Semantics for Cyberinfrastructure: The iPlant Collaborative’s Semantic Web Platform

VIRTUAL WORKSHOP
SEMANTICS IN GEOSPATIAL AND OTHER ARCHITECTURES: DESIGN AND IMPLEMENTATION
May 2013

DAMIAN GESSLER, Ph.D.
SEMANTIC WEB ARCHITECT
UNIVERSITY OF ARIZONA

dgessler (at) iplantcollaborative (dot) org
The iPlant Collaborative

NSF-funded, multi-institutional, collaborative project similar in scope to EarthCube.

Mission:
To build a cyberinfrastructure for the nation’s plant scientists.

Successfully completed its first five years and is anticipatory of the next (May 2013)

www.iplantcollaborative.org
All-too-Familiar Challenges

1. Disparate data
2. Varying research agendas
3. Lack—or proliferation—of standards
4. Disruptive technologies (NGS)
5. Rapidly changing science
6. Global externalities
7. Hierarchical, non-linear systems
8. Finite time, money, and resources
A Constrained Solution Space

Perfect solutions either do not exist, or are not feasibly attainable under the constraints.

Yet even good solutions can be powerfully transformative (e.g., WWW in 1991)
Enterprise and HPC Assets

Enterprise Class
Discovery Environment, Atmosphere

Foundational Infrastructure
iPlant Data Store, HPC, etc.

Enterprise Class
Virtual Workdesk
Cloud, Virtual machines

High Performance Computing
PFLOPS-scale compute power
Petabyte scale storage
Bridging Enterprise, HPC, and Web Assets

Web

Enterprise Class
Discovery Environment, Atmosphere

Semantic Integration
Semantic Pipelining

Foundational Infrastructure
iPlant Data Store, HPC, etc.

Distributed
Semantic Web Services
Logic-driven semantics

Enterprise Class
Virtual Work desk
Cloud, Virtual machines

High Performance Computing
PFLOPS-scale compute power
Petabyte scale storage
The Actors

Semantic Mediation Layer

The World

You

Your lab

The World

iPlant Computational Resources

Community MODS (Model Organism Databases) and CODS (Clade Oriented Databases)
The Antibody Analogy as a Semantic Mediation Layer

Variable region

- service parameters
- input data
- output data
-resource
- subject
- object
- performs mapping
- maps to

Conserved region

antigen binding site
light chain
heavy chain
Establish the framework for web resources to describe themselves and their offerings
Establish the framework for ontological integration
Engage first-order, description logic reasoning
Provide a semantically enabled Discovery Server for service and pipeline coordination

http://sswap.info/protocol
iPlant Semantic Architecture

Semantic Broker / Discovery Server

Repository

Broker

Explicit Interface

Indirection Layer

Explicit Interface

Interpreter

Indirection Layer

Semantic documents described in this talk:
PDG: Provider Description Graph
RDG: Resource Description Graph
RIG: Resource Invocation Graph
RRG: Resource Response Graph
RQG: Resource Query Graph

Data and Service Providers

Ontologies

Clients

Web Resource

Ontology Servers

Protocol Ontology

Client

Data

Algorithm

OWL RDF/XML

OWL RDF/XML

OWL RDF/XML
sswap.info/example
Semantic Integration from Third-party Web sites
TreeGenes’ DiversiTrie

Javascript snippet to launch data for Web Discovery with the press of a button
Web Discovery into Semantic Pipelines

Reasoner-assisted Web workflows

Reasoner uses first-order, description logic to present services and pipelines that can operate on the data at any given step.
Just-In-Time Ontology Hosting

sswap.info/jit
RESTful Pipeline Execution
Data Tree View

“Ontologized” data and metadata

SSWAP enables on-the-fly Data Tree views
Semantic Integration into Third-party Web sites
TreeGenes’ CartograTree

Third-party web sites can engage as renderers on result sets ...
CartograTree
Custom user interface data selection and analysis

... enabling directed user experiences back into Web Discovery
Semantic Integration from Third-party Web sites
Data slicing and contextual augmentation
Direct and Indirect Data Referencing
URI dereferencing of arbitrarily large data sets

Serialize the data itself, or a URI to where the data is located
High Performance Computing
Services engage like any other Web services
Publishing Pipelines
Private data, shared service parameterization
Phylogenetics
Pipeline runs are persisted in OWL and can start new pipelines
TreeViz
Multi-state, multi-institution, web/HPC run
Quick Vitals

• 185,000 lines of code
• 100+ libraries
• Open source and freely available
• Three 48 GB RAM servers
• More info: sswap.info/wiki
Acknowledgements

Special thanks to:

- iPlant Collaborative
- Semantic Web engineering by Clark and Parsia, LLC
- NSF grants #0943879 and #EF-0735191