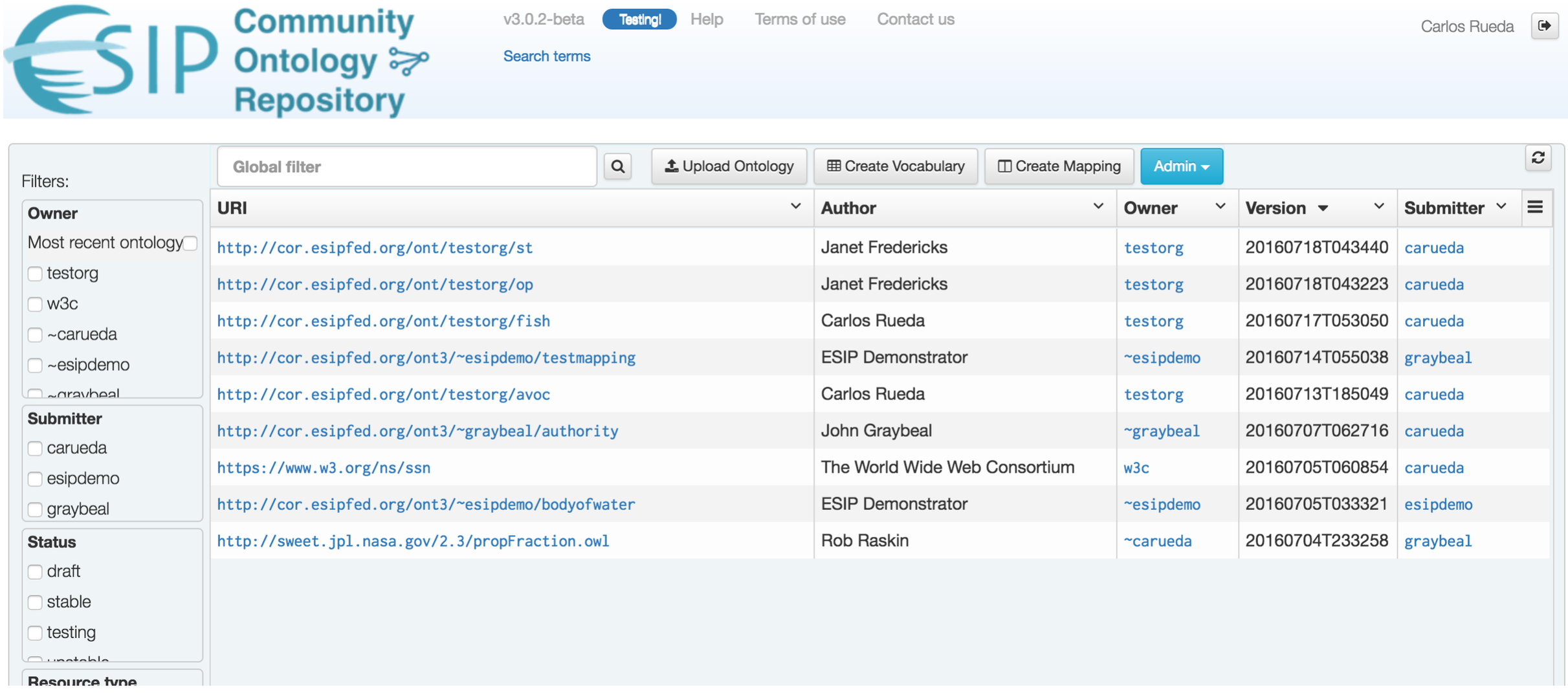
- ORR born as part of MMI's vision for a *Semantic Framework*

ORR Origins

- MMI's vision for a *Semantic Framework*



MMI ORR (v.3)


Community Ontology Repository
v3.0.2-beta
Testing!
Help
Terms of use
Contact us
Search terms
Carlos Rueda
Filters:
Global filter
Upload Ontology
Create Vocabulary
Create Mapping
Admin
Owner
Most recent ontology
testorg
w3c
~carueda
~esipdemo
Submitter
carueda
esipdemo
graybeal
Status
draft
stable
testing
Resource type
URI
Author
Owner
Version
Submitter
http://cor.esipfed.org/ont/testorg/st
Janet Fredericks
testorg
20160718T043440
carueda
http://cor.esipfed.org/ont/testorg/op
Janet Fredericks
testorg
20160718T043223
carueda
http://cor.esipfed.org/ont/testorg/fish
Carlos Rueda
testorg
20160717T053050
carueda
http://cor.esipfed.org/ont3/~esipdemo/testmapping
ESIP Demonstrator
~esipdemo
20160714T055038
graybeal
http://cor.esipfed.org/ont/testorg/avoc
Carlos Rueda
testorg
20160713T185049
carueda
http://cor.esipfed.org/ont3/~graybeal/authority
John Graybeal
~graybeal
20160707T062716
carueda
https://www.w3.org/ns/ssn
The World Wide Web Consortium
w3c
20160705T060854
carueda
http://cor.esipfed.org/ont3/~esipdemo/bodyofwater
ESIP Demonstrator
~esipdemo
20160705T033321
esipdemo
http://sweet.jpl.nasa.gov/2.3/propFraction.owl
Rob Raskin
~carueda
20160704T233258
graybeal

MMI ORR (v.3)

- Enhanced user/organization/permission model
- Overhauled authentication mechanism
- Enhanced performance
- RESTful backend endpoint
- MongoDB; AllegroGraph
- Backend: Scala; comprehensive tests; Travis CI; good coverage
- Front-end: AngularJS
- Docker images for streamlined installation of integrated system
  https://hub.docker.com/r/mmisw/orr-ont/tags/

MMI ORR (v.3)

- Status
  - Recently transitioned to *beta*
    …mostly according to internal testing
  - So, please help us as we make progress toward a stable version. Your feedback is most welcome!

ORR

- Registry
  - ORR is a catalog of pointers to ontologies and associated metadata
- Repository
  - ORR hosts the registered ontologies

ORR Capabilities

- Repository of controlled vocabularies and term mappings
- Web resolvable identifiers for ontologies and terms
- Enable added-value applications with semantic and inference
- Ontology metadata
- Versioning

Key requirements

- Community driven, collaborative creation
- Easy-to-use tools

Client applications–ORR interactions

- **Data Portals** create/use ontologies that capture categories to be exposed
- **Data providers** create/use ontologies:
  - For the terms (concepts) used in their data products and services
  - With mappings between **Data Provider**'s terms and **Data Portal** categories
- **Data Portal and client applications**
  - Access; Resolve; Query; Aggregate; Archive; ...

ORR instances

- mmisw.org – MMI ORR

- cor.esipfed.org – ESIP COR

- sensorml.com – SensorML ORR