SEMANTIC GRID

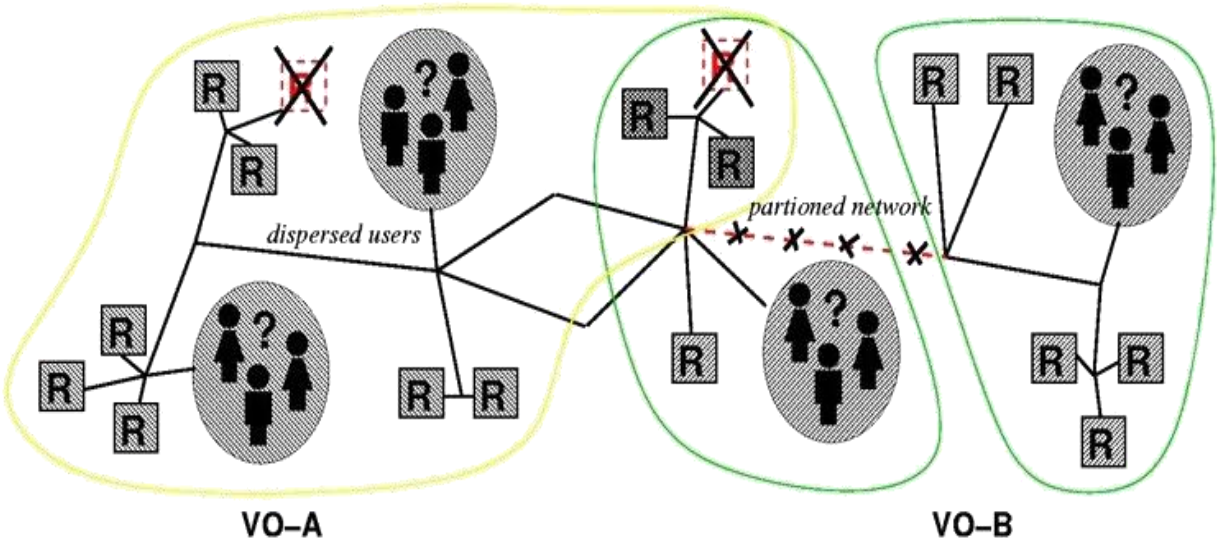
Pham Thi Bich Lien

Nguyen Van Nhat

Nguyen Duy Thien

What is Grid?

- flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources - virtual organizations.



Why we need Semantic Grid?

- **Problem:** enabling scientists to generate, analyse, share and discuss their insights, experiments and results in a more **effective manner**.
- there is a high degree of **easy-to-use** and seamless automation and in which there are flexible collaborations and computations on a global scale.

Why we need Semantic Grid?

- **Example**: To illustrate, consider if a machine's operating system is described as "SunOS" or "Linux." To query for a machine that is "Unix" compatible, a user either has to:

1. Explicitly incorporate the Unix compatibility concept into the request requirements by requesting a disjunction of all Unix-variant operating systems, e.g., (OpSys="SunOS" || OpSys="Linux")

Why we need Semantic Grid?

2. Wait for all interesting resources to advertise their operating system as Unix as well as either Linux or SunOS, e.g., (OpSys="SunOS," "Unix"), and then express a match as set-membership of the desired Unix value in the OpSys value set, e.g., hasMember(OpSys, "Unix").

- **Apply Semantics…**

  - Knowledge base: "SunOS and Linux are types of Unix operating system"

  - Request: "Need the Unix compatibility OS"

Why we need Semantic Grid?

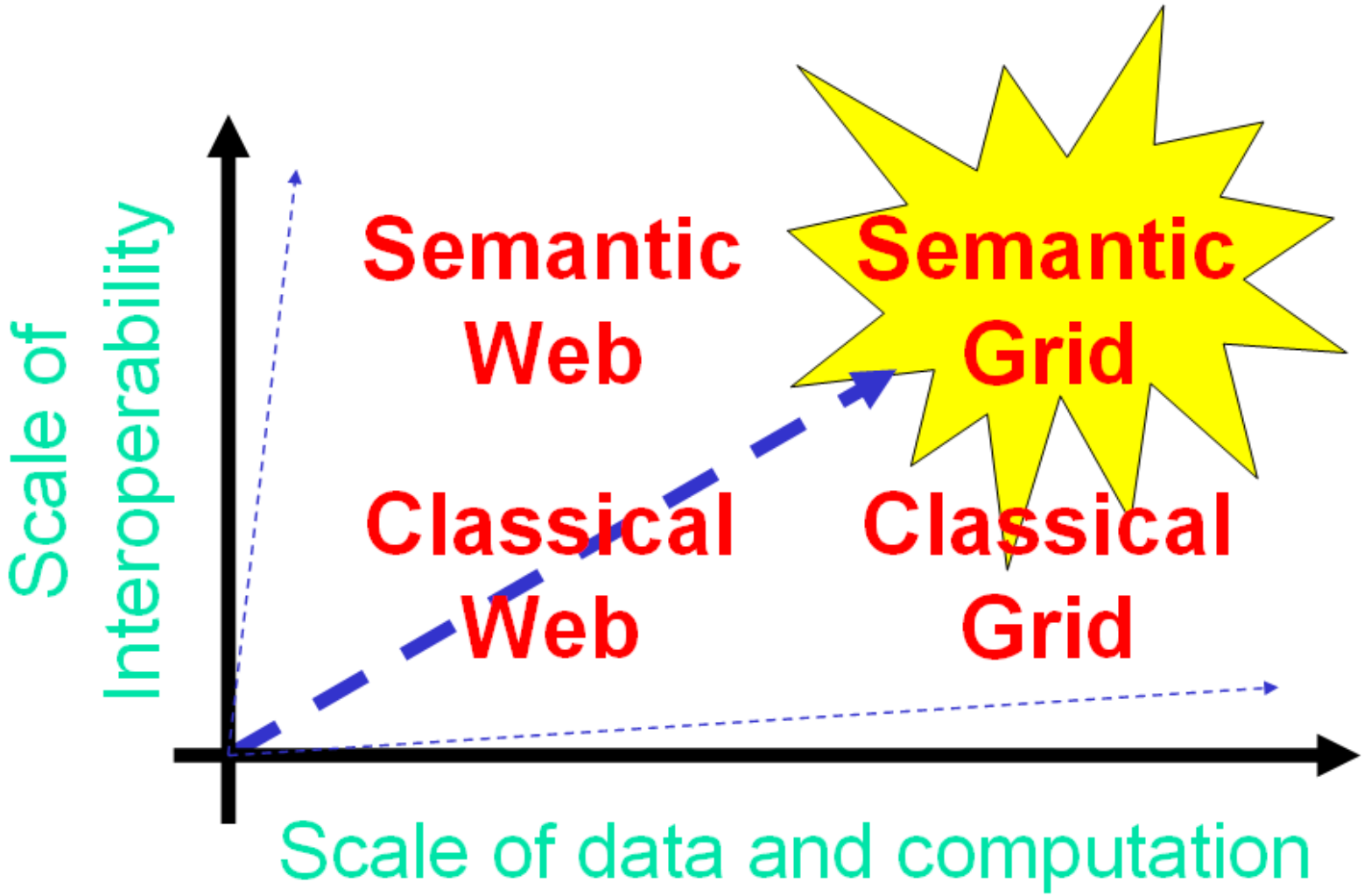
Grid

"It is a truth universally acknowledged, that an application in possession of good middleware, must be in want of meaningful metadata."

Semantic

-- prof. C. Goble

What is Semantic Grid?



What is Semantic Grid?

- An extension of the current Grid in which information and services are given well-defined and explicitly represented meaning, so that it can be shared and used by humans and machines, better enabling them to work in cooperation

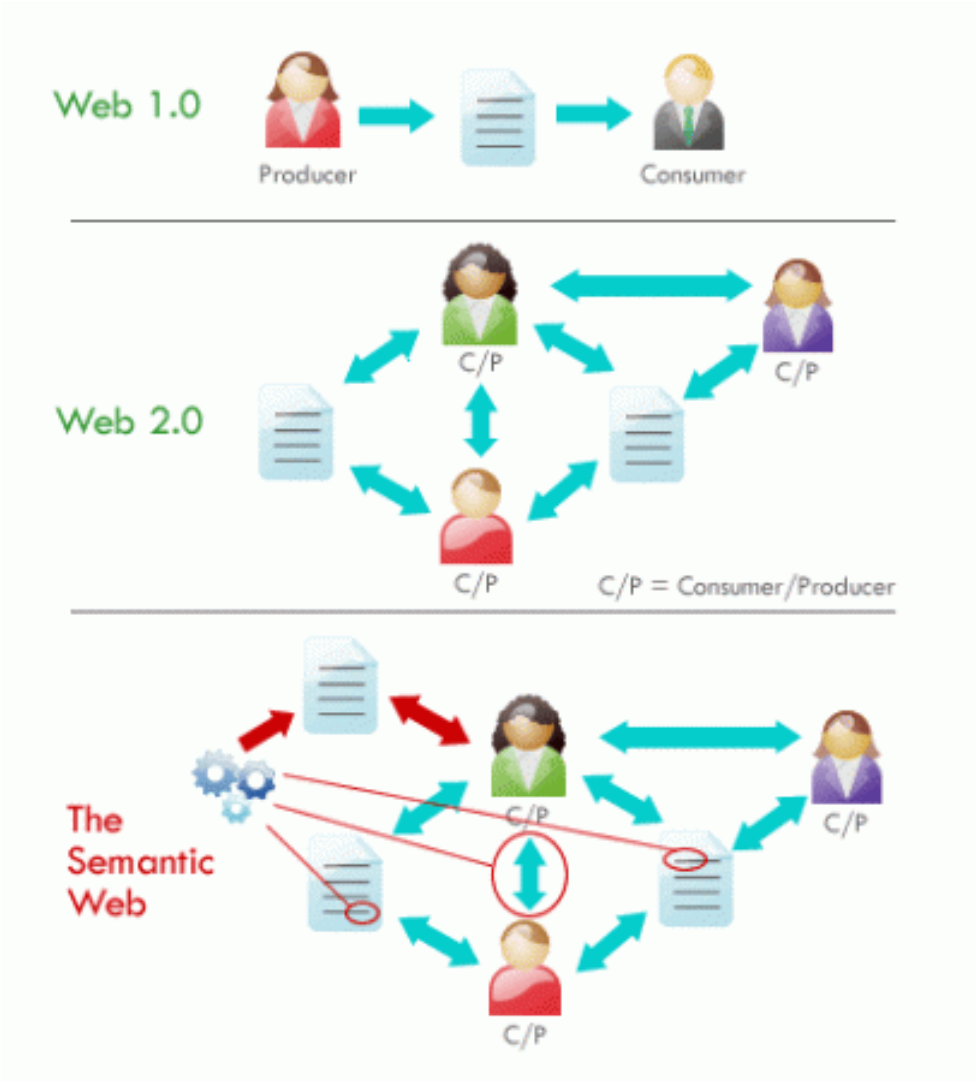
What is Semantic Grid?

- An extension of the Grid
- Rich metadata is exposed and handled explicitly, shared, and managed via Grid protocols

What is Semantic Grid?

- The Semantic Grid uses metadata to describe information in the Grid.
- Turning information into something more than just a collection of data means understanding the context, format, and significance of the data.
- Therefore:
  - Understand information
  - Discovery and reuse

Semantic Web



Semantic Web

- "…to create a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data.
- Facilities to put **machine-understandable data** on the Web are quickly becoming a high priority for many organizations, individuals and communities.

Semantic Web

- The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people.
- For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently.”

Semantic Web

- Current Web (WWW)
  - Is a huge library of interlinked documents that are transferred by computers and presented to people
  - Anyone can contribute to it
  - Quality of information or even the persistence of documents cannot be generally guaranteed
  - Contains a lot of information and knowledge, but machines usually serve only to deliver and present the content of documents describing the knowledge
  - People have to connect all the sources of relevant information and interpret them themselves

Machine **can Process the content**

But

Machine **can't Understand content**

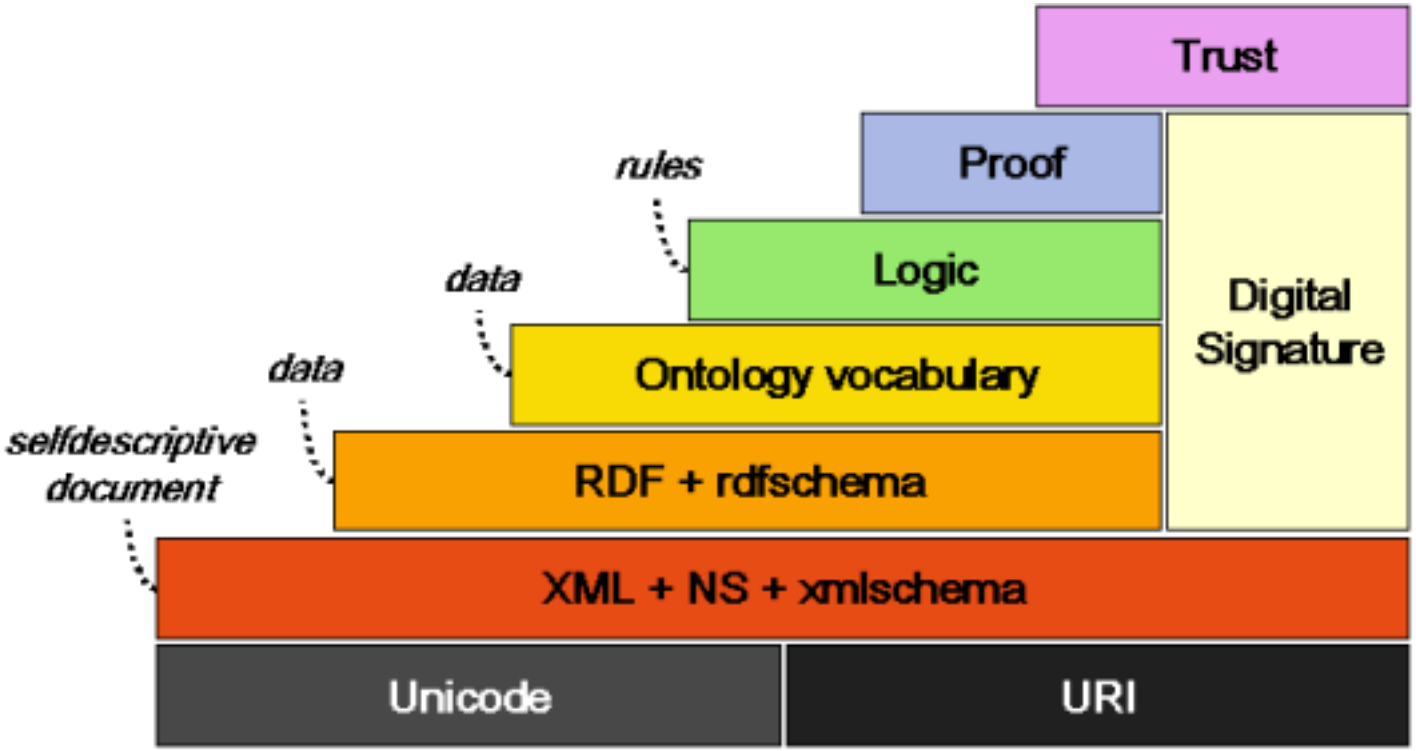
Semantic Web

- Definition

  The **Semantic Web** is an extension of the current web in which the semantics of information and services on the web is defined, making it possible for the web to understand and satisfy the requests of people and machines to use the web content.

  --- Tim Berners-Lee

Semantic Web

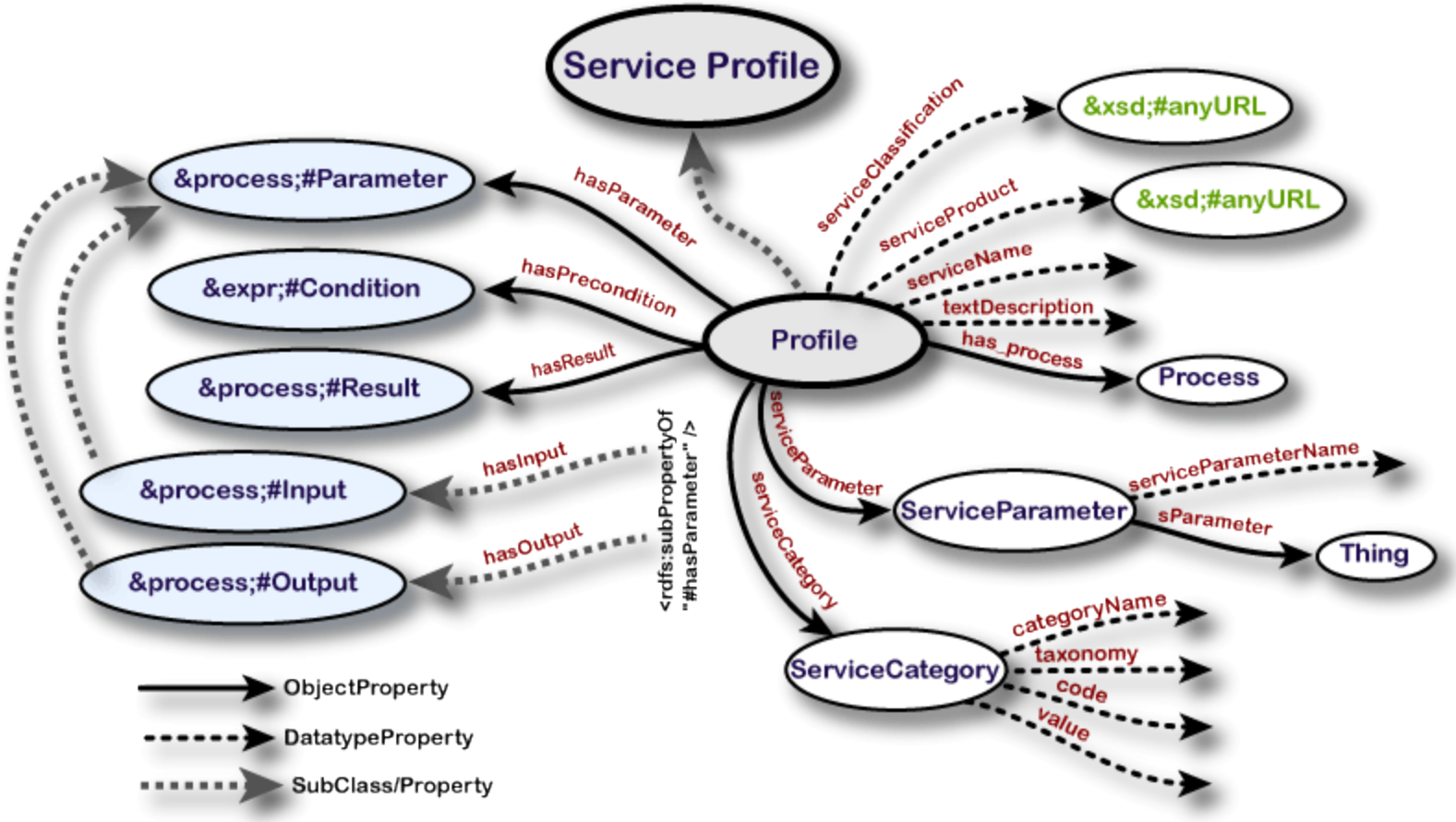


Semantic Web

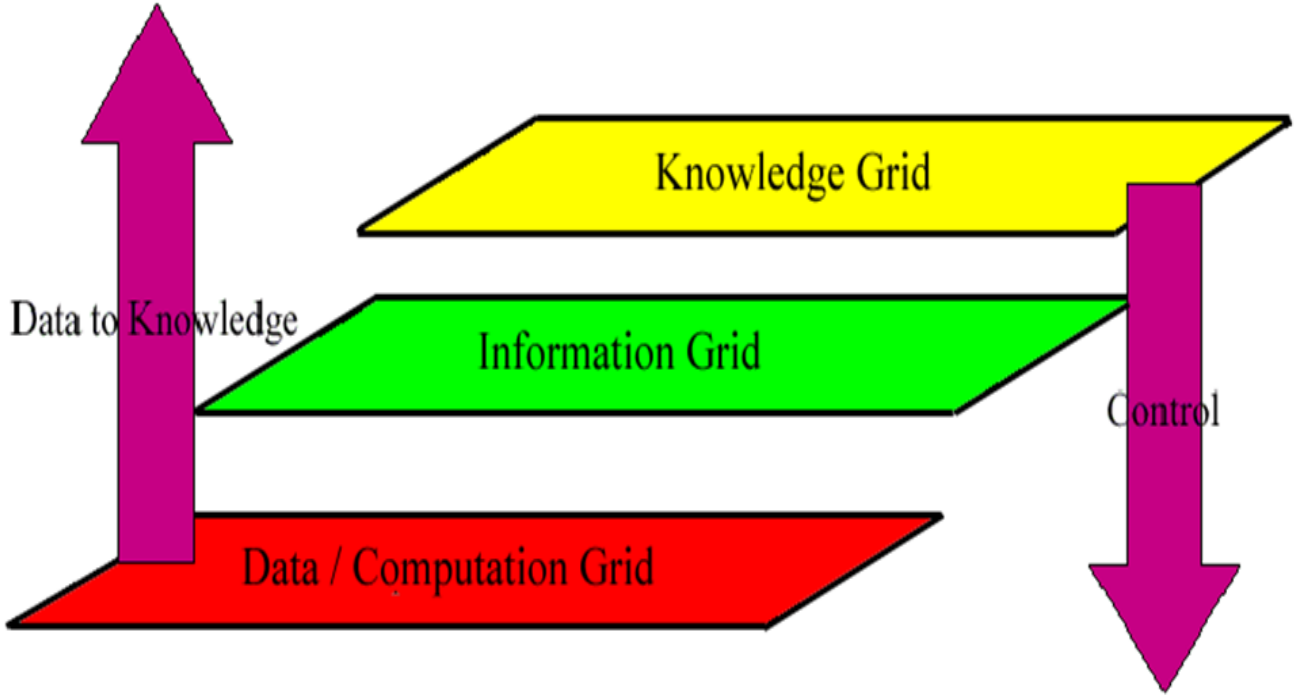
- **Ontology** is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to describe the domain.
- Implement by XML, XML Namespace, XML Schema, RDF, RDF Schema và OWL

Semantic Web

- **Ontology example**



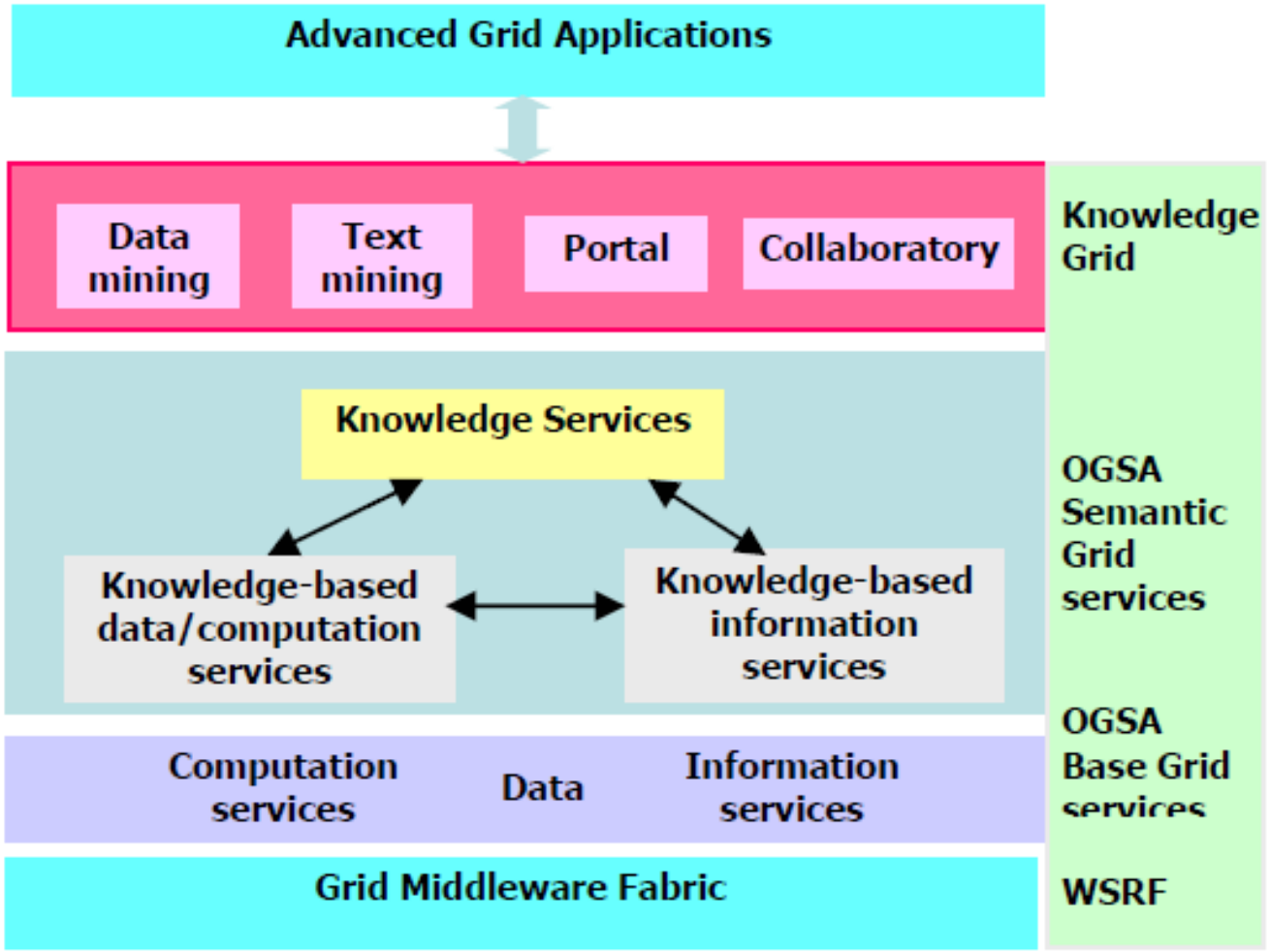
Semantic Grid Requirement



Semantic Grid Requirement

- *Resource description, discovery and use.*
- *Process description and enactment.*
- *Autonomic behaviour.*
- *Security and trust.*
- *Annotation.*
- *Information Integration.*
- *Synchronous information streams and fusion.*
- *Context-aware decision support.*
- *Communities.*
- *Smart environments.*

Semantic Grid Architecture



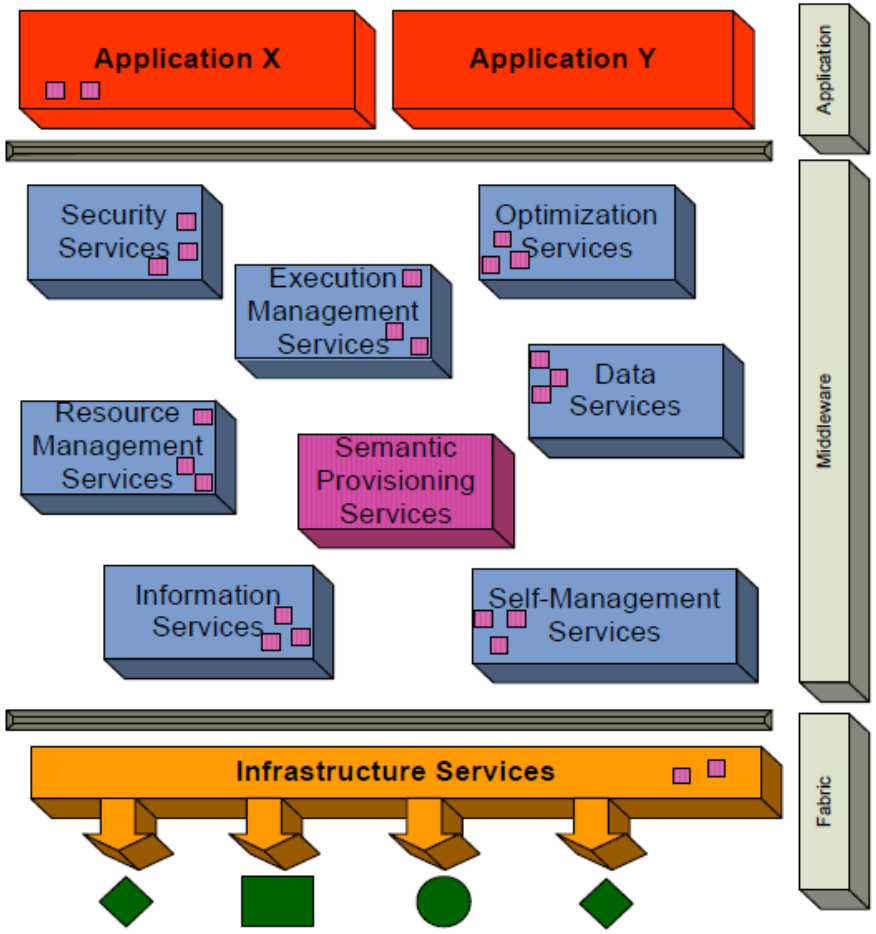
S-OGSA

- A Grid usually consist of several different services by OGSA:
  - VO management service
  - Resource discovery and Management service
  - Job Management service
  - Security service
  - Data Management service
- The S-OGSA should (will) provide the metadata + semantic services to those services.

S-OGSA

- The Solution:
  - Attached the semantic to Grid entities.
  - Binding them together by semantic binding service.
  - Normal grid services can be “semantic” by the semantic binding service.

S-OGSA


Application X
Application Y
Application
Security Services
Optimization Services
Execution Management Services
Data Services
Resource Management Services
Semantic Provisioning Services
Information Services
Self-Management Services
Middleware
Infrastructure Services
Fabric

DESIGN PRINCIPLES FOR A REFERENCE SEMANTIC GRID ARCHITECTURE

- **Parsimony** of architecture elements.
- **Extensibility** of the framework.
- **Uniformity** of the mechanisms.
  - Similar to the Grid resources they are associated with, knowledge and metadata should exhibit manageability aspects.
  - S-OGSA must encapsulate both stateless and stateful Grid services, as OGSA does.
  - Knowledge services in S-OGSA are OGSA-observant Grid services.

DESIGN PRINCIPLES FOR A REFERENCE SEMANTIC GRID ARCHITECTURE

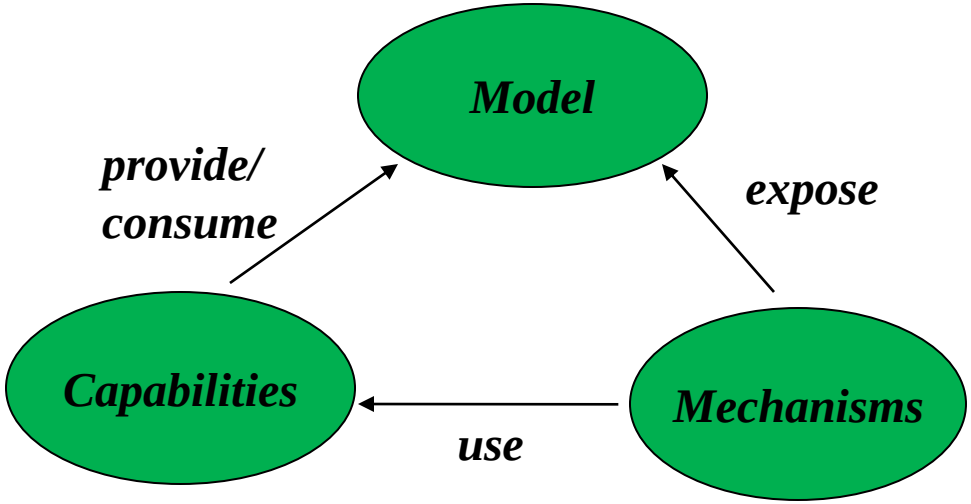
- **Diversity** of semantic capabilities
  - Ignorant of the associated semantics of another entity.
  - Aware that another entity has explicit associated semantics but incapable of processing it.
  - Aware that another entity has explicit associated semantics and capable of processing it, partially or completely.
- **Heterogeneity** of semantic representation.

DESIGN PRINCIPLES FOR A REFERENCE SEMANTIC GRID ARCHITECTURE

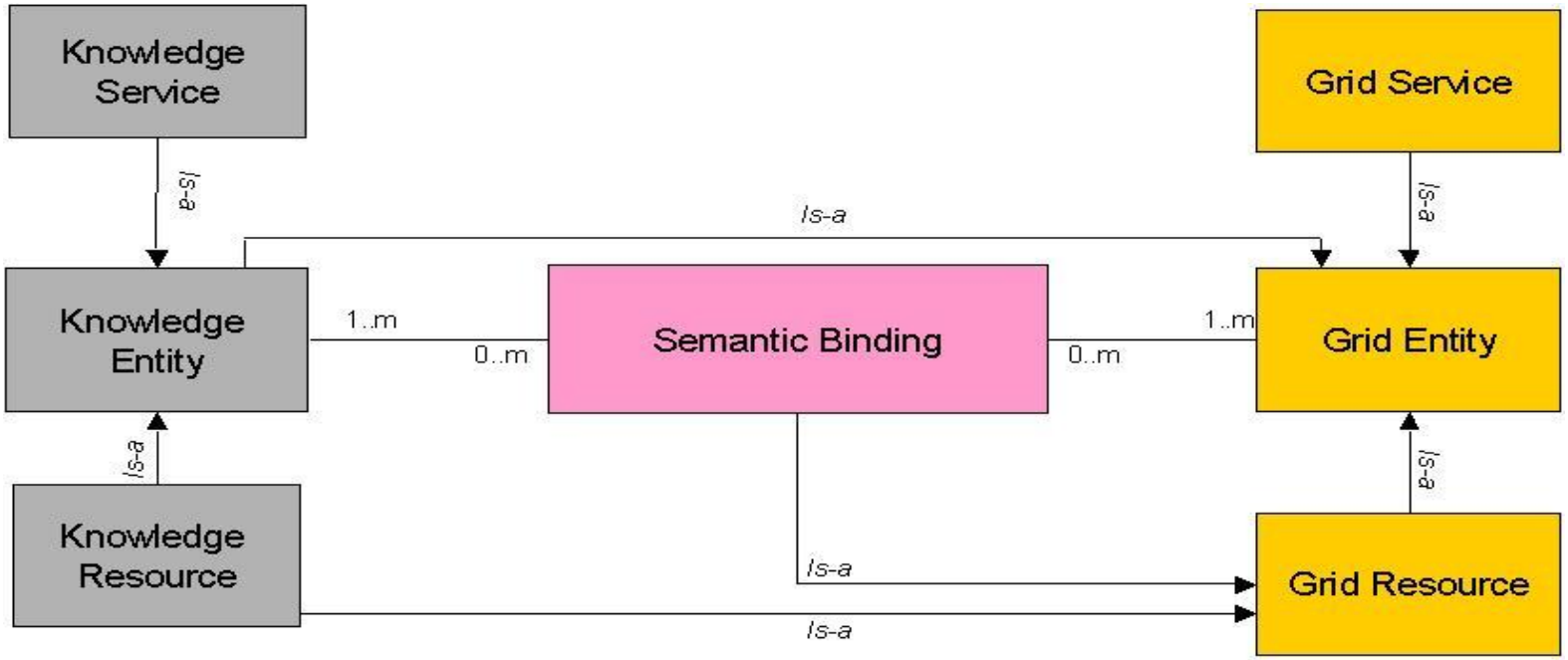
- **Enlightenment** of services
  - S-OGSA should have minimal impact on adding explicit semantics to current Grid entity interfaces or on Grid services that are ignorant of Semantic Grid entities;
  - Grid entities should not break if they can consume and process Grid resources but cannot consume and process their associated semantics (that is, if they are aware of the semantics but incapable of processing it);
  - If a Grid entity understands only part of the knowledge it consumes it should be able to use it as a best effort;
  - During their lifetime, Grid entities can incrementally acquire, lose and reacquire explicit semantics;

S-OGSA

- Defined by
  - Model
    - New entities
  - Capabilites
    - New functionalities
  - Mechanisms
    - How it is delivered

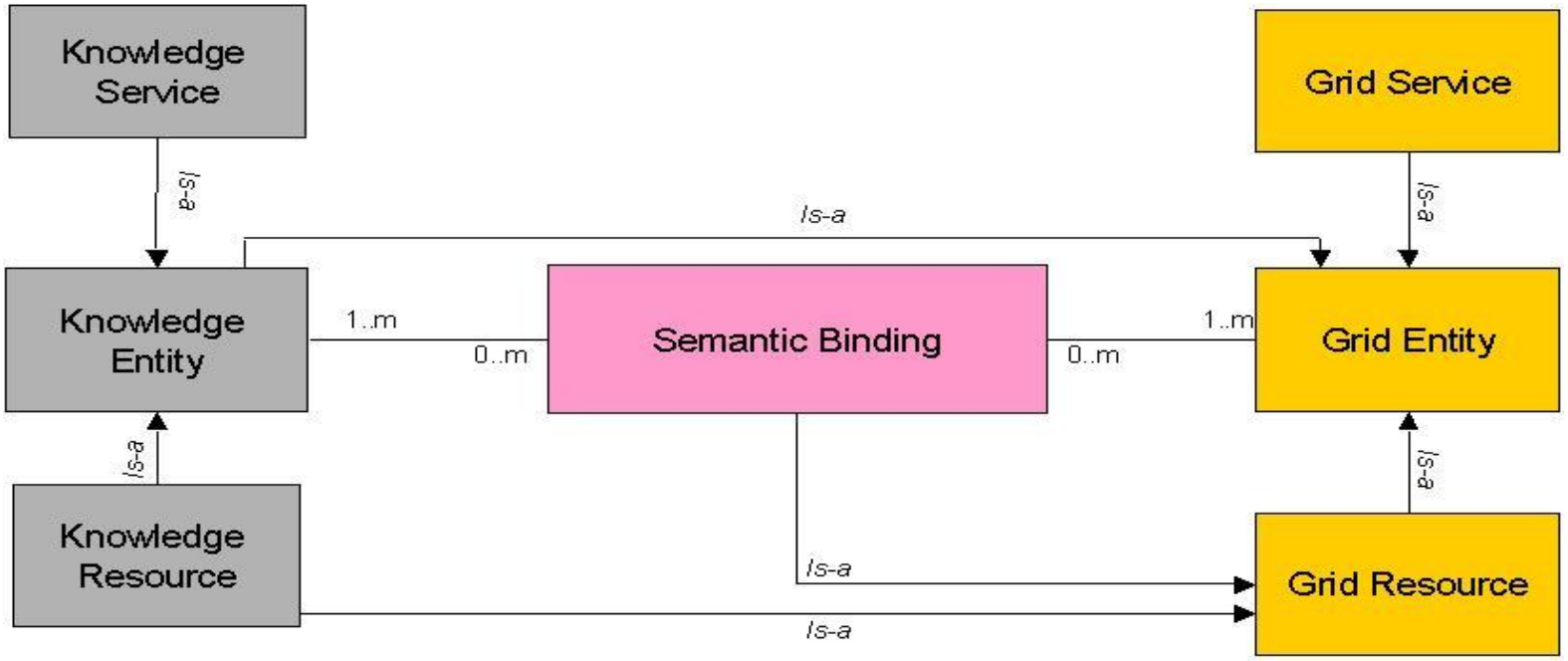


S-OGSA MODEL



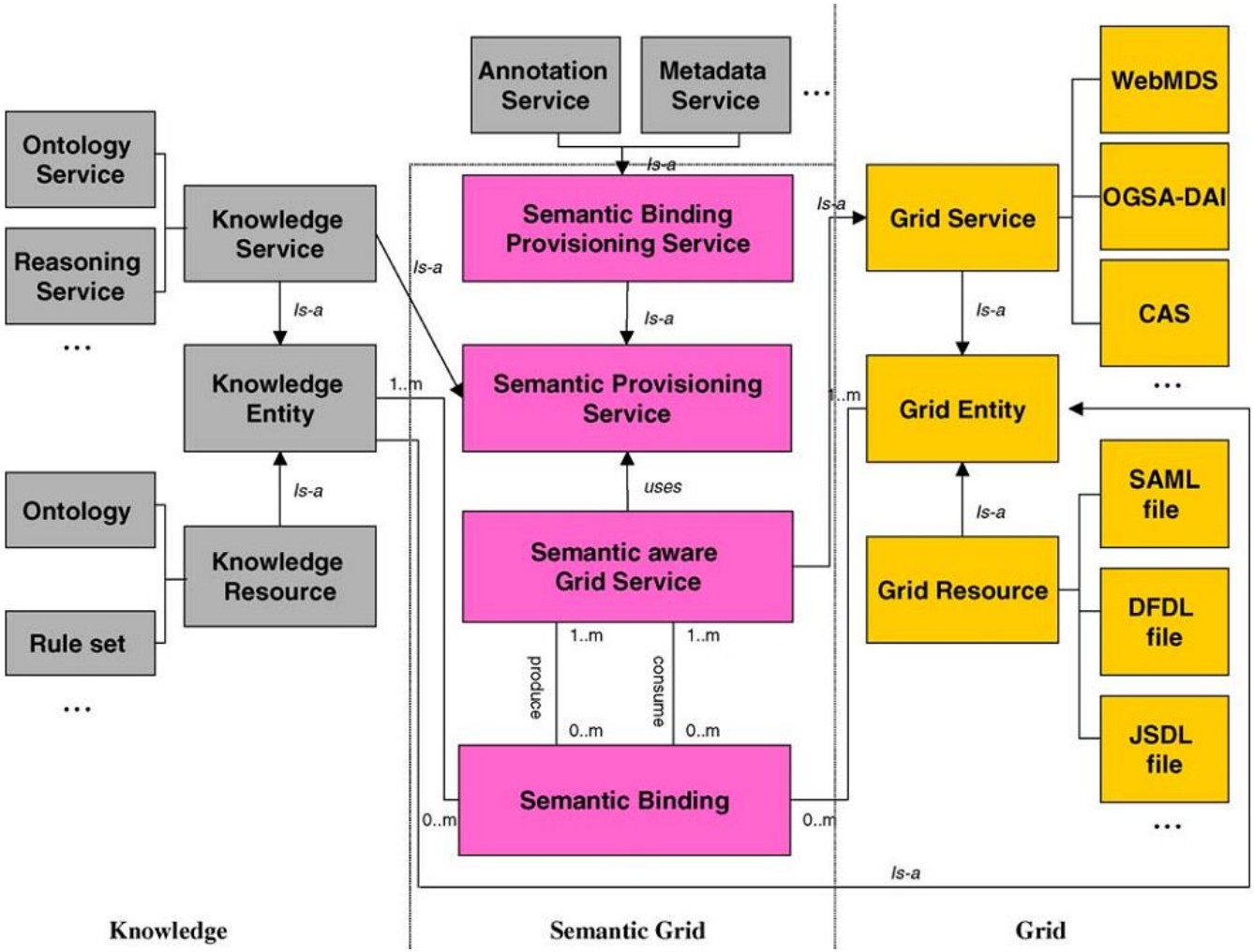
Entities in S-OGSA and their relationships.

S-OGSA MODEL



- **Grid Entities:** Grid resources and services
- **Knowledge Entities:** represent/operate with some form of knowledge (e.g ontologies, rules, knowledge bases …)
- **Semantic Bindings:** entities associate of a Grid Entity with one or more Knowledge Entities

S-OGSA CAPABILITIES



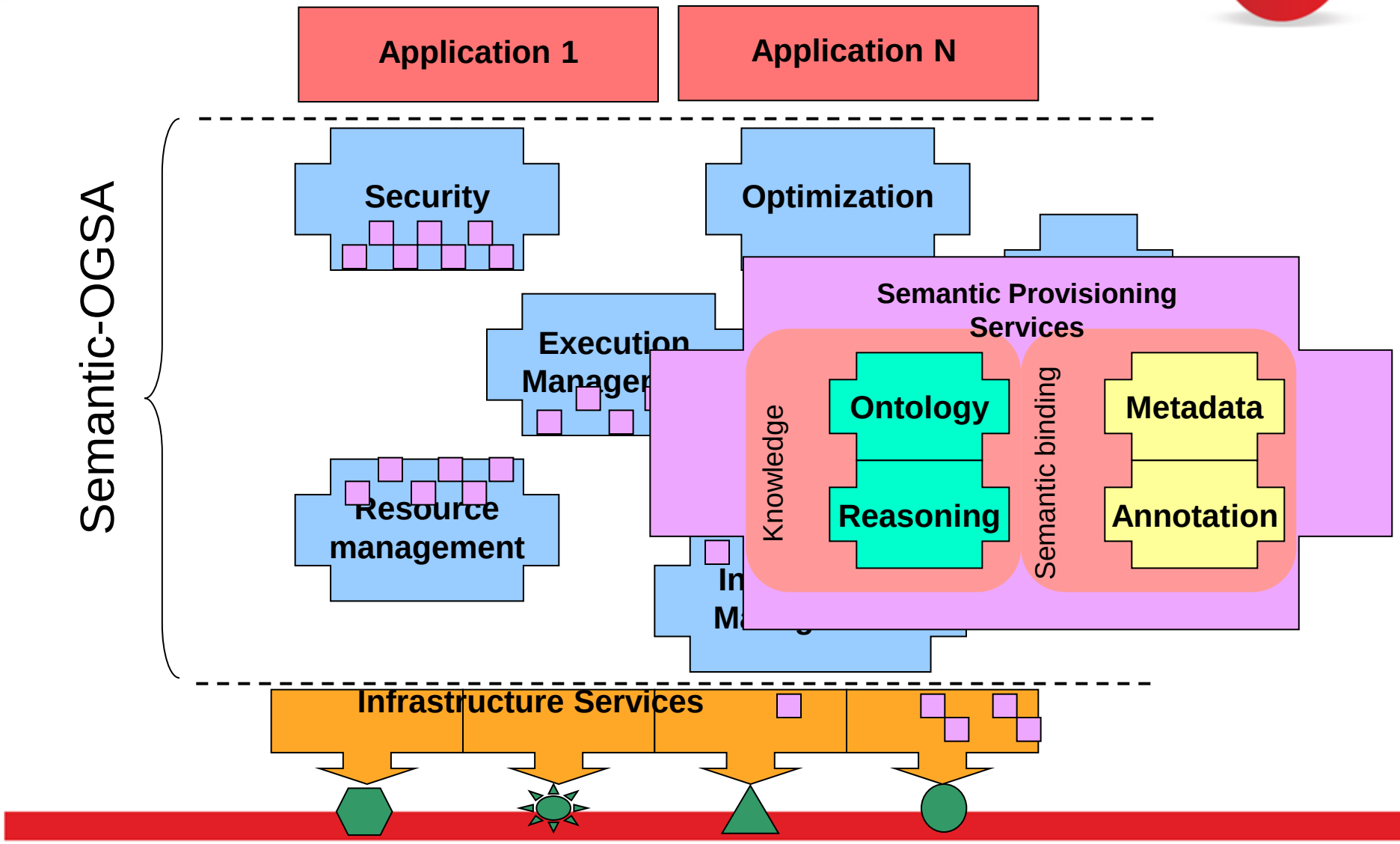
S-OGSA CAPABILITIES

- **Semantic Provisioning Services – SPS**
  - provisioning and management of explicit semantics and its association with Grid entities
  - creation, storage, update, removal and access of different forms of knowledge and metadata
  - **Knowledge provisioning services**
  - ontology services , reasoning services .
  - **Semantic binding provisioning services**
    - metadata services, annotation services .

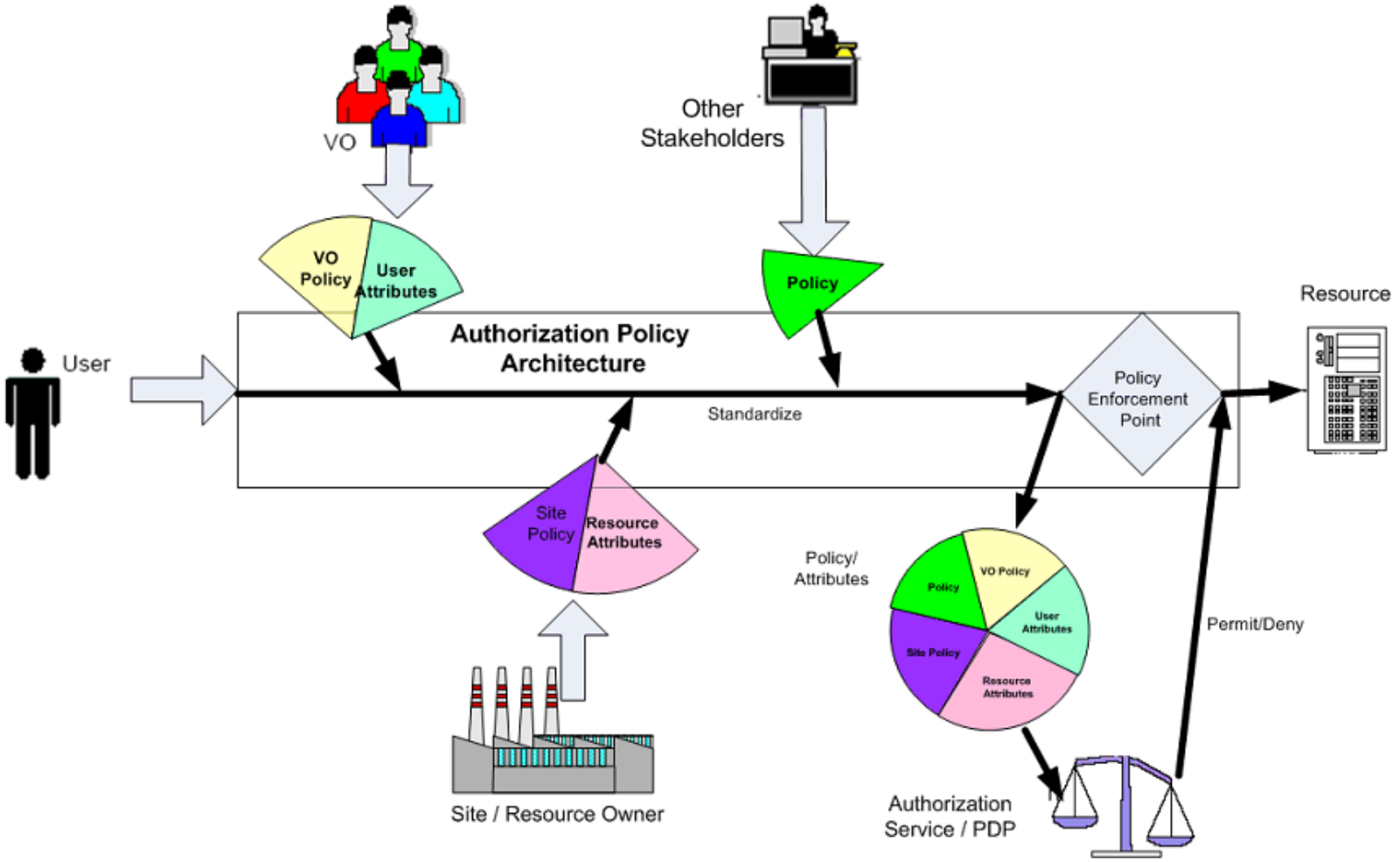
S-OGSA CAPABILITIES

- **Semantically Aware Grid Services**
  - Be able to consume Semantics Bindings and being able to take actions based on knowledge and metadata
  - Sample Actions :
    - Metadata aware authorization of a given identity by a VO Manager service
    - Execution of a search request over entries in a semantic resource catalogue
    - Incorporation of a new concept in to an ontology hosted by an ontology service

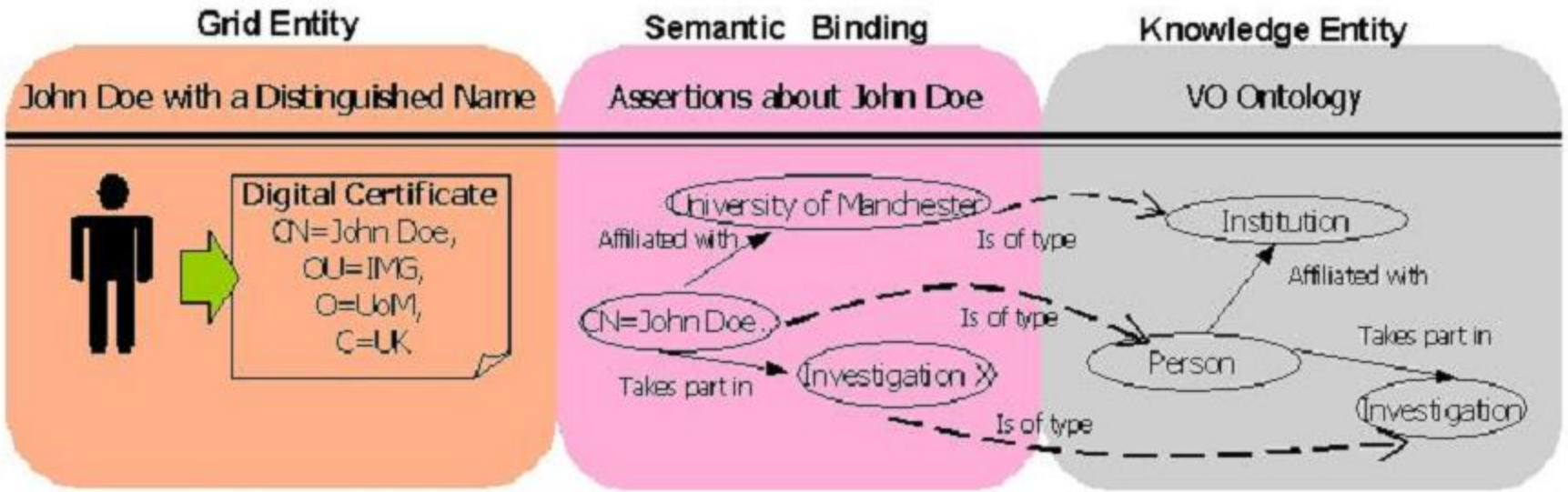
OGSA TO S-OGSA



EXAMPLE: S-OGSA MODEL AND CAPABILITIES



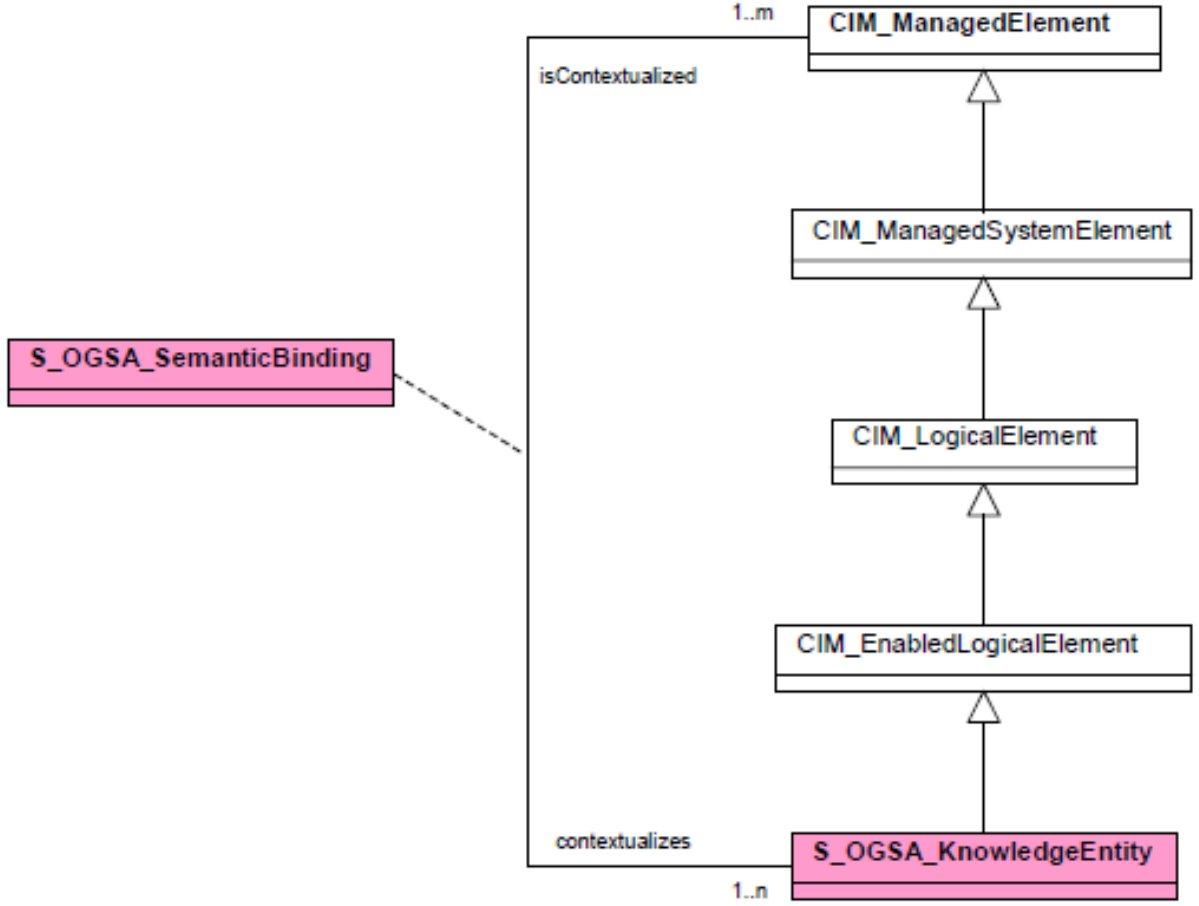
EXAMPLE: S-OGSA MODEL AND CAPABILITIES



S-OGSA MECHANISMS

- Treating Knowledge Entities and Semantic Bindings as Grid Resources
  - Common Information Model (CIM) Resource Model
  - Grid Entities : class CIM-ManagedElement in the CIM Model.
  - Knowledge Entities : class S-OGSA-KnowledgeEntity
  - S-OGSA-SemanticBinding:Semantic Binding, the association between a Grid Entity (CIM-ManagedElement) and a Knowledge Entity (S-OGSA-KnowledgeEntity)

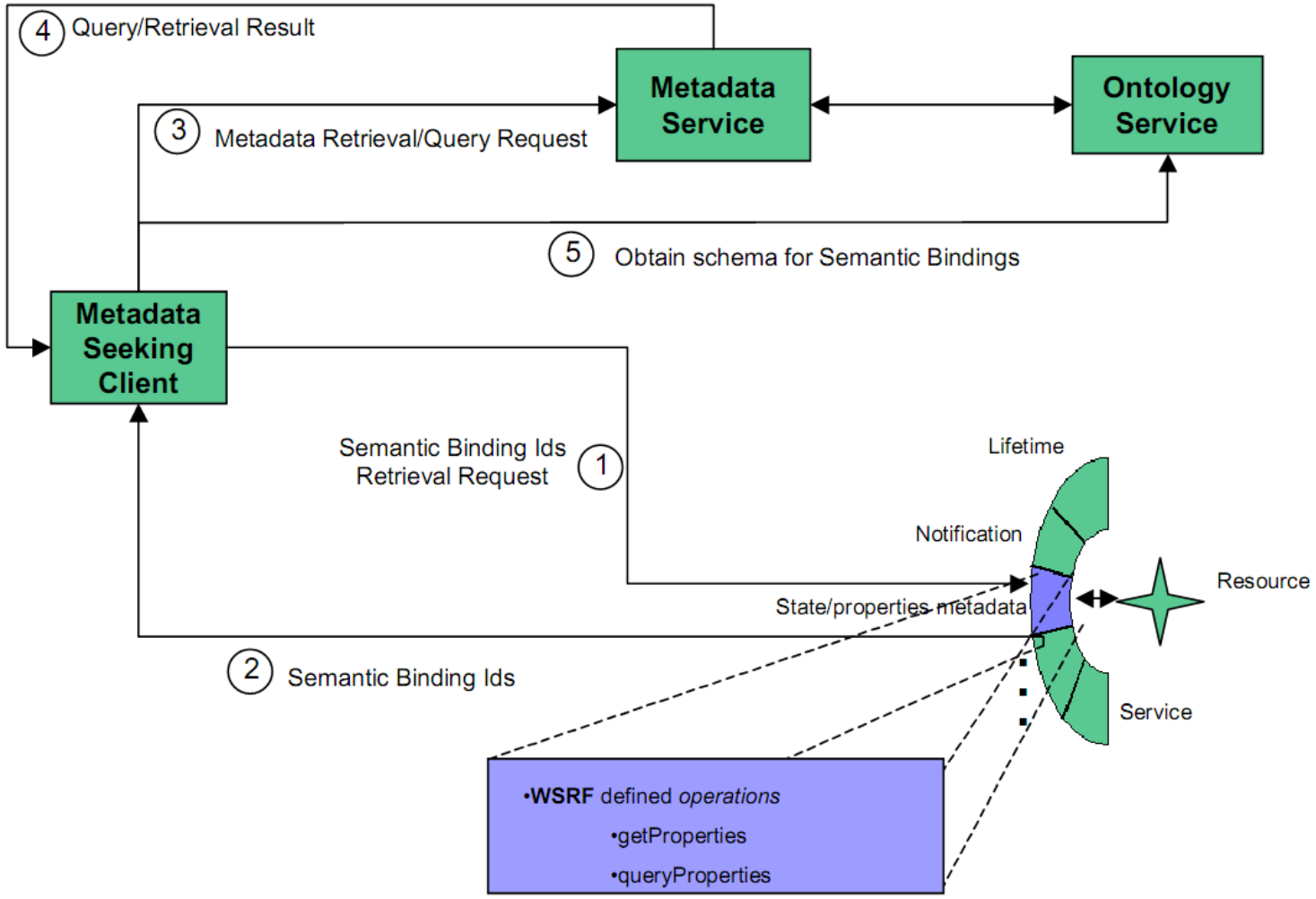
S-OGSA MECHANISMS



S-OGSA MECHANISMS

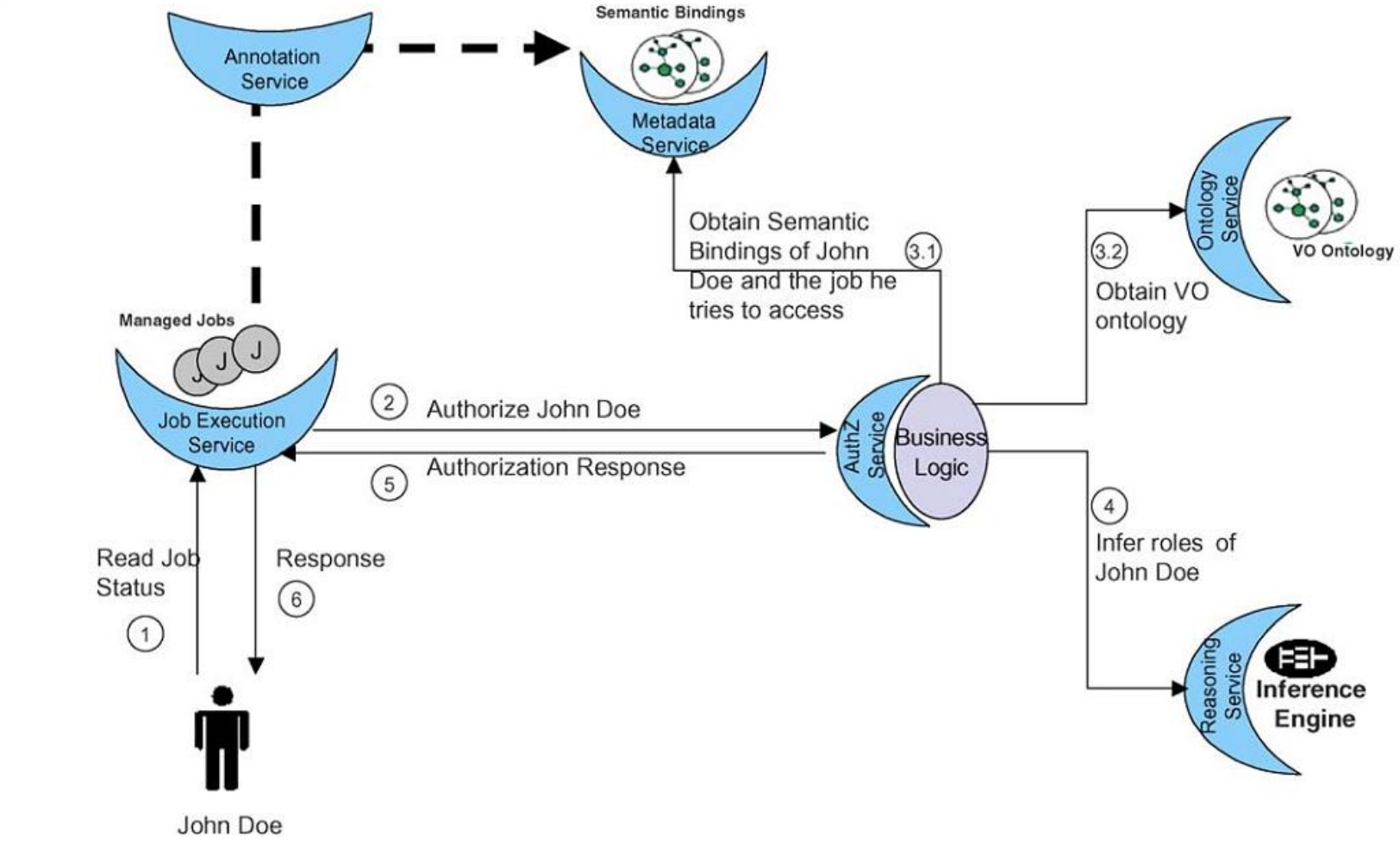
- S-Stateful Services: Delivery of Semantic Bindings by Grid Services.
  - Virtualise Grid resources
  - Set of mapping
  - Delivery mechanism rather than a descriptive framework

S-OGSA MECHANISMS



Retrieving and Querying Semantic Bindings of Resources

S-OGSA MECHANISMS



A semantic aware authorization service consuming an input that could in relation with S-OGSA Semantic Provisioning Services.

- What is it?
- How it is built?
- Towards the e-Laboratory

The social process of Science 2.0

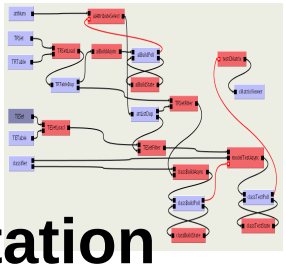
Virtual Learning Environment

Undergraduate Students

Digital Libraries

scientists

Graduate Students



Reprints

Peer-Reviewed Journal & Conference Papers

Preprints & Metadata

Technical Reports

experimentation

myexperiment

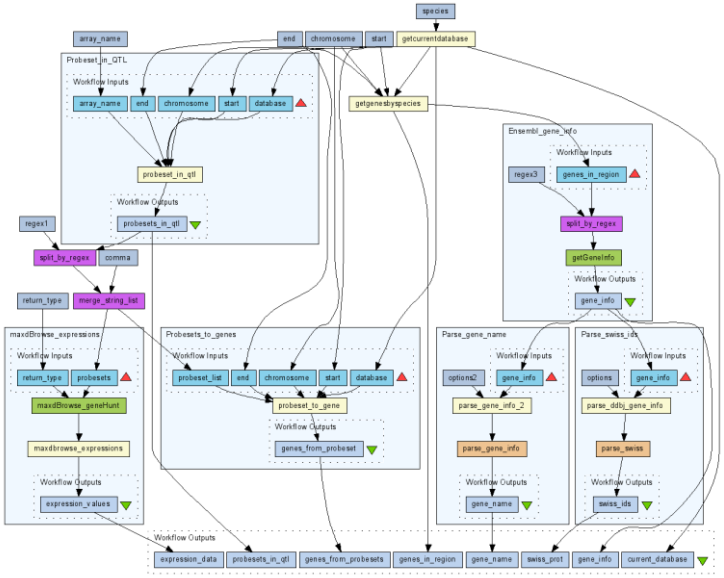
Repositories

Local Web

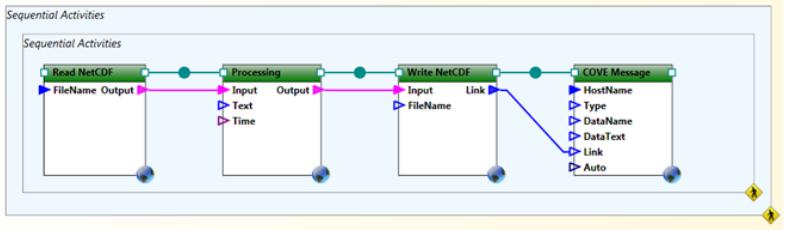
Certified Experimental Results & Analyses

Data, Metadata Provenance Workflows Ontologies

Sharing pieces of process



http://www.mygrid.org.uk/tools/taverna/



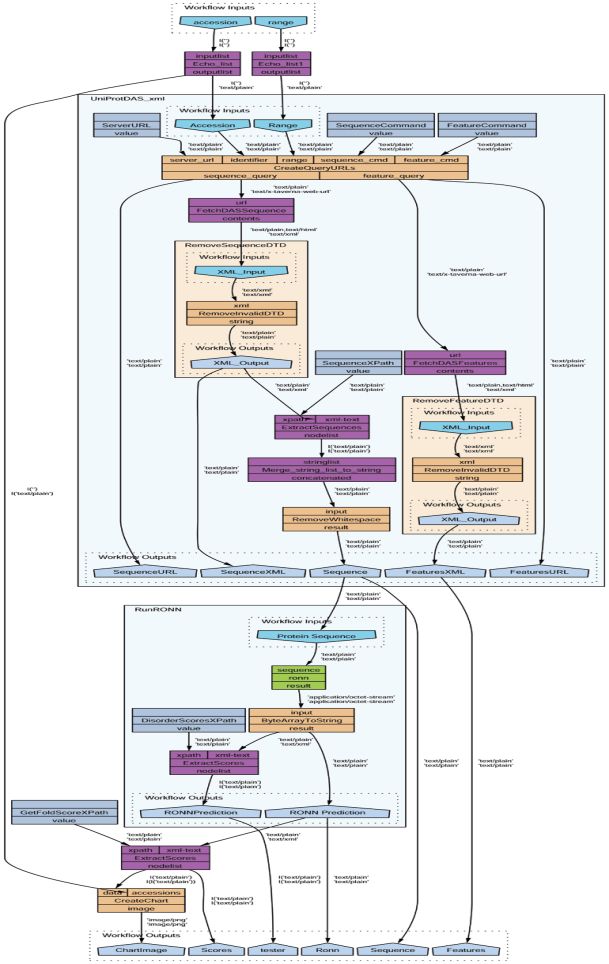
http://www.microsoft.com/mscorp/tc/trident.mspx

Monitor the formation of an aromatic imine by HMR and CMR in CDCl3

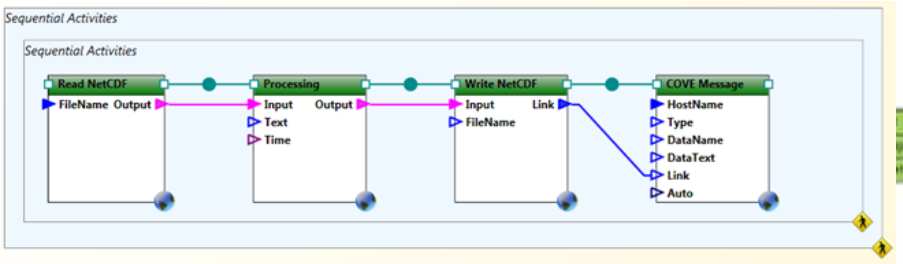
1. Make up separate 1 mL of 1M solutions of piperonal and 5-methylfurfurylamine in CDCl3.
2. Take HMRs and CMRs of the aldehyde and amine. Use 5 sec relaxation time and acquire for about 15 mins for the CMR. This should be good enough based on James' results at 1M in methanol.
3. Combine the two solutions into a 1 dram vial and shake vigorously then transfer to an NMR tube.
4. Take HMR at 5, 10 and 20 minutes after mixing.
5. Take CMR at 25 min after mixing.
6. Take HMR at 40 min after mixing.
7. Take CMR at 45 min after mixing.
8. Take HMR at 80 mins after mixing.
9. Take CMR at 85 mins after mixing.
10. Continue to take NMRs after interval doubling until no more change is observed.

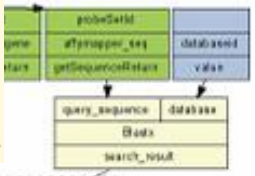
http://usefulchem.wikispaces.com/page/code/EXPLAN001

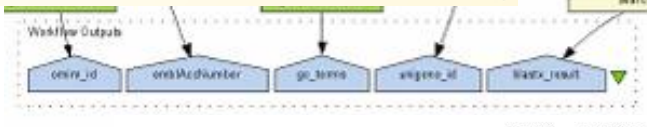
E. Science laboris

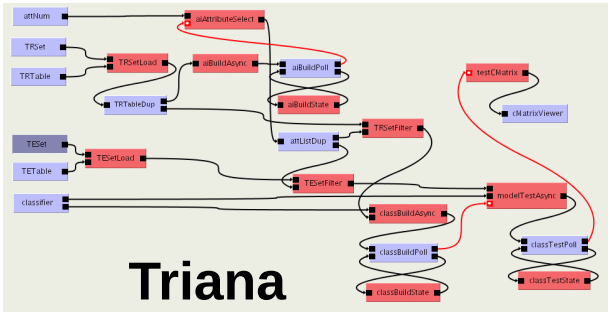


- Workflows are the new rock and roll
- Machinery for coordinating the execution of (scientific) services and linking together (scientific) resources
- The era of Service Oriented Applications
- Repetitive and mundane boring stuff made easier

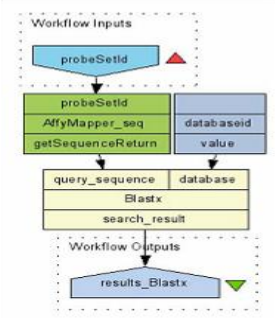


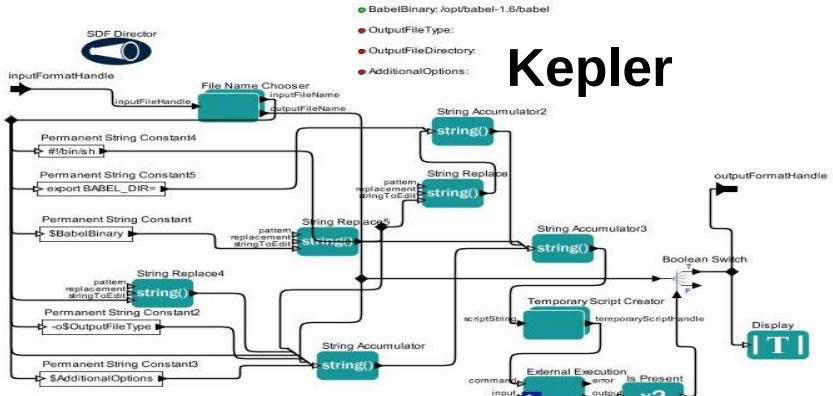


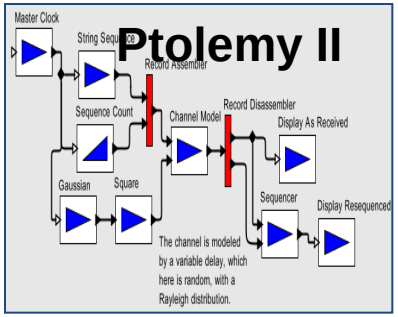




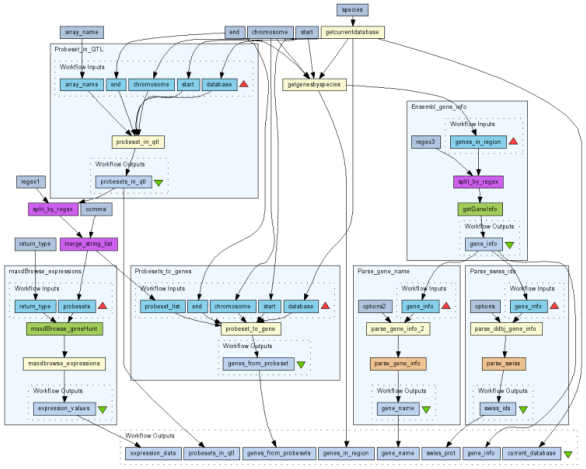
Trident

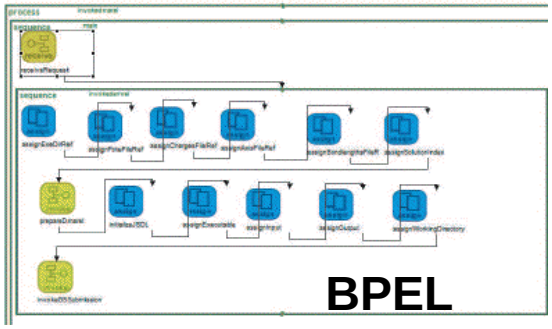


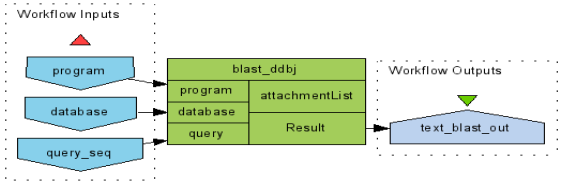


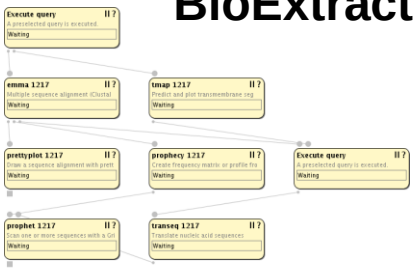


Taverna



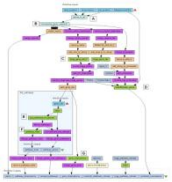






Reuse, Recycling, Repurposing

- Paul writes workflows for identifying biological pathways implicated in resistance to Trypanosomiasis in cattle
- Paul meets Jo. Jo is investigating Whipworm in mouse.
- Jo reuses one of Paul's workflow without change.
- Jo identifies the biological pathways involved in sex dependence in the mouse model, believed to be involved in the ability of mice to expel the parasite.
- Previously a manual two year study by Jo had failed to do this.



myexperiment

- "Facebook for Scientists" ...but different to Facebook!
- A repository of research methods
- A community social network
- A Virtual Research Environment

- Open source (BSD) Ruby on Rails application with HTML, REST and SPARQL interfaces
- Project started March 2007
- Closed beta since July 2007
- Open beta November 2007

myExperiment currently has 1712 registered users, 141 groups, 584 Taverna workflows plus 81 others, and 51 packs

Go to www.myexperiment.org to access publicly available content or create an account

Keep up to date

Get the latest news about what your online community is doing and what's happening with your Research Objects.

Form Friends & Groups

Explore and manage the social network. You have fine control over the privacy and sharing of your Research Objects.

Find Workflows

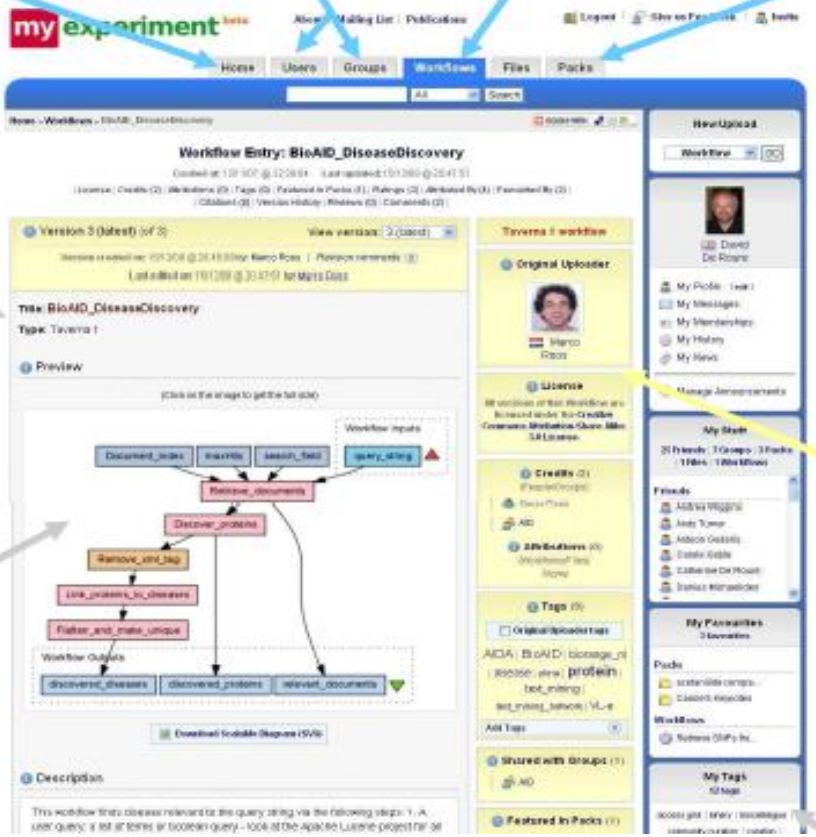
See the latest and most popular workflows: discover, view, download, run, tag and rate. Upload your workflows.

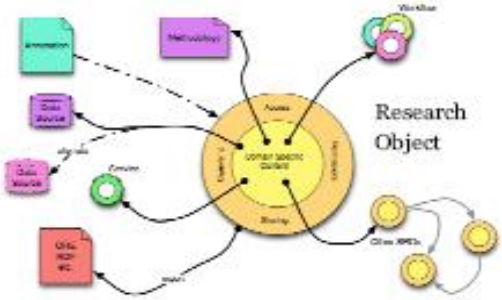
Build Packs

Share collections of items as individual packs — like all the digital items in an experiment. Include external items too.

Content types

We provide special support for workflow systems including Taverna and Trident, as well as experiment plans, providing a foundation for the e-Laboratory.





Curating process

Workflows capture pieces of research process which are curated by their authors, experts and the community. These curation models are also used in the Biocatalogue service registry.

BioCatalogue

All about the Research Object

See and manage all the essential extrinsic information and 'social metadata' — licence, tags, sharing, ratings.

Credits and attributions are an essential feature to support flow of rights and reputation.

All about me

Easy navigation using a dashboard of all the things relating to me and my social network.

JISC EPSRC

Microsoft

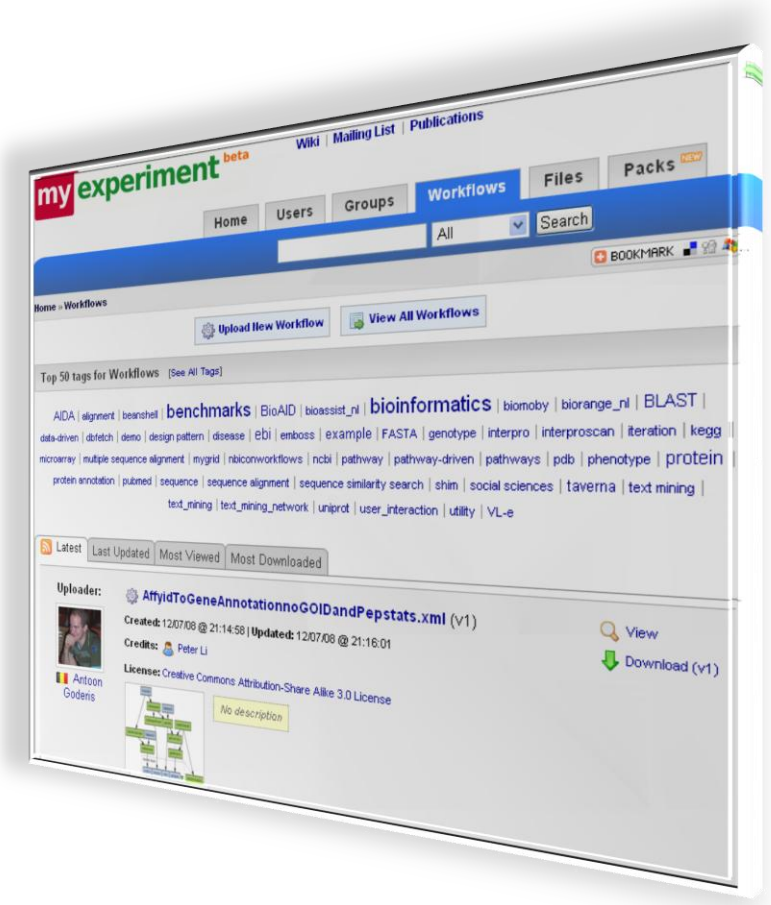
myexperiment

myExperiment Features

- User Profiles
- Groups
- Friends
- Sharing
- Tags
- Workflows
- Developer interface

Distinctives

- Credits and Attributions
- Fine control over privacy
- Packs
- Federation
- Enactment



Version 1 (of 1)

Version created on: 08/05/08 @ 15:25:29 **by:** Saeedeh | Revision comments

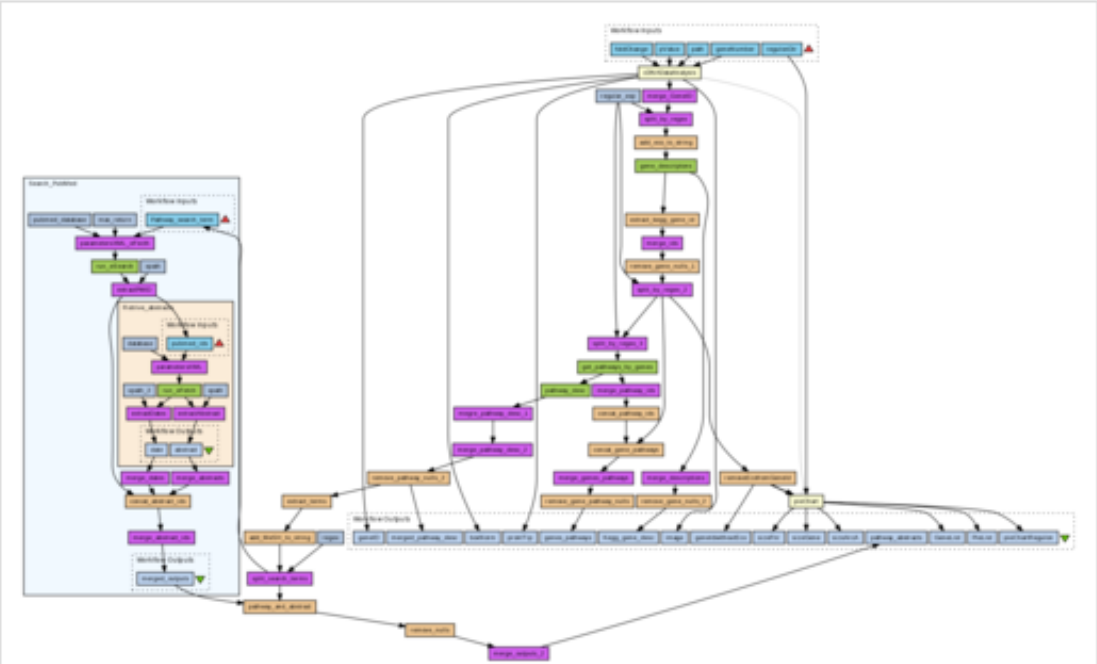
Last edited on: 12/05/08 @ 09:01:37 **by:** Saeedeh

Title: **Escherichia coli : From cDNA Microarray Raw Data to Pathways and Published Abstracts**

Type: Taverna 1

Preview

(Click on the image to get the full size)



Download Scalable Diagram (SVG)

Taverna 1 workflow

Original Uploader

Saeedeh

License

Credits (2)

(People/Groups)

- Saeedeh
- Paul Fisher

Attributions (1)

(Workflows/Files)

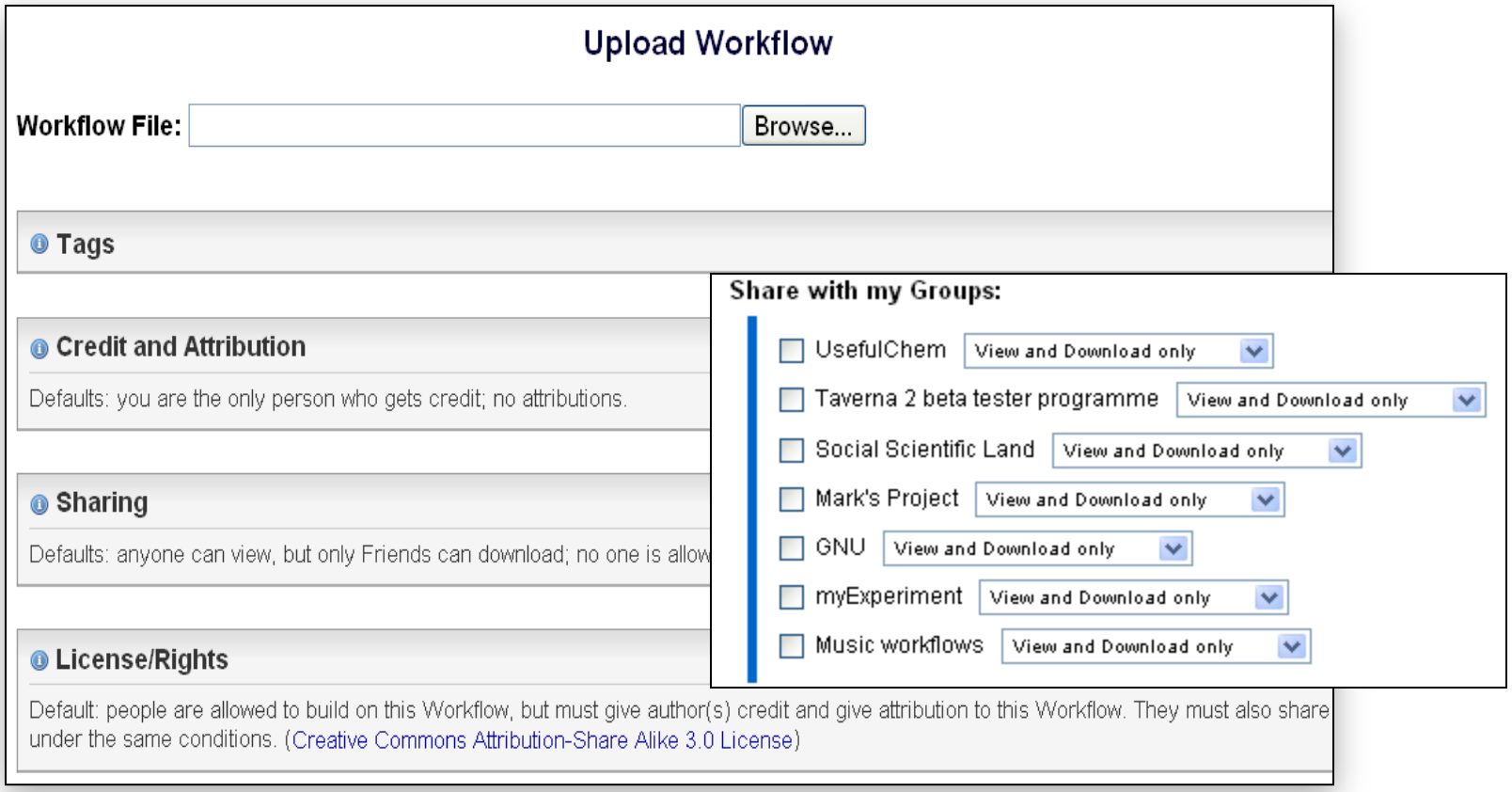
- HUMAN Microarray CEL file to candidate pathways

Tags (5)

Original Uploader tags

e.coli | kegg | Kegg Pathways | pathways | pubmed

Control over sharing



The most important aspect of myExperiment

Designed by scientists

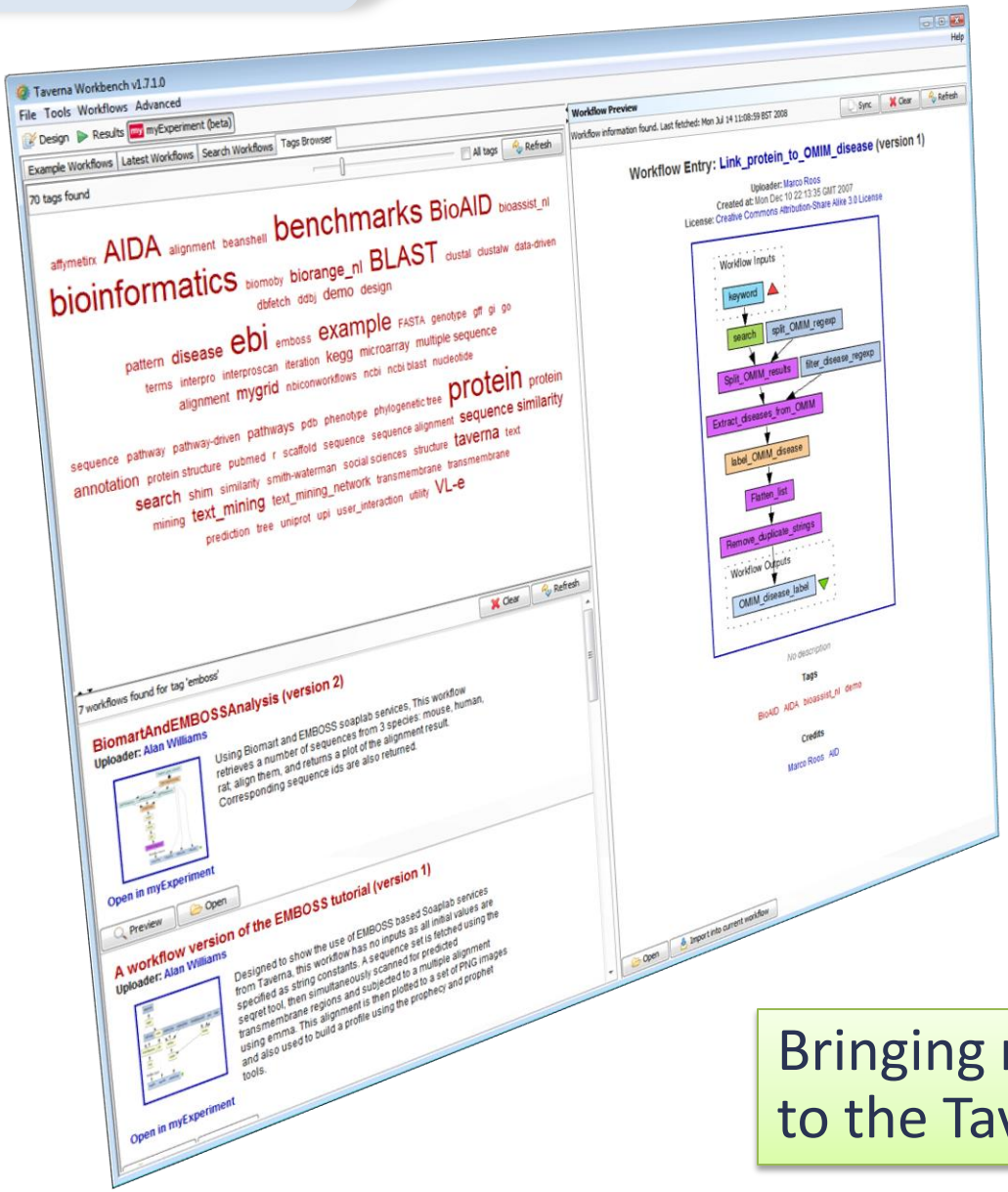
For Developers

- All the myExperiment services are accessible through simple RESTful programming interfaces
  - use your existing environment and augment it with myExperiment functionality
  - build entirely new interfaces and functionality mashups
- The Ruby on Rails codebase is open source (BSD) so you can run your own myExperiment – perhaps for your own lab or to develop new funcionality
- Go to wiki.myexperiment.org for information about our Developer Community

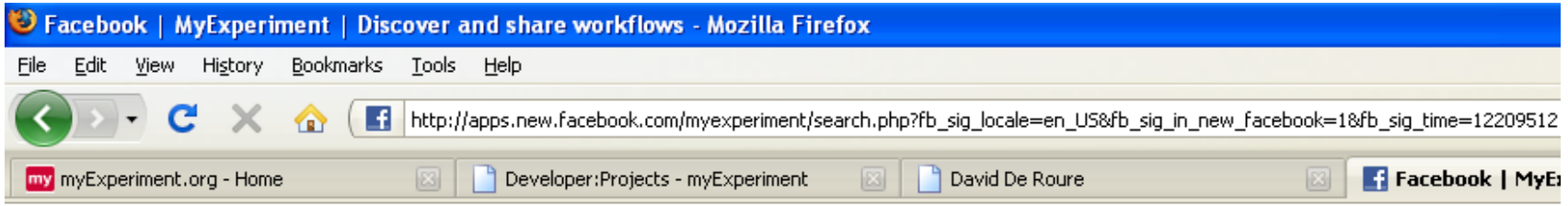
Google Gadgets

Bringing myExperiment to the iGoogle user

Taverna Plugin



Bringing myExperiment to the Taverna user



my experiment beta

Home | Search | Upload workflows | My Workflows | My Friends 's Workflows | Invite | Feedback User Guide

MyExperiment

marco roos All Search

Advanced Search

Search result for keyword: marco roos

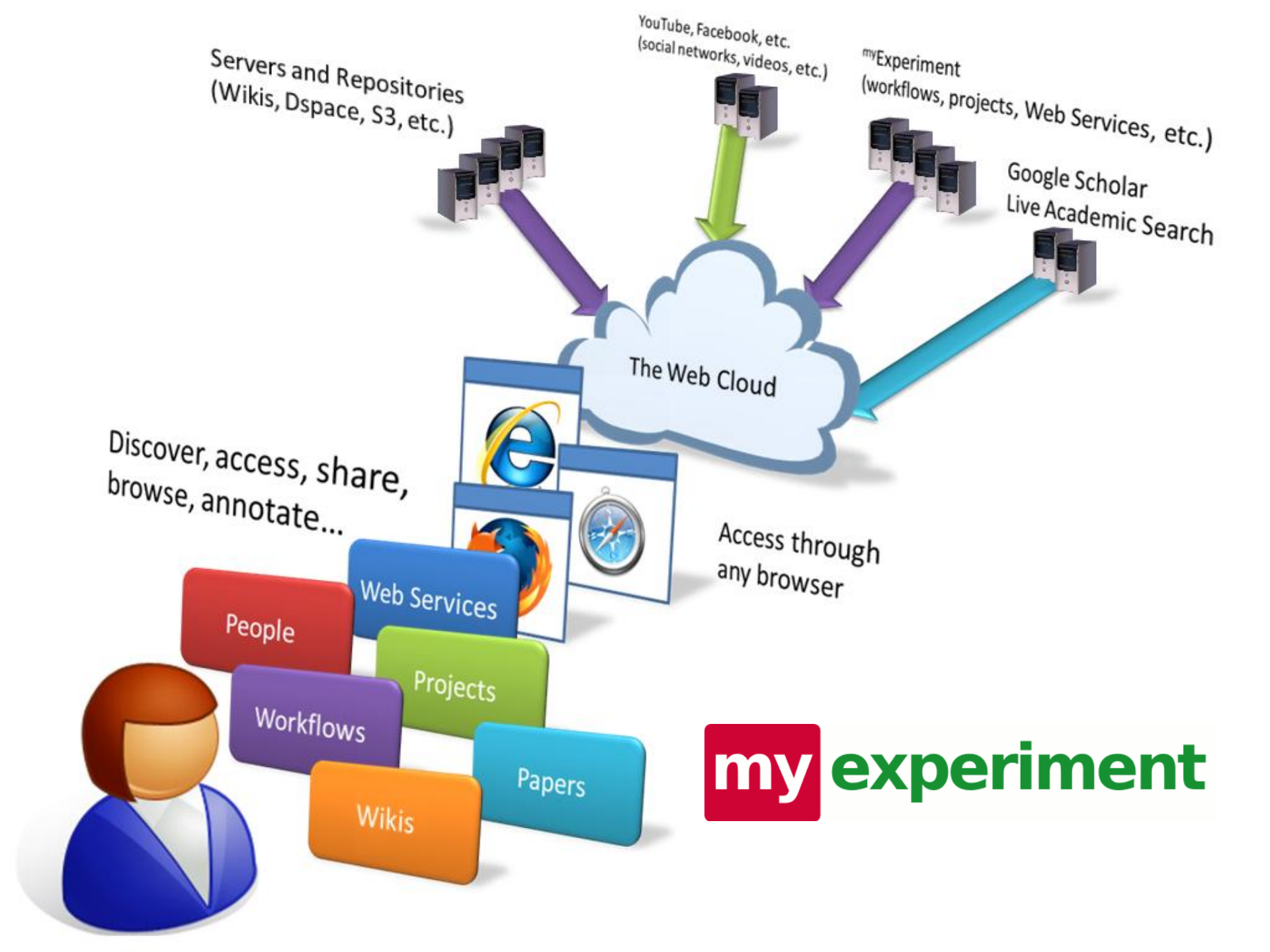
workflows(28) users(1) groups(10)

Page 1

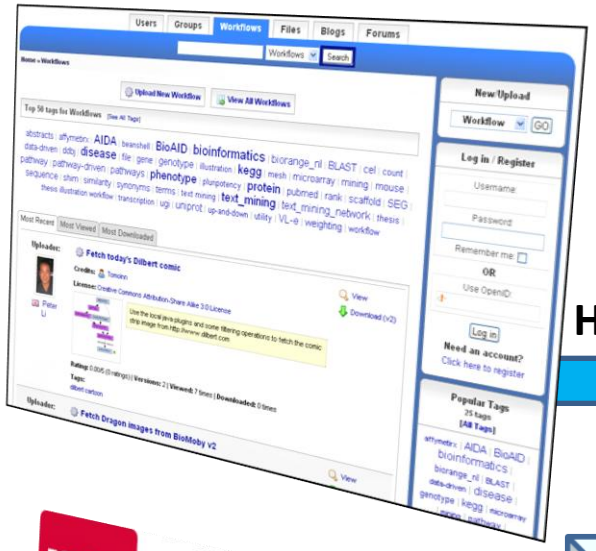
Name: Marco Roos
Joined: 21/Jul/2007 @ 02:43:23
Email: roos@science.uva.nl
Website: http://home.medewerker.uva.nl/m.roos1

Facebook

- What is it?
- How it is built?
- Towards the e-Laboratory


Servers and Repositories
(Wikis, Dspace, S3, etc.)
YouTube, Facebook, etc.
(social networks, videos, etc.)
myExperiment
(workflows, projects, Web Services, etc.)
Google Scholar
Live Academic Search
The Web Cloud
Discover, access, share,
browse, annotate...
Access through
any browser
People
Web Services
Projects
Workflows
Papers
Wikis
my experiment

For Developers



myexperiment

XML

facebook

iGoogle

android

HTML

API config

ORE
FOAF
SIOC

Managed REST API

Search API

Search Engine

SPARQL endpoint

tags ratings reviews profiles
workflows credits groups
files packs friendships

RDF Store

EPrints
DSpace
Fedora
S3
SRB

mySQL

Enactor API

Enactor

SPARQL endpoint

myExperiment SPARQL Endpoint

Querying

Query Type: SELECT

Results Format: SPARQL Results

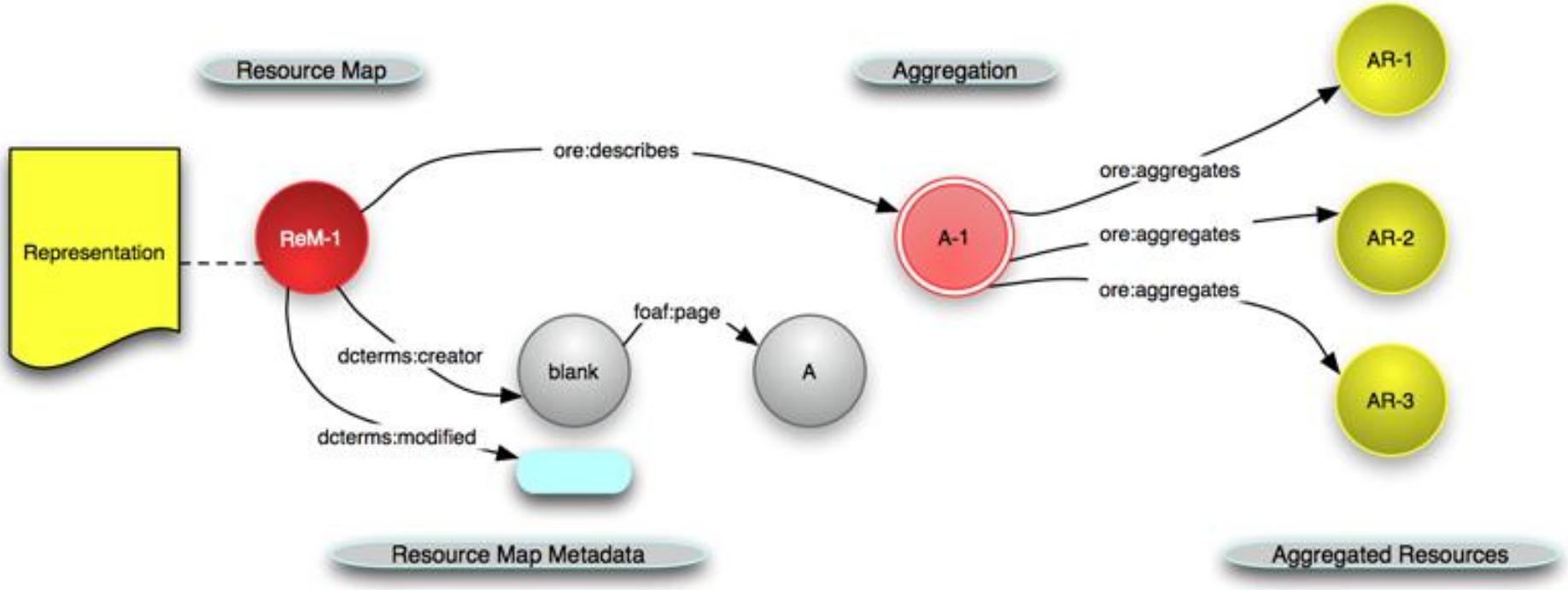
Display Format: In page

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX myexp: <http://rdf.myexperiment.org/ontology#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
select ?friend1 ?friend2 ?acceptedat where {?z rdf:type
<http://rdf.myexperiment.org/ontology#Friendship> . ?z myexp:has-requester
?x .
?x sioc:name ?friend1 . ?z myexp:has-accepter ?y . ?y sioc:name ?friend2 .
?z myexp:accepted-at ?acceptedat }
```

All accepted Friendships including accepted-at time

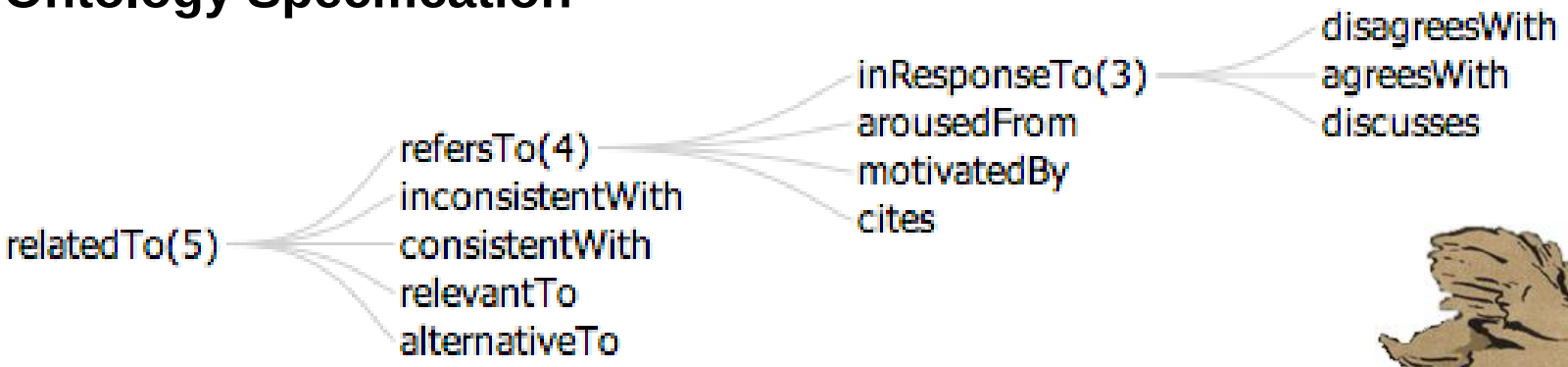
Semantically-Interlinked Online Communities

Exporting packs



Open Archives Initiative
Object Reuse and Exchange

Scientific Discourse Relationships Ontology Specification

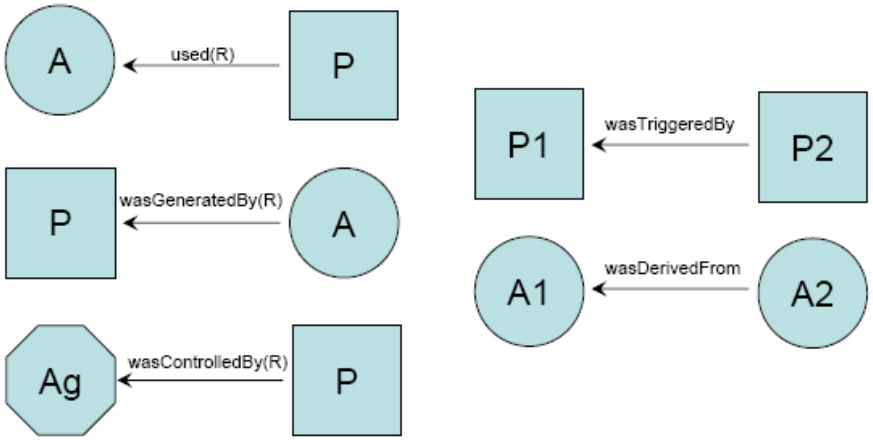


THE PROVENANCE OF ELECTRONIC DATA

It would include details of the processes that produced electronic data as far back as the beginning of time or at least the epoch of provenance awareness.

Provenance is well understood in the study of fine art where it refers to the documented history of some art object. Given that documented history, the object attains an authority that allows scholars to understand and appreciate its importance and context relative to other works. Art objects that lack a proven history may be viewed with skepticism by those who study them.

If the provenance of data produced by computer systems could be determined, then users would be able to understand how documents had been assembled, how simulation results were determined, and how financial analyses were carried out. Computer applications should thus



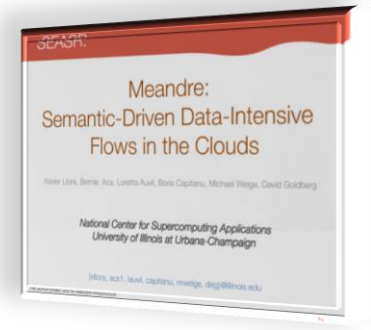
Open Provenance Model

Communications of the ACM 51, 4 (Apr. 2008), 52-58

Phase 2

Phase 2

- Repository integration (institutional: EPrints, Fedora)
- Controlled vocabularies
- Relationships between items (in and between packs)
- Recommendations
- Improved search ranking and faceted browsing
- Indexing of packs
- New contribution types (Meandre, Kepler, e-books)
- Further blog / wiki integration
- Biocatalogue integration



Reuse and Symbiosis

Content Capture and Curation

Self by Service Providers

refine
validate

seed

Experts

refine
validate

seed

my experiment

BioCatalogue

refine
validate

seed

seed

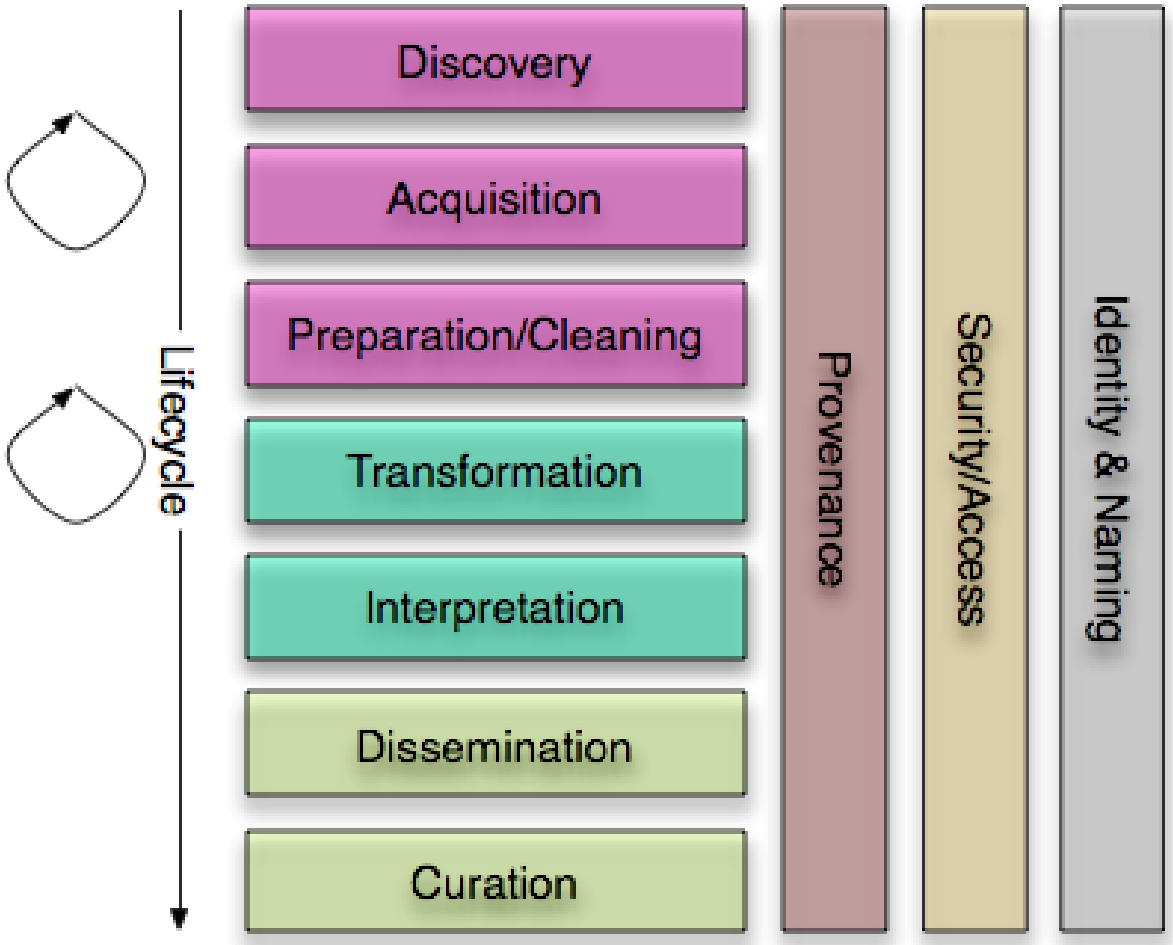
refine
validate

Social by User Community

Automated

- What is it?
- How it is built?
- Towards the e-Laboratory

e-Laboratory Lifecycle



Local projects using Taverna and/or myExperiment

SysMO

Ondex

Obesity eLab

Shared Genomics

NEMA

CombeChem

LifeGuide

IBBRE

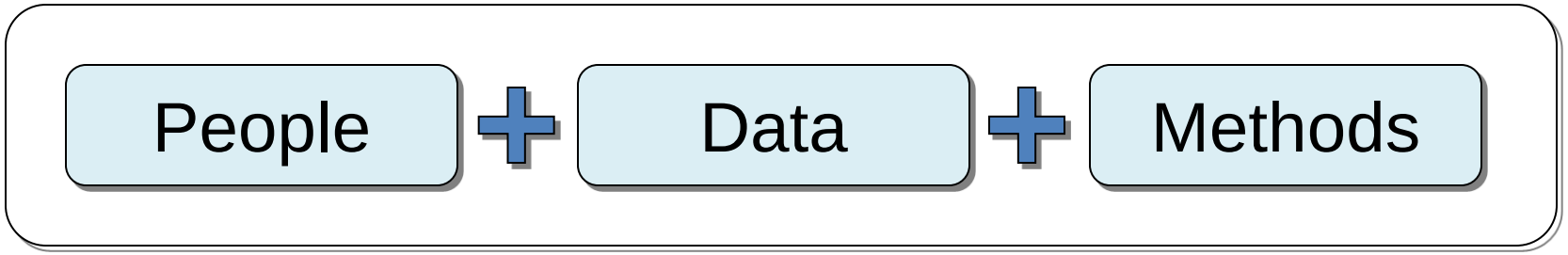
What is an e-Laboratory?

- A **laboratory** is a facility that provides *controlled conditions* in which scientific research, experiments and measurements may be performed, offering a work space for researchers.
- An **e-Laboratory** is a set of integrated components that, used together, form a distributed and collaborative space for e-Science, enabling the planning and execution of *in silico* experiments -- processes that combine data with computational activities to yield experimental results

e-Labs

An e-Lab consists of:

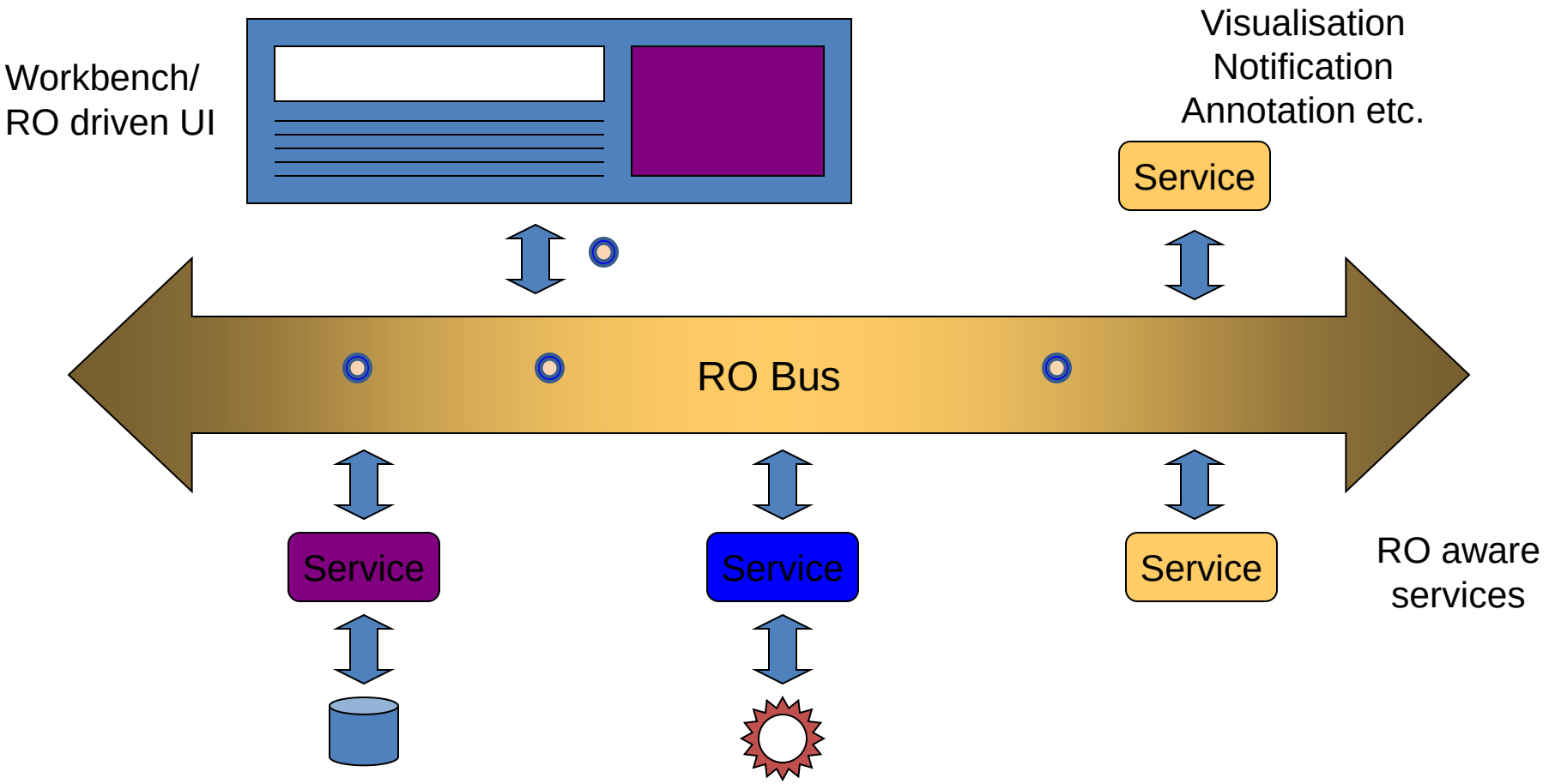
1. a community
2. work objects
3. generic resources for building and transforming work objects



Sharing infrastructure *and* content across projects

e-Labs + Research Objects

- An e-Lab is built from a collection of **services**, consuming and producing **Research Objects**

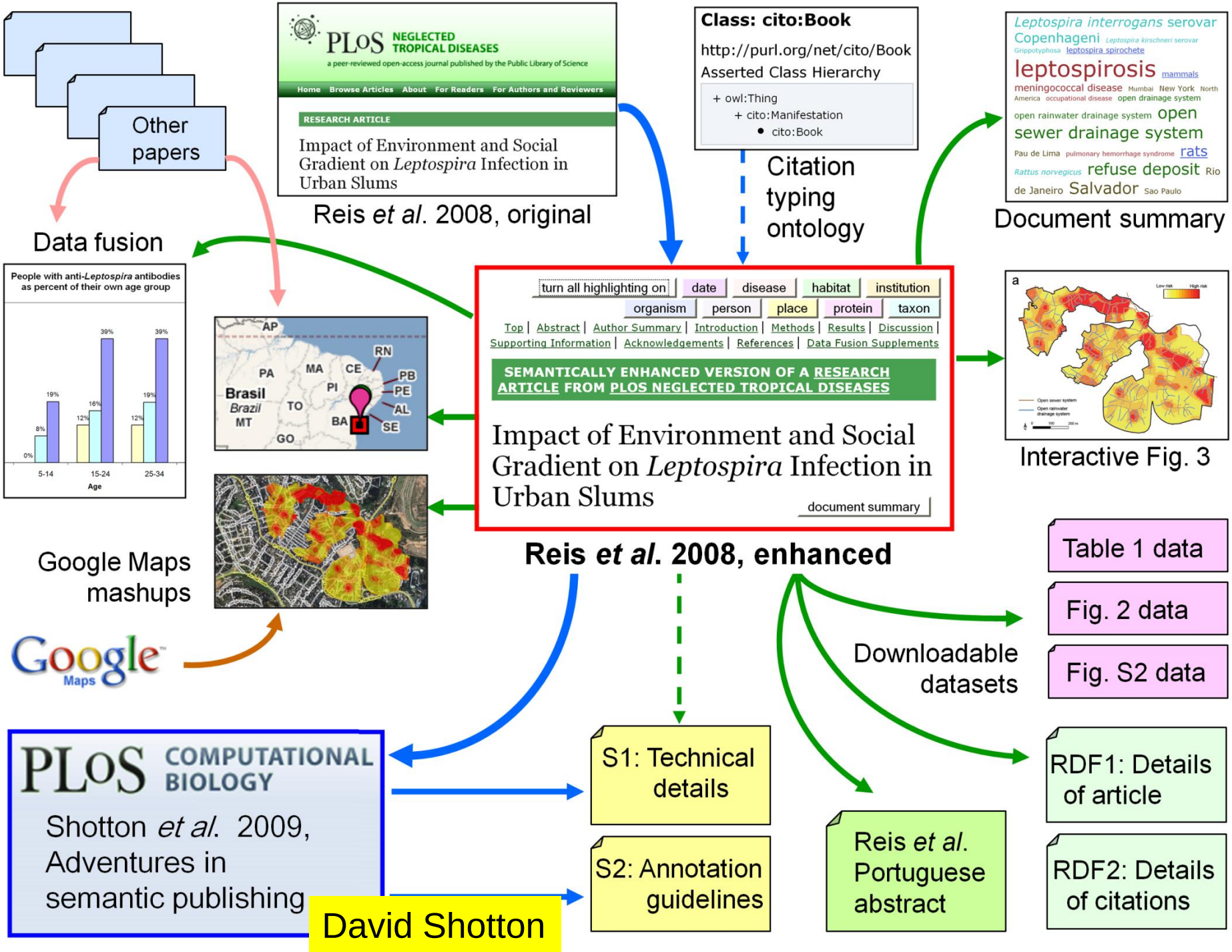


Assembling e-Laboratories

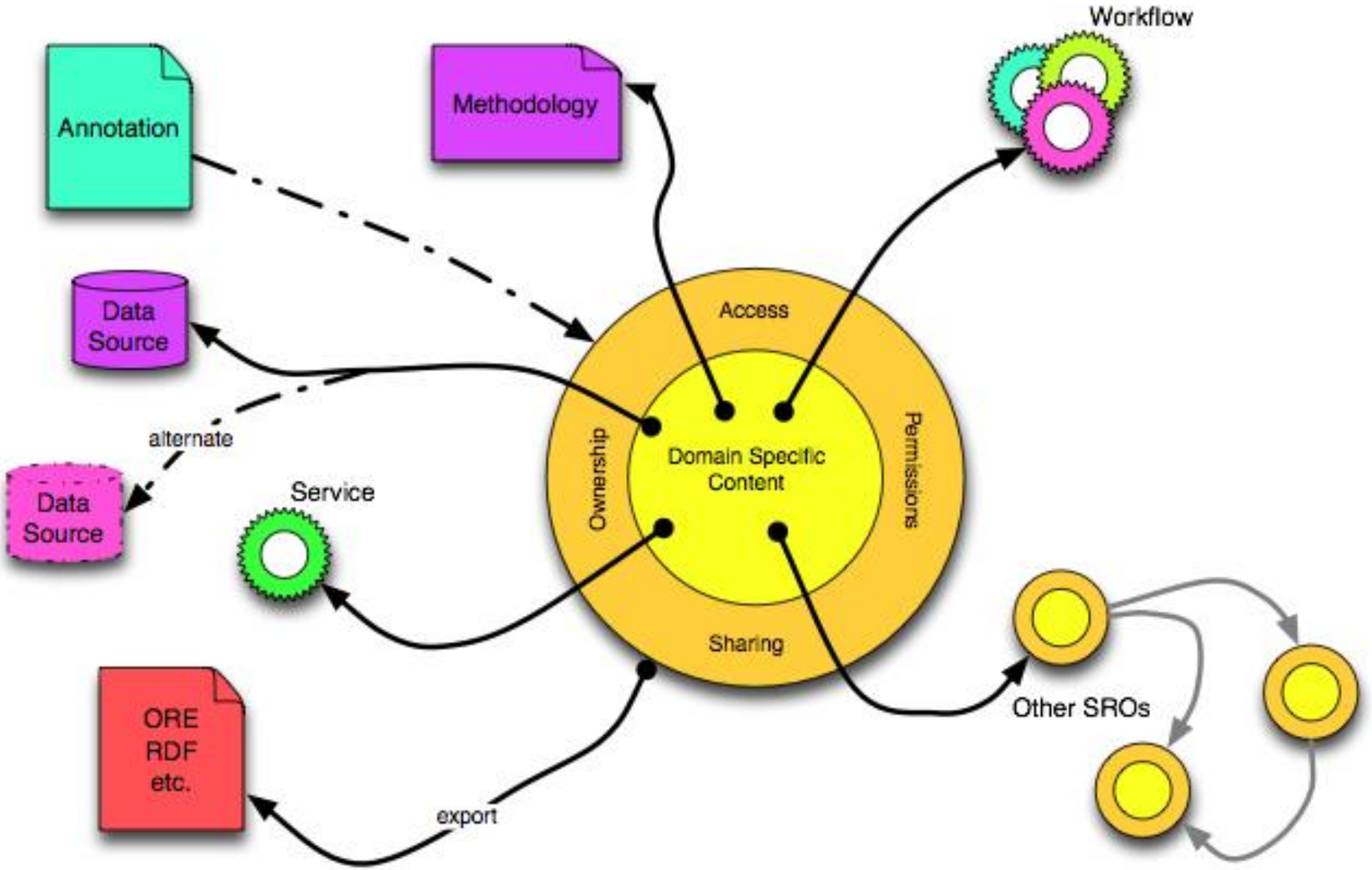
- An e-Lab is a set of components and resources
  - An open system, not a software monolith
  - Utility of components transcends their immediate application
- We envisage an ecosystem of cooperating e-Laboratories
- What are the e-Lab components and services?
- What are the Research Objects?

Example Core Services

Workflow Monitoring
Event Logging
Social Metadata
Annotation Service
Search, ranking
User Registration
Distributed Data Query
Job Execution
Naming and Identity
Anonimisation
Text Mining
Research Object Management
Probity
Coreference Resolution


Other papers
Data fusion
PLoS NEGLECTED TROPICAL DISEASES
a peer-reviewed open-access journal published by the Public Library of Science
Home Browse Articles About For Readers For Authors and Reviewers
RESEARCH ARTICLE
Impact of Environment and Social Gradient on Leptospira Infection in Urban Slums
Reis et al. 2008, original
Class: cito:Book
http://purl.org/net/cito/Book
Asserted Class Hierarchy
+ owl:Thing
+ cito:Manifestation
• cito:Book
Citation typing ontology
Leptospira interrogans serovar Copenhageni
leptospirosis
meningococcal disease
open sewer drainage system
open rainwater drainage system
rats
refuse deposit
Salvador
Document summary
People with anti-Leptospira antibodies as percent of their own age group
5-14
15-24
25-34
Age
Brasil
Brazil
turn all highlighting on
date
disease
habitat
institution
organism
person
place
protein
taxon
Top | Abstract | Author Summary | Introduction | Methods | Results | Discussion | Supporting Information | Acknowledgements | References | Data Fusion Supplements
SEMANTICALLY ENHANCED VERSION OF A RESEARCH ARTICLE FROM PLOS NEGLECTED TROPICAL DISEASES
Impact of Environment and Social Gradient on Leptospira Infection in Urban Slums
document summary
Interactive Fig. 3
Google Maps mashups
Google Maps
Reis et al. 2008, enhanced
Table 1 data
Fig. 2 data
Fig. S2 data
Downloadable datasets
PLoS COMPUTATIONAL BIOLOGY
Shotton et al. 2009, Adventures in semantic publishing
David Shotton
S1: Technical details
S2: Annotation guidelines
Reis et al. Portuguese abstract
RDF1: Details of article
RDF2: Details of citations

Anatomy of a Research Object



Characteristics of a Research Object

1. **Composite**. Contain typed interrelationships and dependencies between resources but are in turn labelled and identifiable as an individual resource.
2. **Distributed**. Structured collections of references to locally managed and externally located resources. Implications for reliability, consistency, mixed stewardship, versioning and identity resolution.
3. **Annotated**. Carry metadata concerning provenance profile, lifecycle profile, sharing profile (permissions, licensing, downloads, views), curation profile (tags, comments, ratings) and usage profile.
4. **Repeatable**. Capture information about the lifecycle of the investigation facilitating experiments to be *repeatable (without change), reusable (with reconfiguration), replayable and/or repurposable (as new components or templates).*
5. **Interoperable**. Publishable and exchangeable units that facilitate interoperability; OAI-ORE standards increase interoperability and facilitate the consumption of Research Objects in between applications.

Thoughts

- myExperiment provides social infrastructure – it facilitates sharing and enables scientists to “collaborate in order to compete”
- myExperiment has growing community and growing content
- New content types: Meandre, Kepler, R, matlab, spreadsheets, ... SPARQL queries?
- We are targetting how we believe research will be conducted in the future, through the assembly of e-Laboratories which share Research Objects
- SPARQL endpoint is an effective alternative to the API – provides any service you want!
- Workflows for Semantic Web scripting?

References

- David De Roure, Nicholas R. Jennings and Nigel R. Shadbolt: The Semantic Grid: Past, Present and Future
- David De Roure, Nicholas R. Jennings and Nigel R. Shadbolt: The Semantic Grid: A Future e-Science Infrastructure
- Fran Berman, Anthony J. G. Hey, Geoffrey C. Fox: Grid Computing Making the Global Infrastructure a Reality
- Oscar Corcho∗, Pinar Alper, Ioannis Kotsiopoulos, Paolo Missier, Sean Bechhofer, Carole Goble: An overview of S-OGSA: A Reference Semantic Grid Architecture, 2006
- http://www.semanticgrid.org/vision.html
- MIT - HCLS IG F2F - myExperiment

THANK YOU!