Repositories and Linked Open Data: the view from myExperiment

David De Roure

Overview

- Motivation: the primacy of method
- myExperiment and Other Animals
- Design and implementation
- The future of research

The social process of Science 1.0 2.0

Virtual Learning Environment

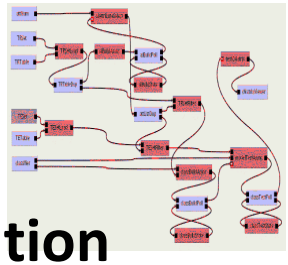
Undergraduate Students

Next Generation Researchers

Digital Libraries

scientists

Graduate Students



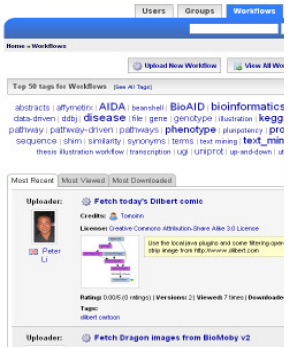
Reprints

Peer-Reviewed Journal & Conference Papers

Preprints & Metadata

Technical Reports

experimentation



Repositories

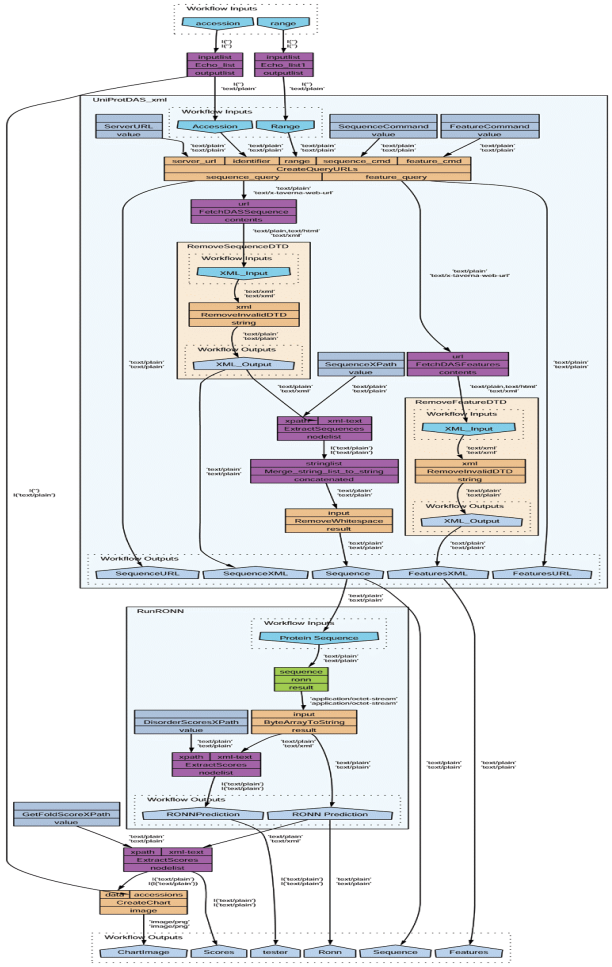
Local Web

Certified Experimental Results & Analyses

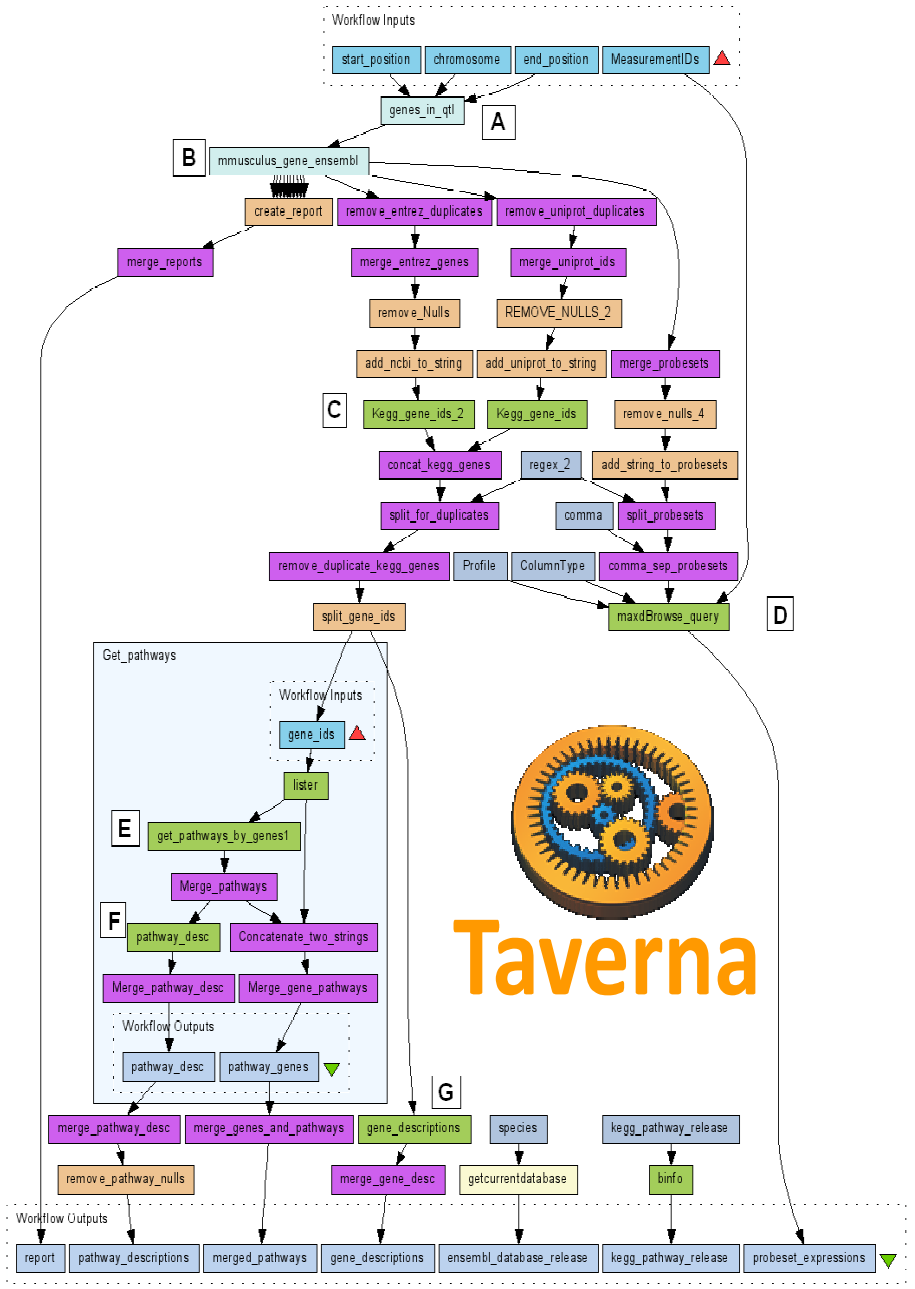
Data, Metadata, Provenance, Scripts, Workflows, Services, Ontologies, Blogs, ...

myexperiment

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

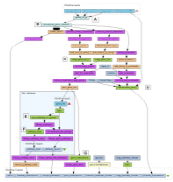


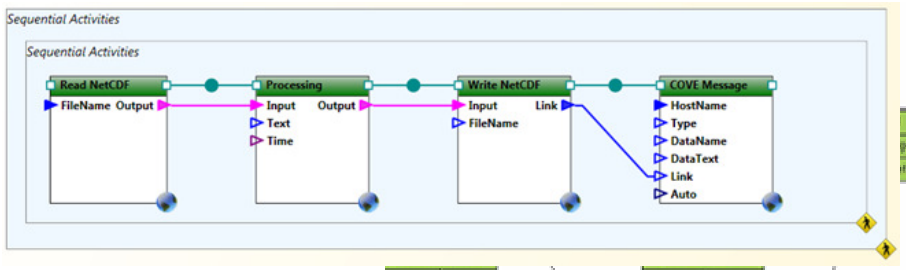
Taverna Workflows

- Access to distributed and local resources
- Automation of data flow
- Iteration over data sets
- Interactive
- Agile software development
- Experimental protocols
- Declarative mashups
- But...
  - Can be hard to build
  - Can “decay” as services change

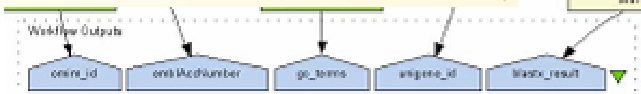
Reuse, Recycling, Repurposing

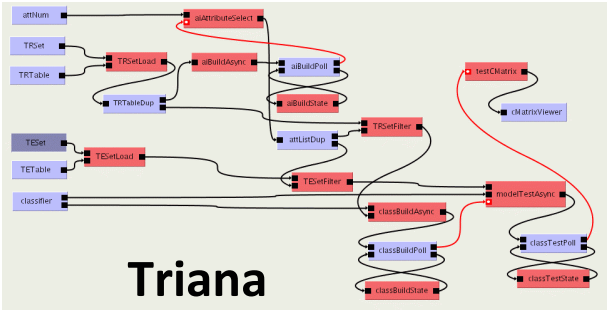
- Paul writes workflows for identifying biological pathways implicated in resistance to Trypanosomiasis in cattle
- Paul meets Jo. Jo is investigating Whipworm in mouse.
- Jo reuses one of Paul's workflow without change.
- Jo identifies the biological pathways involved in sex dependence in the mouse 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.

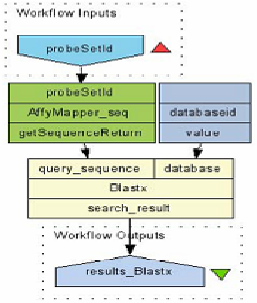


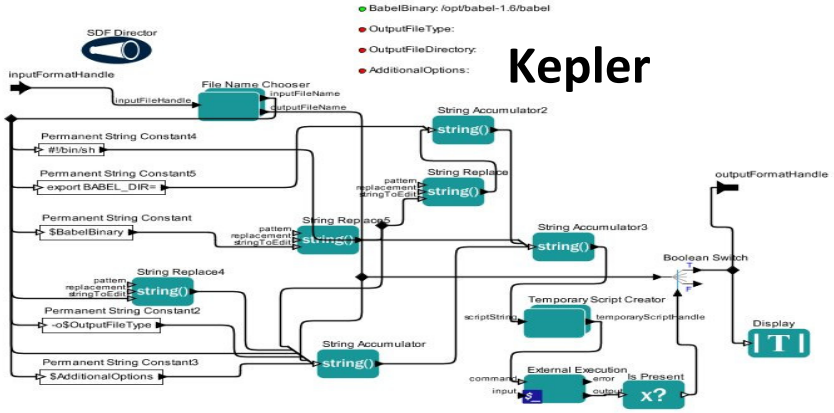


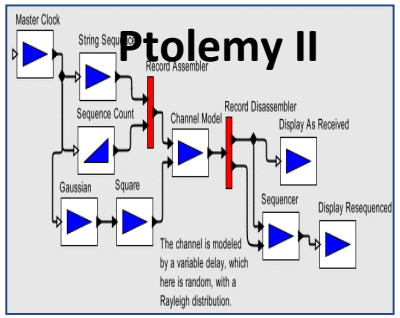
Trident



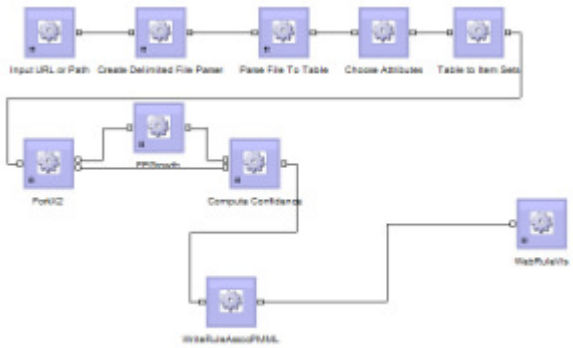


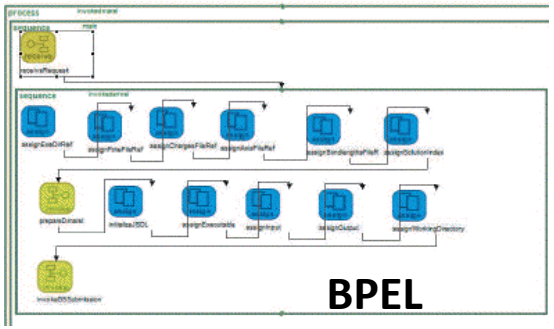


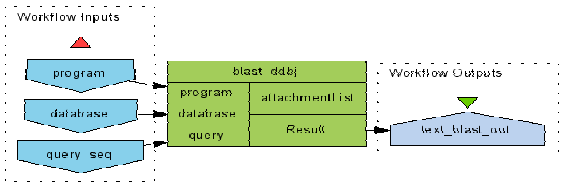




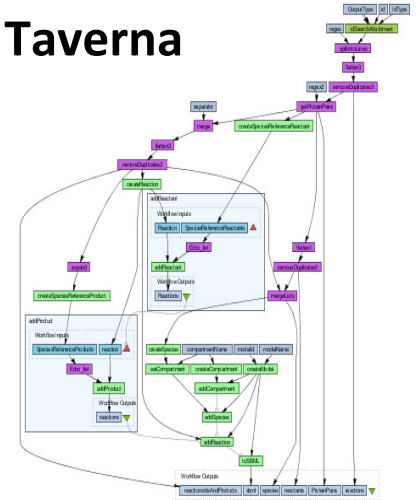
Meandre



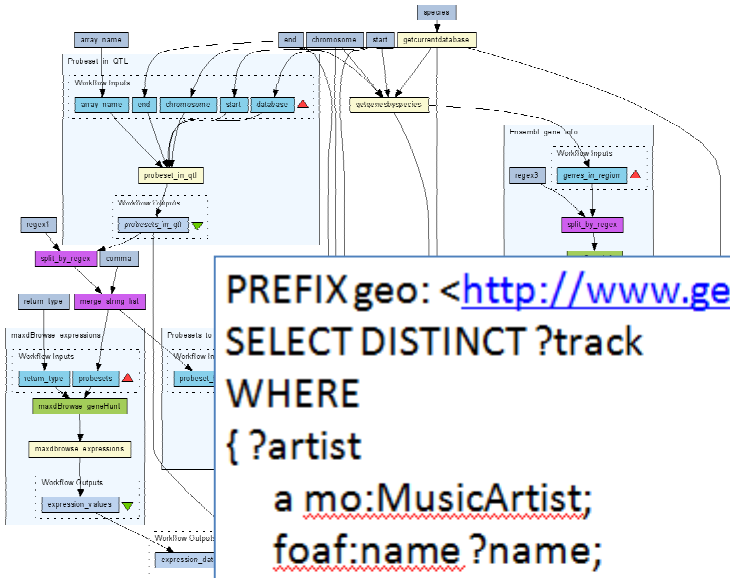




Taverna



Sharing pieces of process

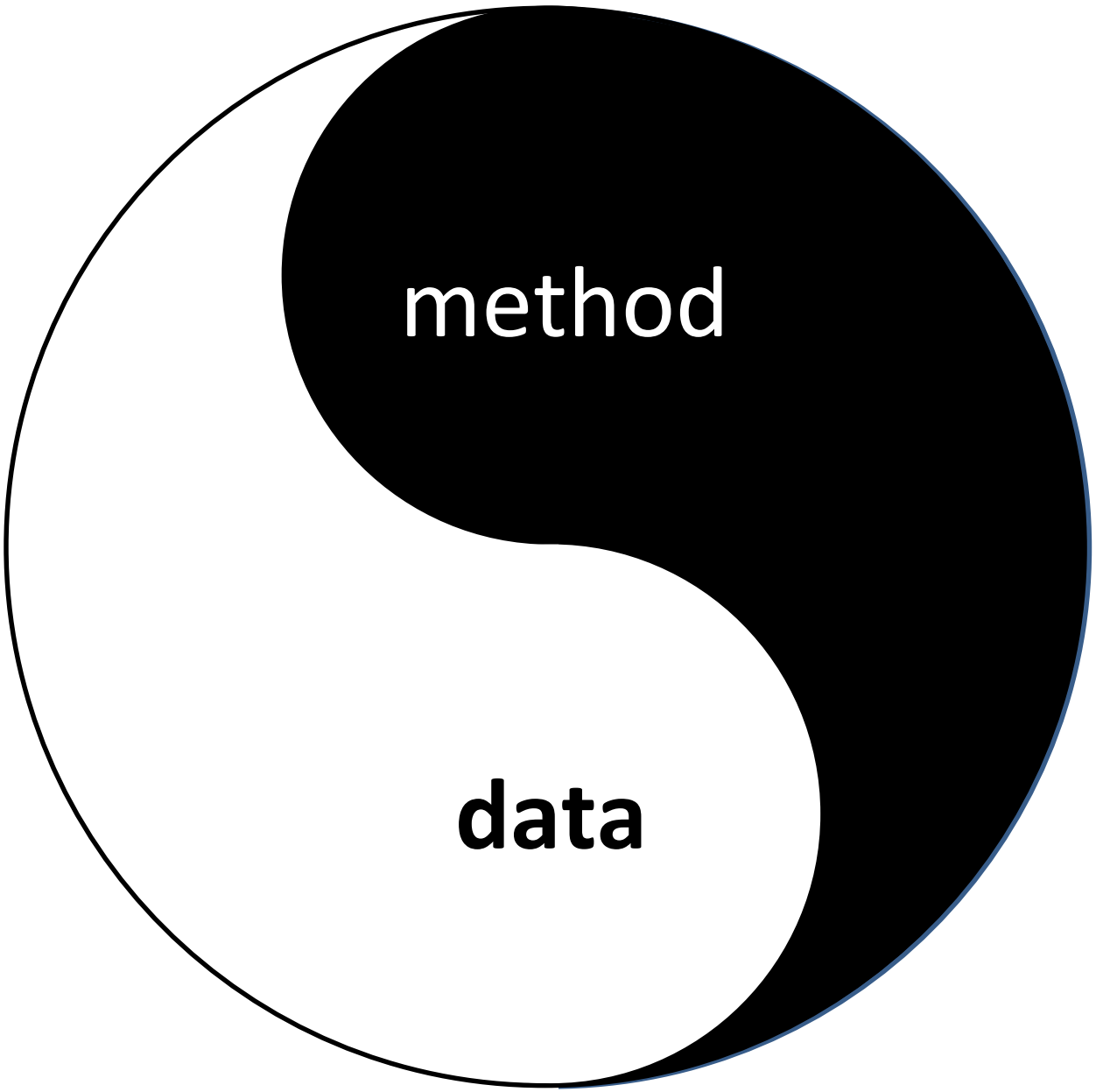


```
PREFIX geo: <http://www.geonames.org/ontology#>
SELECT DISTINCT ?track
WHERE
{ ?artist
    a mo:MusicArtist;
    foaf:name ?name;
    foaf:based_near ?place.
  ?place
    geo:inCountry <http://www.geonames.org/countries/#FR>.
  ?record
    a mo:Record;
    foaf:maker ?artist;
    mo:track ?track }
ORDER BY ?track
```

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.

nd acquire for about 15
ould be good enough
1M in methanol.
s into a 1 dram vial and
sfer to an NMR tube.
minutes after mixing.
mixing.
mixing.
mixing.
r mixing.
r mixing.
ter interval doubling
served.


method
data

Carole Goble “e-Science is me-Science: What do Scientists want?”, EGEE 2006

Home | Tech | Science in Society | News

MySpace for the dudes in lab coats

19 October 2006

Magazine issue 2574 Subscribe and get 4 free issues

Their kids may own version of My deas and research tool

Dubbed MyE ls such as social ne s. "There are these gre (-to- front," says R nester, UK, and a me

As well as sha ace for swapping and modifying the software tools that bioinformaticians use to identify and characterise genes.

“There are these great collaboration tools that 12-year-olds are using. It’s all back to front.”

Robert Stevens

“A biologist would rather share their toothbrush than their gene name”

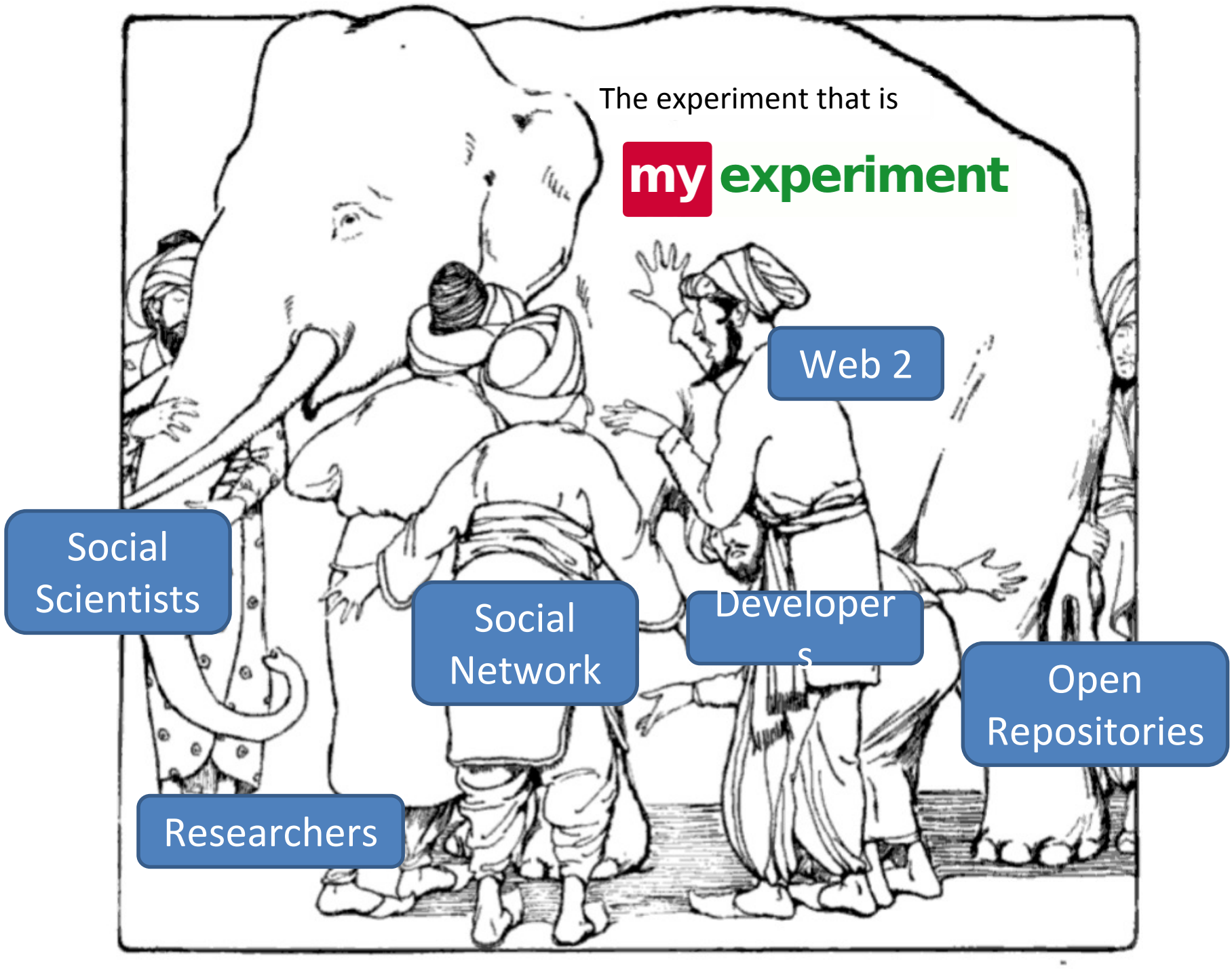
Mike Ashburner and others
Professor in Dept of Genetics,
University of Cambridge, UK

“Data mining: my data’s mine and your data’s mine”

Overview

- Motivation: the primacy of method
- myExperiment and Other Animals
- Design and implementation
- The future of research

Not Facebook for scientists!


The experiment that is
myexperiment
Web 2
Social Scientists
Social Network
Developers
Open Repositories
Researchers

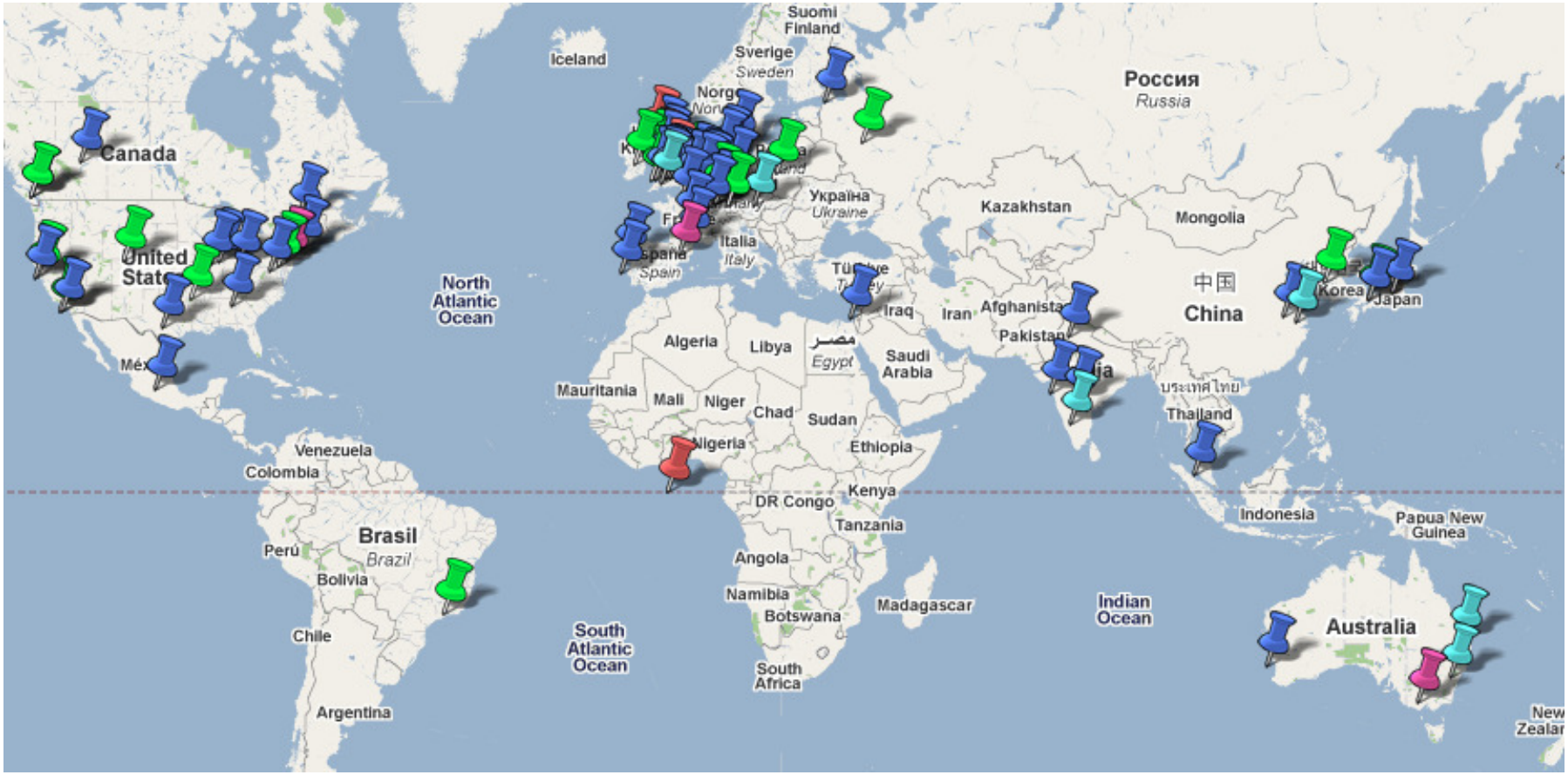
myexperiment

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

- Open source (BSD) Ruby on Rails app
- REST and SPARQL interfaces, Linked Data compliant
- Basis or inspiration for multiple projects including BioCatalogue, MethodBox and SysmoDB

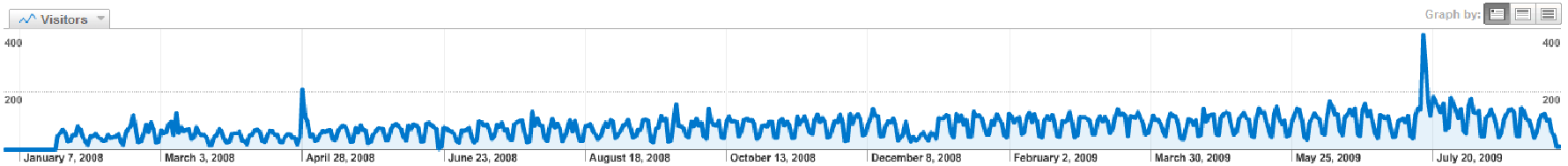
myExperiment currently has 4034 members, 231 groups, 1165 workflows, 326 files and 118 packs

myGrid MANCHESTER 1824 University of Southampton School of Electronics and Computer Science JISC EPSRC Microsoft Research omii-uk www.omii.ac.uk e-Research South



Visitors Overview

Jan 1, 2008 - Sep 8, 2009



36,139 people visited this site

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

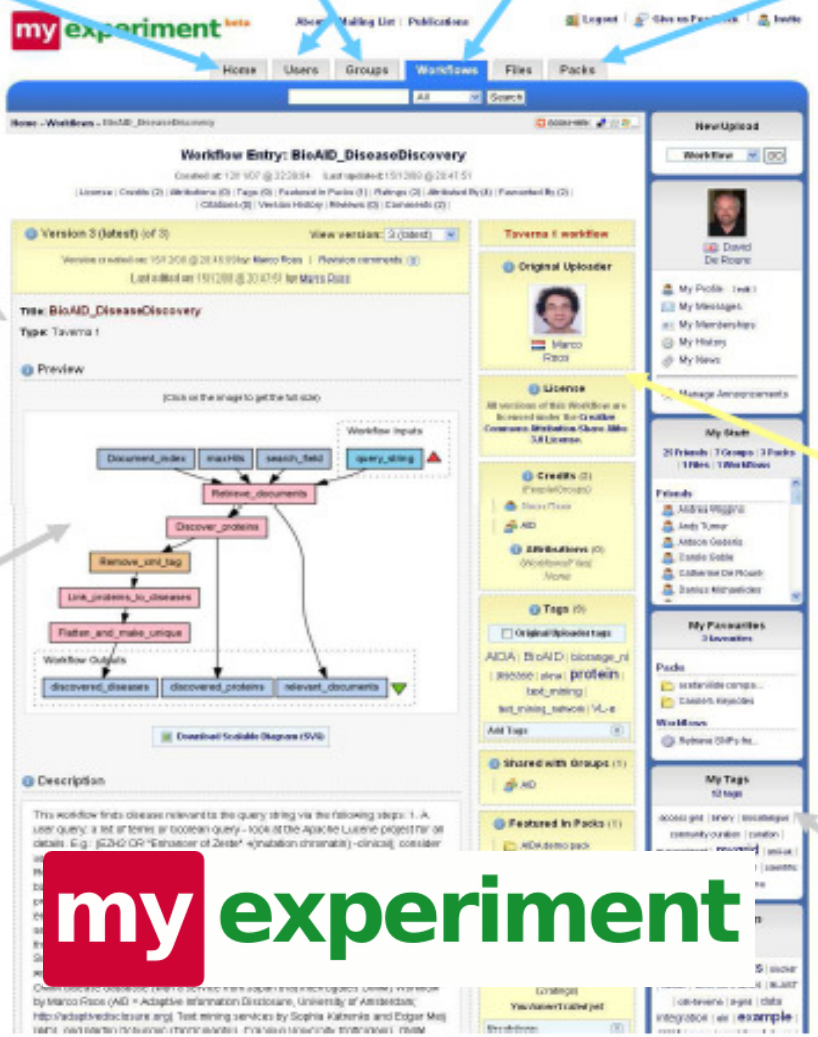
See the latest and most popular workflows: discover, view, download, run, tag and rate. Upload your workflows.

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.

myexperiment

JISC EPSRC

Microsoft

myExperiment Features

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

Distinctives

- Credits and Attributions
- Fine control over privacy
- Packs
- Multiple instances
- Enactment

Logs

Analysis Protocol for Candidate Genes and Pathways

This protocol is aimed at providing a guide to the interpretation of the results obtained from both the QTL and microarray workflows. Each workflow provides a series of text files, which are to be used as a means of obtaining the pathways which relate to differentially expressed genes in the microarray study and genes located within the chosen QTL region.

The output from each workflow consists of the following files:

- ensembl_database_release.text
- pathway_descriptions.text
- gene_descriptions.text
- pathway_descriptions.text
- merged_pathways.text
- kegg_external_gene_reference.text
- report.text
- pathway_list.text

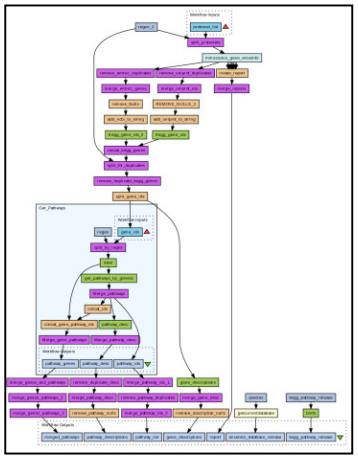
ensembl_database_release.text

The current release of the Ensembl dataset for the chosen species, e.g. *Mus musculus*. Although this uses the programmatic interface of Ensembl, it can be used to identify which release was used to generate the list of genes in the QTL region or mapping of Affymetrix probesets identifiers.

Results

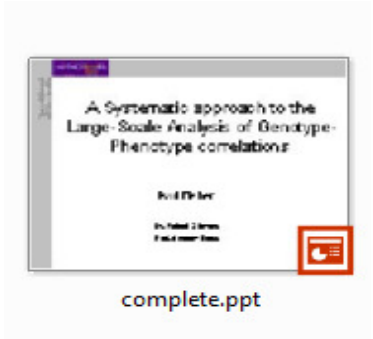
```
path:mmu04060 Cytokine-cytokine receptor interaction - Mus musculus (mouse)
path:mmu00970 Aminoacyl-tRNA biosynthesis - Mus musculus (mouse)
path:mmu00240 Pyrimidine metabolism - Mus musculus (mouse)
path:mmu03010 Ribosome - Mus musculus (mouse)
path:mmu04080 Neuroactive ligand-receptor interaction - Mus musculus (mouse)
path:mmu04210 Apoptosis - Mus musculus (mouse)
path:mmu05220 Chronic myeloid leukemia - Mus musculus (mouse)
path:mmu04612 Antigen processing and presentation - Mus musculus (mouse)
path:mmu00271 Methionine metabolism - Mus musculus (mouse)
path:mmu04912 GnRH signaling pathway - Mus musculus (mouse)
path:mmu04330 Notch signaling pathway - Mus musculus (mouse)
path:mmu04640 Hematopoietic cell lineage - Mus musculus (mouse)
path:mmu00561 Glycerolipid metabolism - Mus musculus (mouse)
path:mmu04110 Cell cycle - Mus musculus (mouse)
path:mmu04530 Tight junction - Mus musculus (mouse)
path:mmu02010 ABC transporters - General - Mus musculus (mouse)
```

Workflow 16



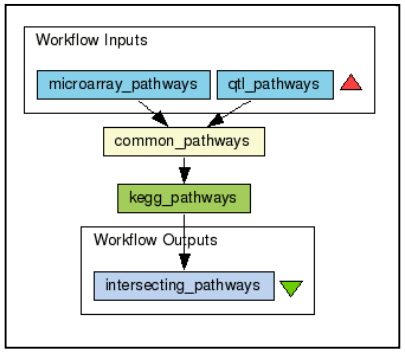
QTL

A Pack



complete.ppt

Metadata



Common pathways

Workflow 13

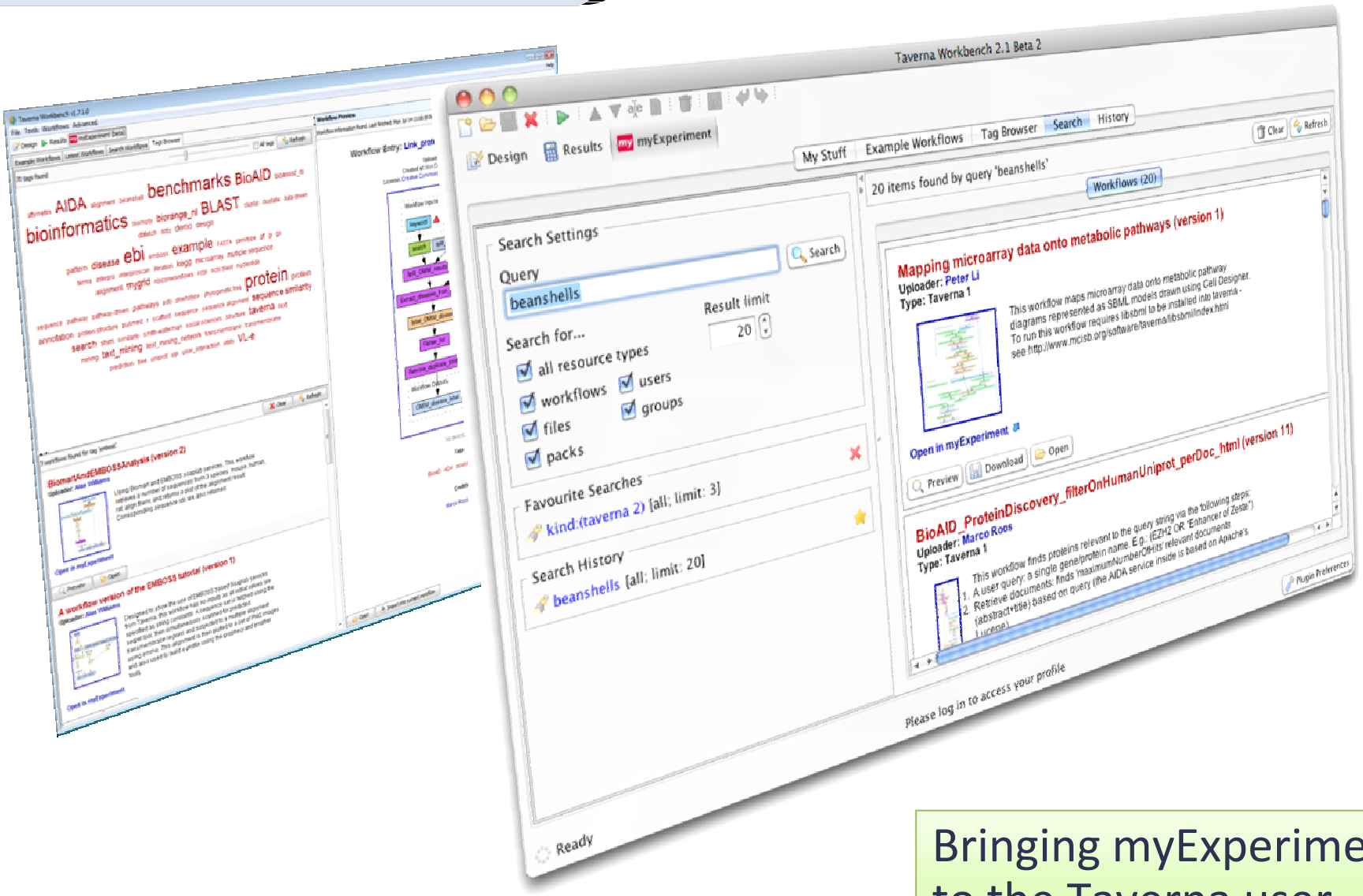
Slides

Paper

Primer Name	Left Flank Nucleotide sequence	Right Flank Nucleotide sequence
DAXX_1274_1812	CAGGAGGAATGGCGAGTG	AGCTTAGTCCTTCCCAAGCC
DAXX_140754_456_1070	CTTGTAGGATTGGGACTGGG	TCTCCTCCTCTTCCTCCTCC
DAXX_2270_2720	TGGGCAGGAGAGATGGTTC	ATGGTTCAAGGGAAGGGAAA
DAXX_2644_3187	TGTGTGATTGGCTGGTTGTT	GCAAATACGAGGAGTCTGGG
DAXX_exon5	TCCTCCTCCTACCAATCAAA	AGCAGAACTAACACCACAAGG
Daxx_Upst_479_1104	CAGGCTTCCTCATCAACACC	TGTCTCTATGGCTGTGCAGG

Results

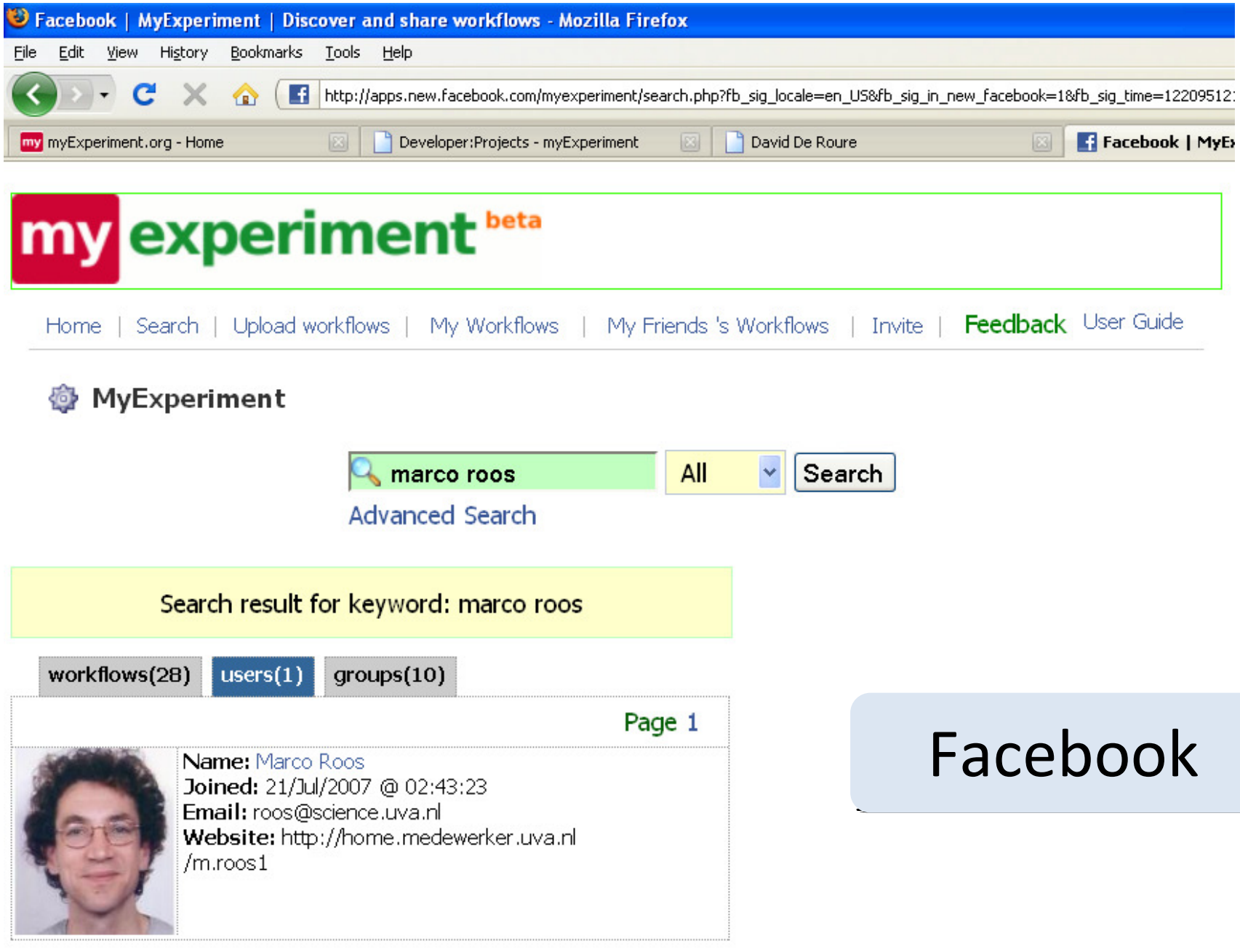
Taverna Plugins

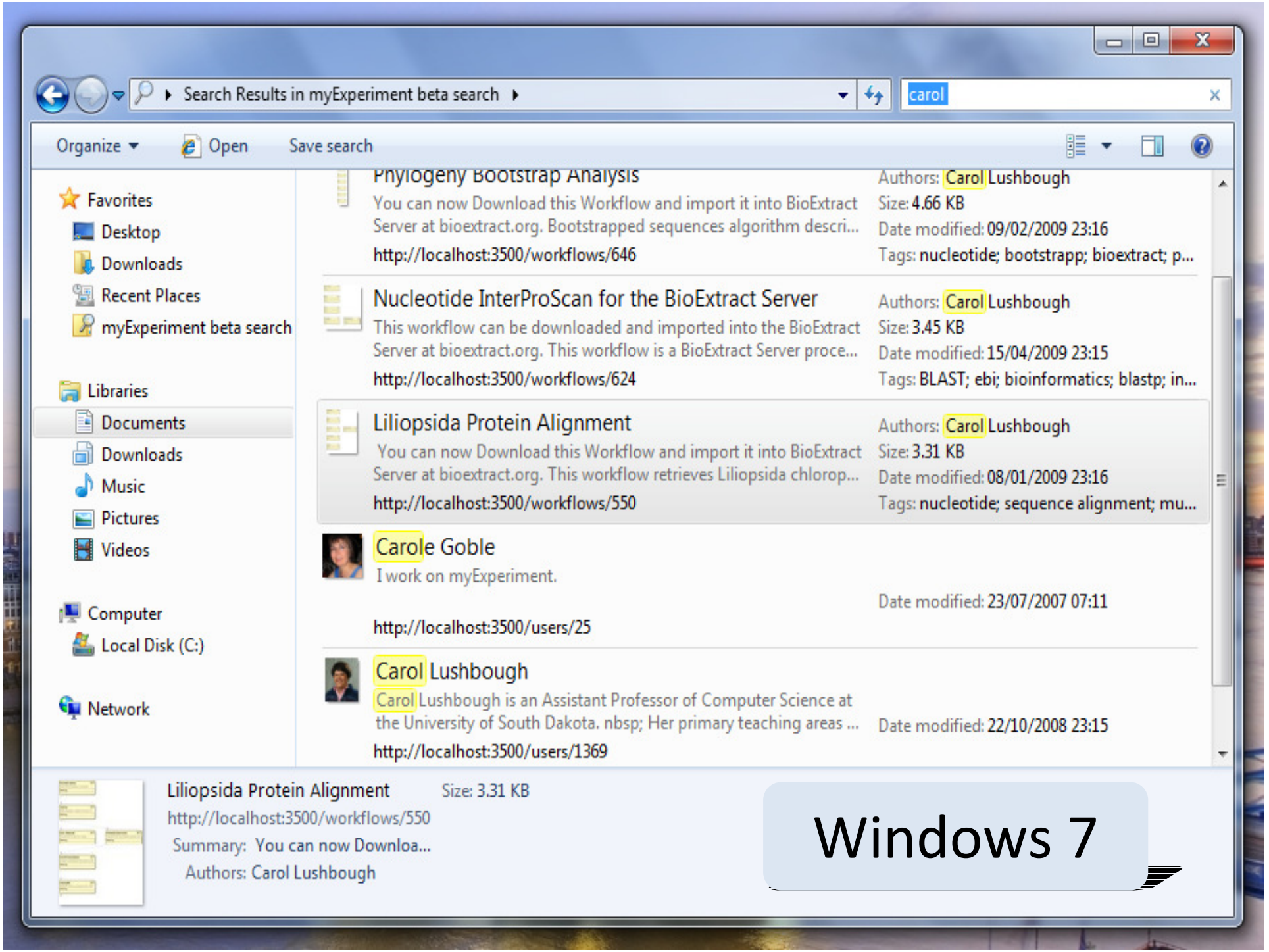


Bringing myExperiment to the Taverna user

Google Gadgets

Bringing myExperiment to the iGoogle user


Facebook | MyExperiment | Discover and share workflows - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://apps.new.facebook.com/myexperiment/search.php?fb_sig_locale=en_US&fb_sig_in_new_facebook=1&fb_sig_time=12209512
myExperiment.org - Home
Developer:Projects - myExperiment
David De Roure
Facebook | MyEx
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


Search Results in myExperiment beta search
carol
Organize
Open
Save search
Favorites
Desktop
Downloads
Recent Places
myExperiment beta search
Libraries
Documents
Downloads
Music
Pictures
Videos
Computer
Local Disk (C:)
Network
Phylogeny Bootstrap Analysis
You can now Download this Workflow and import it into BioExtract Server at bioextract.org. Bootstrapped sequences algorithm descri...
http://localhost:3500/workflows/646
Authors: Carol Lushbough
Size: 4.66 KB
Date modified: 09/02/2009 23:16
Tags: nucleotide; bootstrapp; bioextract; p...
Nucleotide InterProScan for the BioExtract Server
This workflow can be downloaded and imported into the BioExtract Server at bioextract.org. This workflow is a BioExtract Server proce...
http://localhost:3500/workflows/624
Authors: Carol Lushbough
Size: 3.45 KB
Date modified: 15/04/2009 23:15
Tags: BLAST; ebi; bioinformatics; blastp; in...
Liliopsida Protein Alignment
You can now Download this Workflow and import it into BioExtract Server at bioextract.org. This workflow retrieves Liliopsida chlorop...
http://localhost:3500/workflows/550
Authors: Carol Lushbough
Size: 3.31 KB
Date modified: 08/01/2009 23:16
Tags: nucleotide; sequence alignment; mu...
Carole Goble
I work on myExperiment.
http://localhost:3500/users/25
Date modified: 23/07/2007 07:11
Carol Lushbough
Carol Lushbough is an Assistant Professor of Computer Science at the University of South Dakota. nbsp; Her primary teaching areas ...
http://localhost:3500/users/1369
Date modified: 22/10/2008 23:15
Liliopsida Protein Alignment
Size: 3.31 KB
http://localhost:3500/workflows/550
Summary: You can now Downloa...
Authors: Carol Lushbough
Windows 7

http://www.myexperiment.org/packs/112 Open

marbles

http://www.myexperiment.org/packs/112

type	• http://rdf.myexperiment.org/ontologies/packs/Pack
http://purl.org/dc/terms/title	• Presentation - Anchors in the Shifting Sand: The Primacy of Method in the Web of Data
http://www.openarchives.org/ore/terms/aggregates	• http://dl.dropbox.com/u/1202407/Presentations/myExpWebSci2010.ppt • http://www.slideshare.net/dder/anchors-in-shifting-sand-the-primacy-of-method-in-the-web-of-data • http://dl.dropbox.com/u/1202407/Presentations/myExpWebSci2010.pdf • http://journal.webscience.org/325/ • http://eprints.ecs.soton.ac.uk/id/eprint/20817

http://www.openarchives.org/ore/terms/aggregates

http://eprints.ecs.soton.ac.uk/id/eprint/20817

EPrints

http://eprints.ecs.soton.ac.uk/id/eprint/20817 Open

marbles

http://eprints.ecs.soton.ac.uk/id/eprint/20817

type	• http://eprints.org/ontology/EPrint • http://purl.org/ontology/bibo/AcademicArticle • http://purl.org/ontology/bibo/Article • http://eprints.org/ontology/ConferenceItemEPrint
seeAlso	• HTML Summary of of #20817 Anchors in Shifting Sand: the Primacy of Method in the Web of Data
http://purl.org/dc/terms/title	• Anchors in Shifting Sand: the Primacy of Method in the Web of Data
http://purl.org/dc/terms/creator	• Professor David C De Roure (also at eprints.ecs.soton.ac.uk) • http://eprints.ecs.soton.ac.uk/id/person/ext-2b740eab33bb7da43e Source: http://rdf.ecs.soton.ac.uk/person/47
http://purl.org/ontology/bibo/abstract	• The wealth of new government and scientific data appearing on the Web is to be welcomed and makes it possible for citizens and scientists to interpret evidence and obtain new insights. But how will they do this, and how will people trust the results? We suggest the Linked Data Web must embrace the “methods” by which results are obtained as well as the results themselves. By making methods first class citizens, results can be explained, interpreted and assessed, and the methods themselves can be shared, discussed, reused and repurposed. We present the myExperiment.org website, a social network of people sharing reusable methods for processing research data, and make some observations on the nature of first class methods in the Web of Data.

ECS id

http://id.ecs.soton.ac.uk/person/47 Open

marbles

Professor David C De Roure

type	• Person • http://rdf.ecs.soton.ac.uk/ontology/ecs#Person
sameAs	• Professor David C De Roure
is sameAs of	• Professor David C De Roure (http://eprints.ecs.soton.ac.uk/id/person/ext-47)
name	• Professor David C De Roure
personal mailbox	• mailto:dder@ecs.soton.ac.uk
Homepage	http://users.ecs.soton.ac.uk/dder/
image	

<sameAs>

interlinking the Web of Data

The Web of Data has many equivalent URIs.
This service helps you to find co-references
between different data sets.
Enter a known URI, or use Sindice to search first.

<sameAs> http://id.ecs.soton.ac.uk/person/47

sindice Enter a literal (text) search...

Equivalent URIs for http://id.ecs.soton.ac.uk/person/47 –

<sameAs>

1. http://southampton.rkbexplorer.com/id/person-00047
2. http://southampton.rkbexplorer.com/id/person-005e1d76ede30a16c475e8...7d5b58d159
3. http://southampton.rkbexplorer.com/id/person-2126aee62bd8b58f82c6ff...f484814122

Show 502 more

506. http://webscience.org/person/26

rdf+xml · n3 · json · text

Currently serving **35187488** URIs in **11285263** bundles!

about · contact · get the widget

Overview

- Motivation: the primacy of method
- myExperiment and Other Animals
- Design and implementation
- The future of research

Web 2.0 patterns

The Long Tail

Data is the Next “Intel Inside”

Users add value

Network effects by default

Some Rights Reserved

The Perpetual Beta

Cooperate, don’t Control

Software above the level of the single device

Software Design for Empowering Scientists

David De Roure, *University of Southampton*

Carole Goble, *University of Manchester*

The Taverna Workbench, a scientific workflow-management system, and the myExperiment social Web site for sharing scientific experiments follow six principles for designing software for scientists.

Science is becoming increasingly digital. Scientists' tools are not just the experimental apparatus of the laboratory but are also the software apparatus they use to conduct their research, analyze data, search databases, run simulations, and record their scientific process. New scientific techniques—from DNA microarrays to sensor networks in the environment—are generating volumes of data that wouldn't get processed without software assistance.

Watch researchers at work, and you'll see a lot of computer activity as they use applications, services, and data that might be local to the laboratory or enterprise, or accessed on the Web. These new research tools and methods are evident across a broad spectrum of disciplines. Some researchers, such as bioinformaticians working with protein sequences, might conduct research entirely *in silico*. Meanwhile, chemists in the laboratory are using computers for look-

these challenges: Taverna provides automation of scientific data processing tasks, making them systematic and repeatable, whereas myExperiment facilitates the discovery and sharing of scientific digital objects, encouraging reuse and avoiding reinvention. These tools' comparative success owes much to a software design approach that understands the practice and culture of scientists.

Scientific Workflows

Six Principles of Software Design to Empower Scientists

1. Fit in, Don't Force Change
2. Jam today and more jam tomorrow
3. Just in Time and Just Enough
4. Act Local, think Global
5. Enable Users to Add Value
6. Design for Network Effects

1. Keep your Friends Close
2. Embed
3. Keep Sight of the Bigger Picture
4. Favours will be in your Favour
5. Know your users
6. Expect and Anticipate Change

De Roure, D. and Goble, C. "Software Design for Empowering Scientists," IEEE Software, vol. 26, no. 1, pp. 88-95, January/February 2009

For Developers

myexperiment

HTML

XML

facebook

iGoogle

android

API config

ORE
FOAF
SIOC

Managed REST API

tags ratings reviews profiles
workflows credits groups
files packs friendships

Search API

Search Engine

SPARQL endpoint

RDF Store

EPrints
DSpace
Fedora
S3
SRB

mySQL

Enactor API

Enactor

SPARQL endpoint

rdf.myexperiment.org

SPARQL endpoint

tags ratings reviews profiles
workflows credits groups
files packs friendships

Transform

RDF Store

mySQL

myExperiment data model (evolving!)

Modularised myExperiment Ontology

DC, FOAF, SIOC
(Semantically-Interlinked Online Communities)

How To SPARQL

Submit Feedback/Bug Report

myExperiment's SPARQL endpoint allows anyone to query all of myExperiment's public data using the query language SPARQL.

Although SPARQL is quite easy to use it may take a while to get you head round it if you are coming to it new and you are not from a computer science background. This guide is intended to help explain the basics of SPARQL and give usable examples that will return actual results from myExperiment's SPARQL endpoint, which can be compared against example results. Although it should be noted that example results are intended as a guide to the format of results and will probably not be exactly the same as those returned by the SPARQL endpoint.

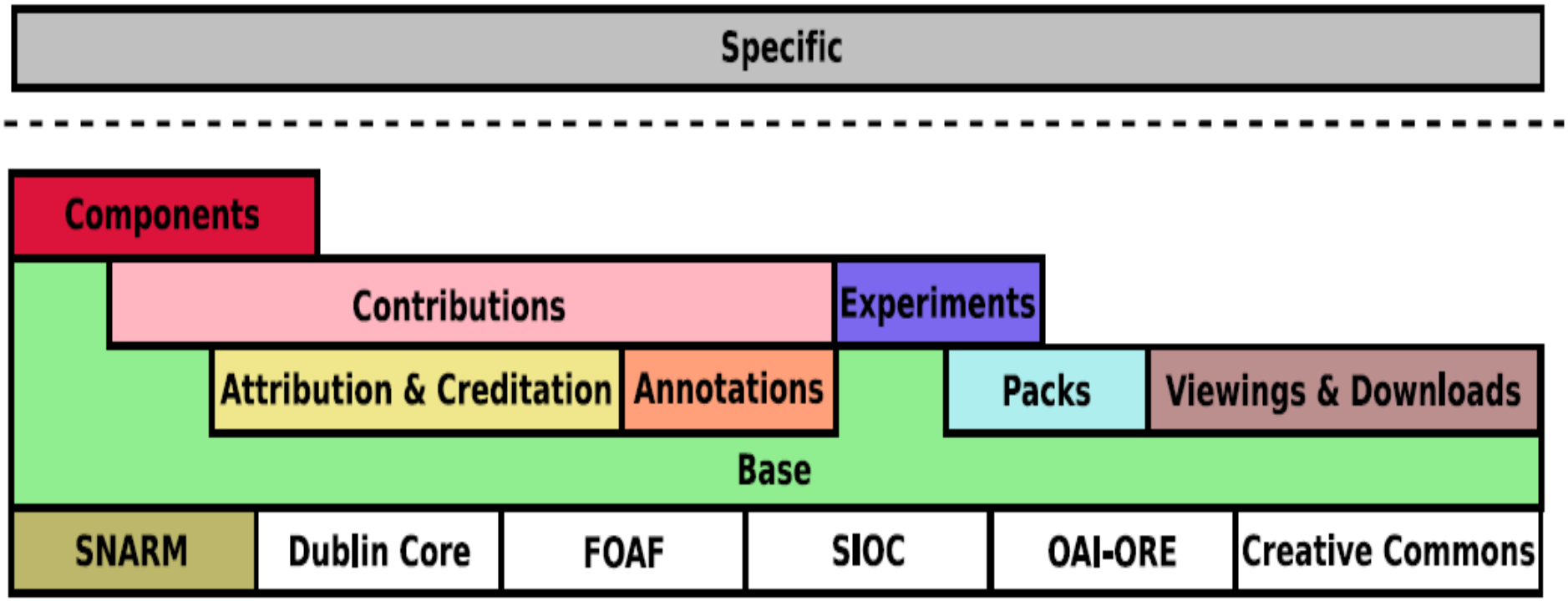
The first section explains how to use the SPARQL endpoint and the unique features the myExperiment SPARQL endpoint has.

Sections 2-7 explain the main clauses used within SPARQL and how they are used to define queries.

Section 8 give some explanation of various warning and error messages you might be prompted with and how to fix you query to eliminate these.

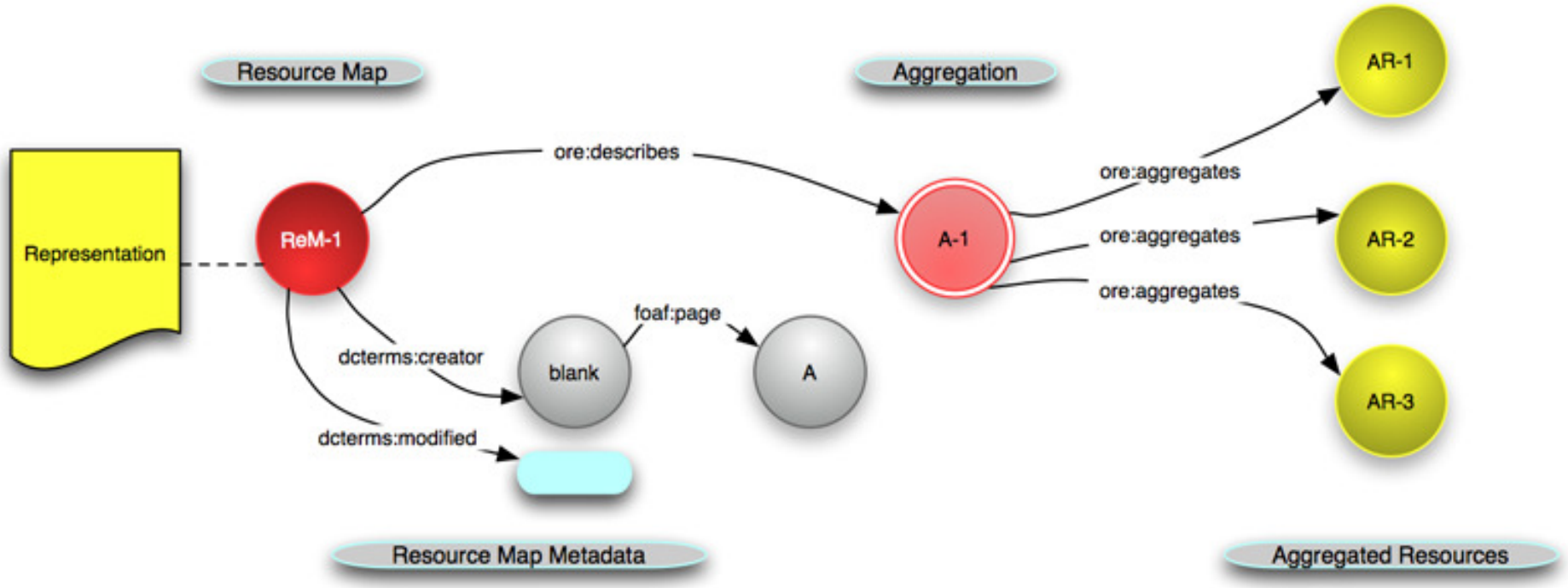
David Newman

myExperiment modularised ontology

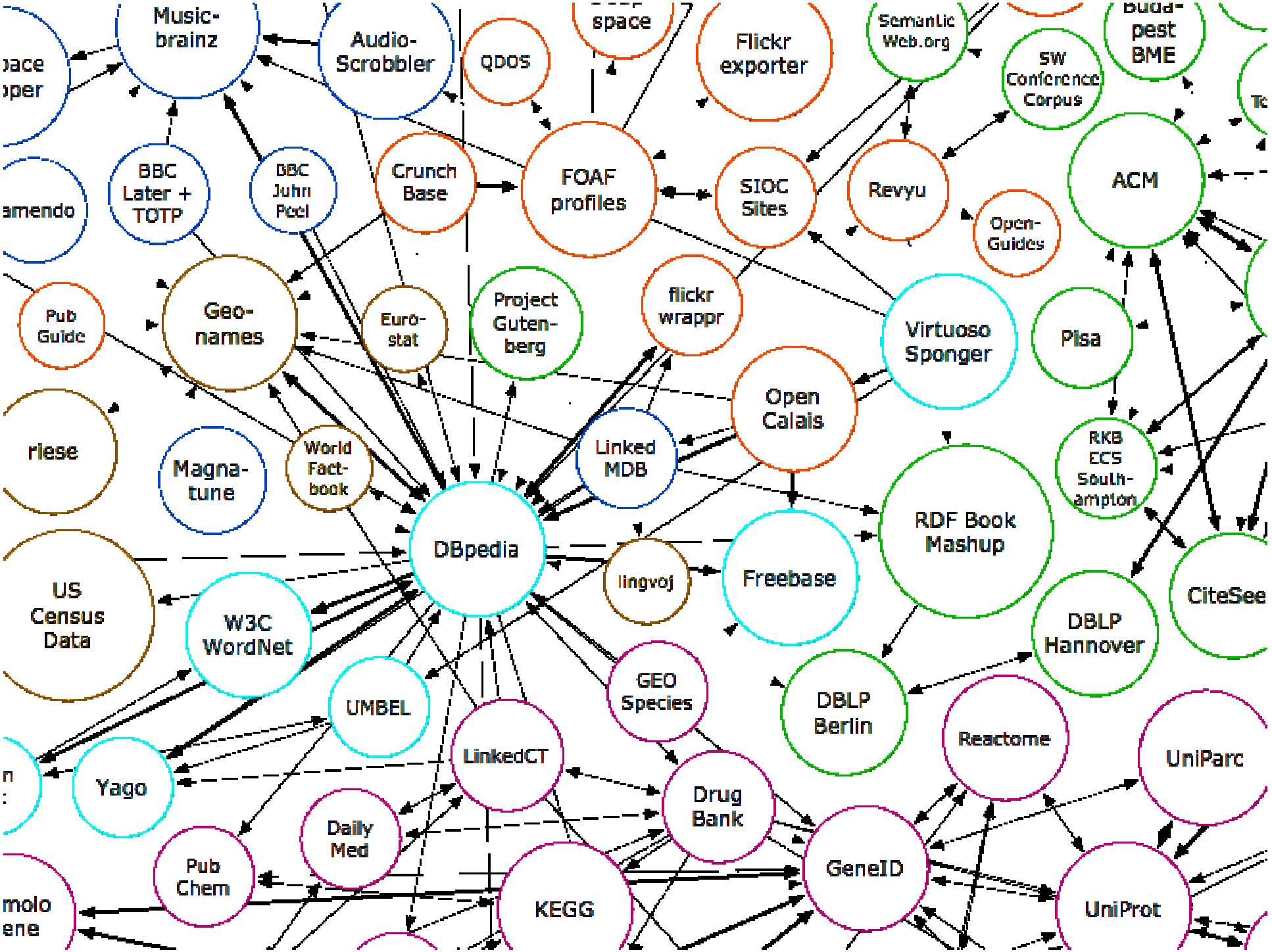


David Newman

Exporting packs



Open Archives Initiative
Object Reuse and Exchange


Music-
brainz
Audio-
Scrobbler
QDOS
space
Flickr
exporter
Semantic
Web.org
SW
Conference
Corpus
Budapest
BME
BBC
Later +
TOTP
BBC
John
Peel
Crunch
Base
FOAF
profiles
SIOC
Sites
Revyu
Open-
Guides
ACM
Pub
Guide
Geo-
names
Euro-
stat
Project
Guten-
berg
flickr
wrappr
Virtuoso
Sponger
Pisa
Open
Calais
RKB
ECS
South-
ampton
Magna-
tune
World
Fact-
book
Linked
MDB
DBpedia
RDF Book
Mashup
US
Census
Data
W3C
WordNet
lingvoj
Freebase
CiteSeer
DBLP
Hannover
UMBEL
GEO
Species
DBLP
Berlin
Reactome
UniParc
Yago
LinkedCT
Drug
Bank
Daily
Med
Pub
Chem
KEGG
GeneID
UniProt

Levels of (social) compliance?

- 303s
- 303s + RDF
- 303s + RDF + SPARQL
- Being on the diagram!

1. Use URIs as names for things

- **Everything**
  - If you don't name something you can't talk about it
  - Things of course
  - Year of publication
  - Ideas
  - …
- **Cool URIs**
  - Think of the consumer/customer
    - https://secure.ecs.soton.ac.uk/gizmos/person_by_username.php?username=hg
    - https://secure.ecs.soton.ac.uk/person/username/hg
- **RESTful Interfaces**
- **Ambiguity**
  - URIs help to avoid it, especially if you…

2. Use HTTP URIs so that people can look up those names

- HTTP URI names come with an excellent look up mechanism
- And ownership, etc.

3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)

- **So they know what you mean**
- **Deliver some human readable data**
  - html
- **Deliver some machine processable data**
  - RDF
  - JSON
  - CSV
  - text

4. Include links to other URIs. so that they can discover more things

- **“Foreign” URIs**
  - **dbpedia:Southampton_University**
- **Equivalence**
  - owl:sameAs
  - skos:exactMatch
  - …

URI	Accept	Redirect	Response URI
http://www.myexperiment.org/workflows/16	text/html	303	http://www.myexperiment.org/workflows/16.html
http://www.myexperiment.org/workflows/16	application/rdf+xml	303	http://www.myexperiment.org/workflows/16.rdf
http://www.myexperiment.org/workflows/16	text/xml	303	http://www.myexperiment.org/workflows/16.xml
http://www.myexperiment.org/workflows/16/versions/n	text/html	303	http://www.myexperiment.org/workflows/16/versions/n.html
http://www.myexperiment.org/workflows/16/versions/n	application/rdf+xml	303	http://www.myexperiment.org/workflows/16/versions/n.rdf
http://www.myexperiment.org/workflows/16/versions/n	text/xml	303	http://www.myexperiment.org/workflows/16/versions/n.xml
http://www.myexperiment.org/workflows/16.html	text/html		http://www.myexperiment.org/workflows/16.html
http://www.myexperiment.org/workflows/16.html	application/rdf+xml	301	http://www.myexperiment.org/workflows/16
http://www.myexperiment.org/workflows/16.html	text/xml	301	http://www.myexperiment.org/workflows/16
http://www.myexperiment.org/workflows/16/versions/n.html	text/html		http://www.myexperiment.org/workflows/16/versions/n.html
http://www.myexperiment.org/workflows/16/versions/n.html	application/rdf+xml	301	http://www.myexperiment.org/workflows/16/versions/n
http://www.myexperiment.org/workflows/16/versions/n.html	text/xml	301	http://www.myexperiment.org/workflows/16/versions/n
http://www.myexperiment.org/workflows/16.rdf	text/html	301	http://www.myexperiment.org/workflows/16
http://www.myexperiment.org/workflows/16.rdf	application/rdf+xml		http://www.myexperiment.org/workflows/16.rdf
http://www.myexperiment.org/workflows/16.rdf	text/xml	301	http://www.myexperiment.org/workflows/16
http://www.myexperiment.org/workflows/16/versions/n.rdf	text/html	301	http://www.myexperiment.org/workflows/16/versions/n
http://www.myexperiment.org/workflows/16/versions/n.rdf	application/rdf+xml		http://www.myexperiment.org/workflows/16/versions/n.rdf
http://www.myexperiment.org/workflows/16/versions/n.rdf	text/xml	301	http://www.myexperiment.org/workflows/16/versions/n
http://www.myexperiment.org/workflows/16.xml	text/html	301	http://www.myexperiment.org/workflows/16
http://www.myexperiment.org/workflows/16.xml	application/rdf+xml	301	http://www.myexperiment.org/workflows/16
http://www.myexperiment.org/workflows/16.xml	text/xml	303	http://www.myexperiment.org/workflow.xml?id=16
http://www.myexperiment.org/workflows/16/versions/n.xml	text/html	301	http://www.myexperiment.org/workflows/16/versions/n
http://www.myexperiment.org/workflows/16/versions/n.xml	application/rdf+xml	301	http://www.myexperiment.org/workflows/16/versions/n
http://www.myexperiment.org/workflows/16/versions/n.xml	text/xml	303	http://www.myexperiment.org/workflow.xml?id=16&version=n
http://www.myexperiment.org/workflow.xml?id=16	text/html	301	http://www.myexperiment.org/workflows/16
http://www.myexperiment.org/workflow.xml?id=16	application/rdf+xml	301	http://www.myexperiment.org/workflows/16
http://www.myexperiment.org/workflow.xml?id=16	text/xml		http://www.myexperiment.org/workflow.xml?id=16

Linked Data

Submit Feedback/Bug Report

Linked Data is a way of interconnecting data published on the web that has not been previously linked. myExperiment is now a publisher of such data.

Every myExperiment entity, whether it be a Workflow, Pack, User, Group, etc. has its own Non-Information Resource (NIR) URI to identify it. This URI can also be used in a HTTP request where the header's accept parameter is set to an appropriate MIME type to retrieve data about it in one of up to three different formats, HTML, RDF and XML. This is done through a process called Content Negiotiation, which redirects to an explicit URL for that particular format. E.g.

- wget --header "Accept: text/html" http://www.myexperiment.org/workflows/16 → http://www.myexperiment.org/workflows/16.html
- wget --header "Accept: application/rdf+xml" http://www.myexperiment.org/workflows/16 → http://www.myexperiment.org/workflows/16.rdf
- wget --header "Accept: application/xml" http://www.myexperiment.org/workflows/16 → http://www.myexperiment.org/workflows/16.xml

If you are using an application that does not allow you to specify parameters of the HTTP request you can use the explicit URL for the different formats to retrieve data in that format.

Ontology

The structure of myExperiment RDF is defined by the myExperiment Ontology. This is a set of modules that borrows classes/properties from FOAF, SIOC, Dublin Core, Creative Commons and OAI-ORE, that can be assembled to build a comprehensive specification for the myExperiment data model. Auto-generated documentation is available that describes the documents and classes defined within the ontology.

SPARQL Endpoint

All myExperiment's public RDF data can queried using the query language SPARQL at myExperiment's SPARQL Endpoint. An introduction to SPARQL with a guide to querying myExperiment RDF is available here.

Vocabulary of Interlinked Datasets (VoID)

A description of myExperiment RDF is specified in voiD. One thing voiD encourages is the publication of RDF datasets so they can easily be reused by others rather than having to crawl each NIR for its RDF. All of myExperiment's Public RDF can be downloaded as a gzipped file.

David Newman

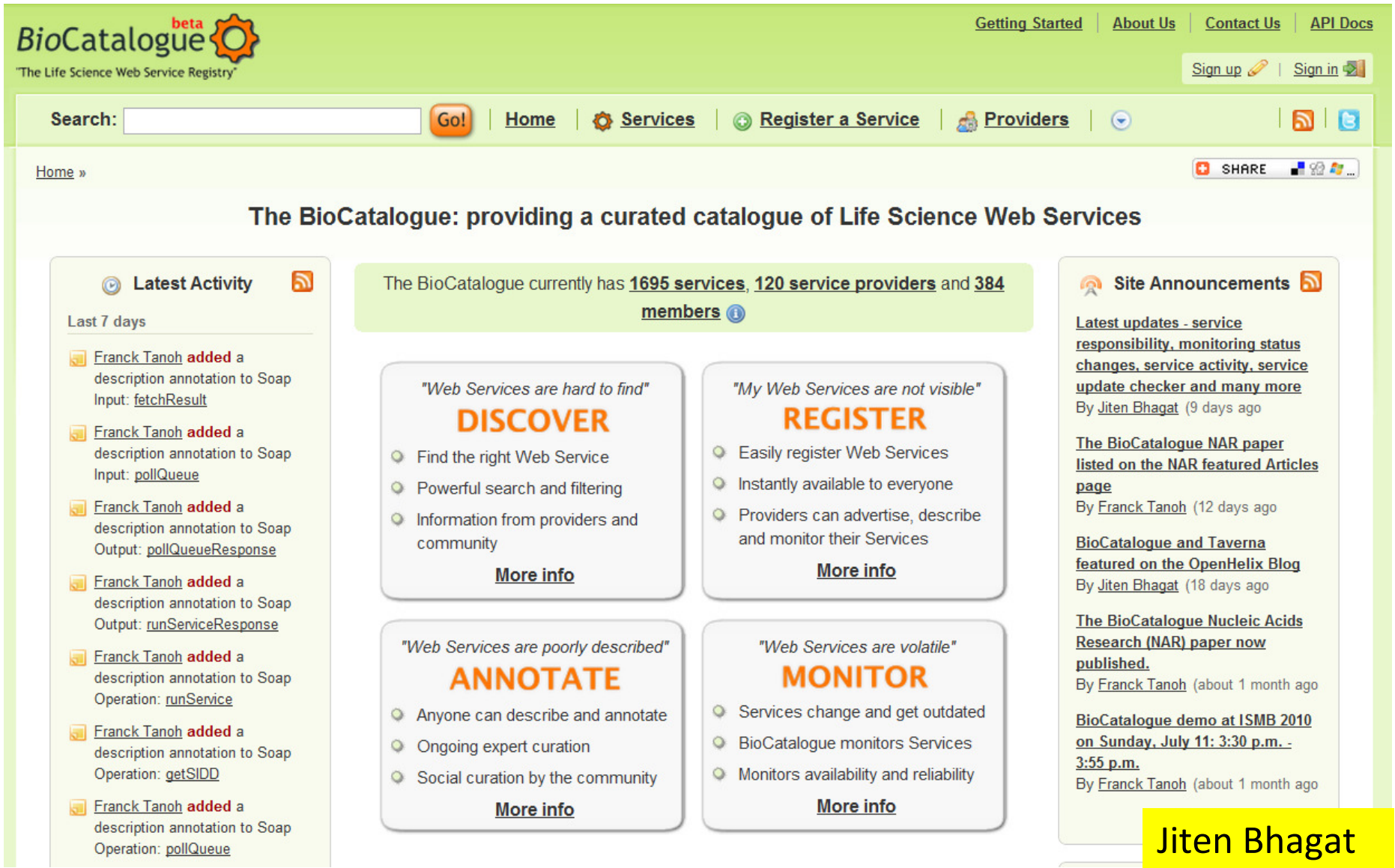
The hidden costs of linked data

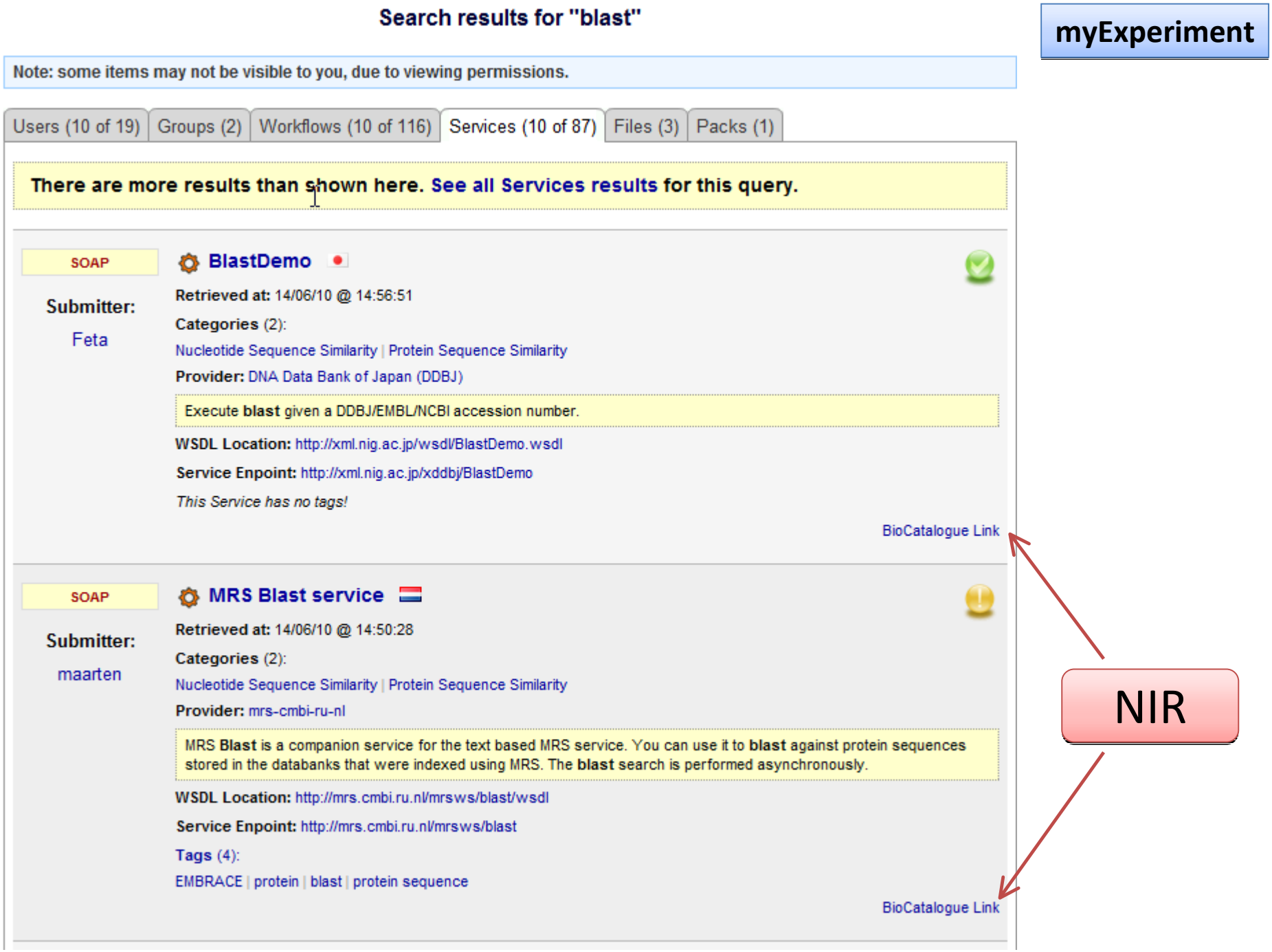
- Usability
  - We had a perfectly good scheme before and now we change it for something more complicated!
- Performance
  - All those 303s!
  - Rumoured that on some sites developers append .xml to save round trips

www.myexperiment.org/packs/112

www.myexperiment.org/packs/112.html

BioCatalogue




Search results for "blast"
myExperiment
Note: some items may not be visible to you, due to viewing permissions.
Users (10 of 19)
Groups (2)
Workflows (10 of 116)
Services (10 of 87)
Files (3)
Packs (1)
There are more results than shown here. See all Services results for this query.
SOAP
BlastDemo
Submitter:
Feta
Retrieved at: 14/06/10 @ 14:56:51
Categories (2):
Nucleotide Sequence Similarity | Protein Sequence Similarity
Provider: DNA Data Bank of Japan (DDBJ)
Execute blast given a DDBJ/EMBL/NCBI accession number.
WSDL Location: http://xml.nig.ac.jp/wsdl/BlastDemo.wsdl
Service Enpoint: http://xml.nig.ac.jp/xddbj/BlastDemo
This Service has no tags!
BioCatalogue Link
SOAP
MRS Blast service
Submitter:
maarten
Retrieved at: 14/06/10 @ 14:50:28
Categories (2):
Nucleotide Sequence Similarity | Protein Sequence Similarity
Provider: mrs-cmbi-ru-nl
MRS Blast is a companion service for the text based MRS service. You can use it to blast against protein sequences stored in the databanks that were indexed using MRS. The blast search is performed asynchronously.
WSDL Location: http://mrs.cmbi.ru.nl/mrsws/blast/wsdl
Service Enpoint: http://mrs.cmbi.ru.nl/mrsws/blast
Tags (4):
EMBRACE | protein | blast | protein sequence
BioCatalogue Link
NIR

Overview

- Motivation: the primacy of method
- myExperiment and Other Animals
- Design and implementation
- The future of research

Packs in Practice

Taverna 1.7.1 starter pack

Created: 17/07/08 @ 21:06:12 | **Last updated:** 20/07/08 @ 15:46:51

Everything to get started with Taverna 1.7.1

16

Comm

Tags

exam

Towards Genotype-Phenotype Correlations

Created: 08/04/09 @ 13:14:54 | **Last updated:** 08/04/09 @ 13:16:23

It is increasingly common to combine Microarray and Quantitative Trait Loci data to aid the search for candidate genes responsible for phenotypic variation. Workflows provide a means of systematically proces
declara
manua
phe...

19 iter

Commer

Tags:

affymetri
KeggID |
swissprc

myExperiment paper for Concurrency Practice and Experience eScience 2008 Special Issue

Created: 10/04/09 @ 13:41:56 | **Last updated:** 14/04/09 @ 08:34:12

This pack contains the materials used in the paper De Roure, D., Goble, C., Aleksejevs, S., Bechhofer, S., Bhagat, J., Cruickshank, D., Fisher, P., Hull, D., Michaelides, D., Newman, D., Procter, R., L
Con
issu
The

8 ite

Comm

Tags:

curati

Workflow discovery benchmarks

Created: 12/07/08 @ 11:20:23 | **Last updated:** 09/10/08 @ 16:29:50

This pack contains benchmarks that measure how bioinformaticians discover Taverna workflows. Several subpacks are available: Collection of workflows by Paul Fisher, used in benchmarks PR2 and C

7

Co

Tag

ben

Creator:

David De Roure

Carole's Keynotes

Created: 27/09/08 @ 19:52:17 | **Last updated:** 15/10/08 @ 08:55:46

Carole Goble's 2008 keynote talks that feature myExperiment, in PowerPoint 97-2003 format. All myExperiment presentation can be found on the wiki on http://wiki.myexperiment.org/index.php/Presentations

5 items in this pack

Comments: 0 | **Viewed:** 86 times | **Downloaded:** 18 times

Tags:

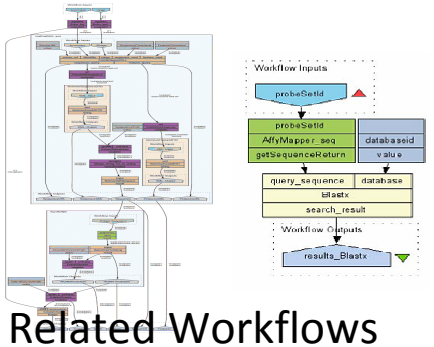
biocatalogue | community curation | curation | myexperiment | mygrid | open science | science 2.0 | scientific workflow | taverna

Knowledge Packages – More than Methods

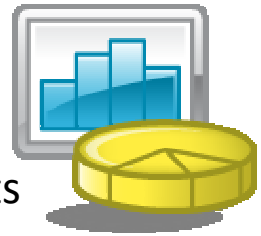
Data & Configuration

Metadata

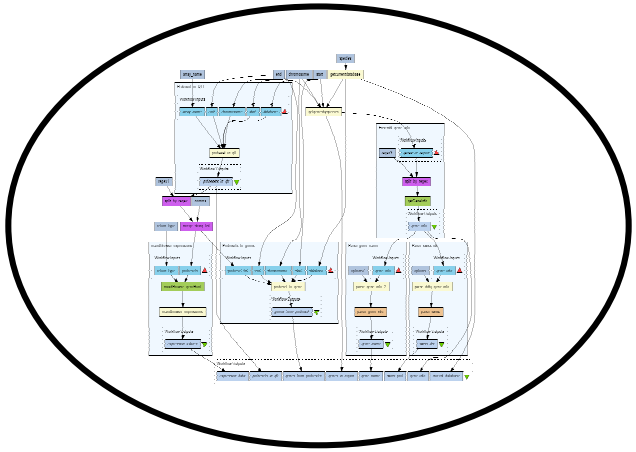
Reviews

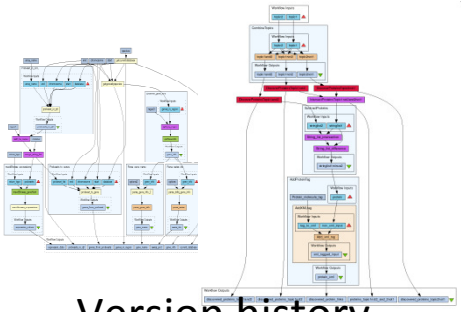


Related Workflows



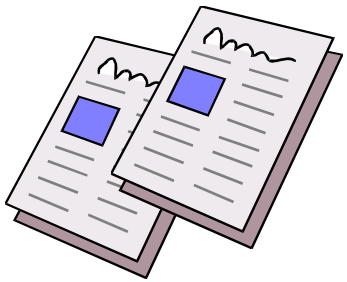
Results





Version history

Log Book
Provenance



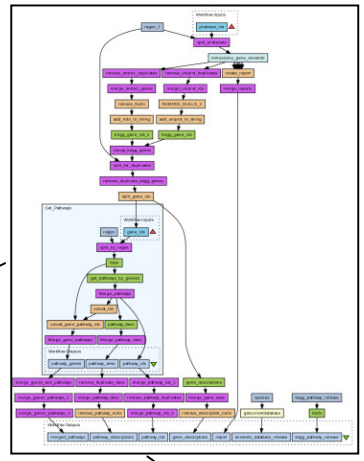
Publications and Presentations

Training
material

Carole Goble

Paul's Research Object

Workflow 16



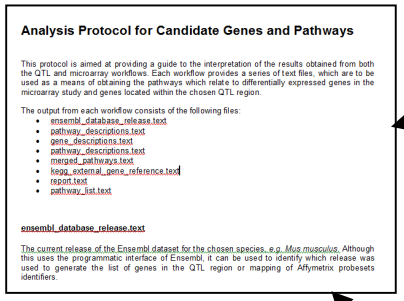
QTL

Results

```
path:mmu04060 Cytokine-cytokine receptor interaction - Mus musculus (mouse)
path:mmu00970 Aminoacyl-tRNA biosynthesis - Mus musculus (mouse)
path:mmu00240 Pyrimidine metabolism - Mus musculus (mouse)
path:mmu03010 Ribosome - Mus musculus (mouse)
path:mmu04080 Neuroactive ligand-receptor interaction - Mus musculus (mouse)
path:mmu04210 Apoptosis - Mus musculus (mouse)
path:mmu05220 Chronic myeloid leukemia - Mus musculus (mouse)
path:mmu04612 Antigen processing and presentation - Mus musculus (mouse)
path:mmu00271 Methionine metabolism - Mus musculus (mouse)
path:mmu04912 GnRH signaling pathway - Mus musculus (mouse)
path:mmu04330 Notch signaling pathway - Mus musculus (mouse)
path:mmu04640 Hematopoietic cell lineage - Mus musculus (mouse)
path:mmu00561 Glycerolipid metabolism - Mus musculus (mouse)
path:mmu04110 Cell cycle - Mus musculus (mouse)
path:mmu04530 Tight junction - Mus musculus (mouse)
path:mmu02010 ABC transporters - General - Mus musculus (mouse)
```

produces

Included in



Analysis Protocol for Candidate Genes and Pathways

This protocol is aimed at providing a guide to the interpretation of the results obtained from both the QTL and microarray workflows. Each workflow provides a series of text files, which are to be used as a means of obtaining the pathways which relate to differentially expressed genes in the microarray study and genes located within the chosen QTL region.

The output from each workflow consists of the following files:

- ensembl_database_release.text
- pathway_descriptions.text
- gene_descriptions.text
- pathway_descriptions.text
- merged_pathways.text
- kegg_external_gene_reference.text
- report.text
- pathway_list.text

ensembl_database_release.text

The current release of the Ensembl dataset for the chosen species, *e.g. Mus musculus*. Although this uses the programmatic interface of Ensembl, it can be used to identify which release was used to generate the list of genes in the QTL region or mapping of Affymetrix probesets identifiers.

Logs

produces

Feeds into

Included in

A Systematic approach to the Large-Scale Analysis of Genotype-Phenotype correlations

complete.ppt

Included in

Published in

PDF

Adobe

Included in

Slides

Paper

Metadata

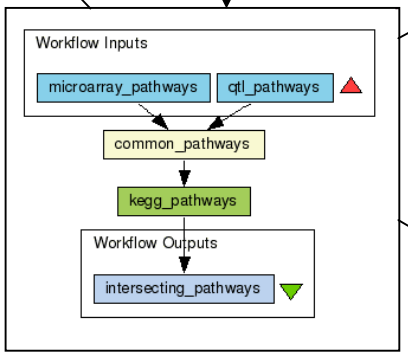
Tags (19)

Creator tags

affymetrix | african trypanosomiasis | cattle | data-driven | disease | entrez | genotype | Kegg Pathways | KeggID | link-integration | microarray | mouse | pathway | pathway-driven | phenotype | sleeping sickness | swissprot | uniprot | web services

[edit]

Add Tags



Workflow Inputs

microarray_pathways | qtl_pathways

common_pathways

kegg_pathways

Workflow Outputs

intersecting_pathways

Common pathways

Workflow 13

produces

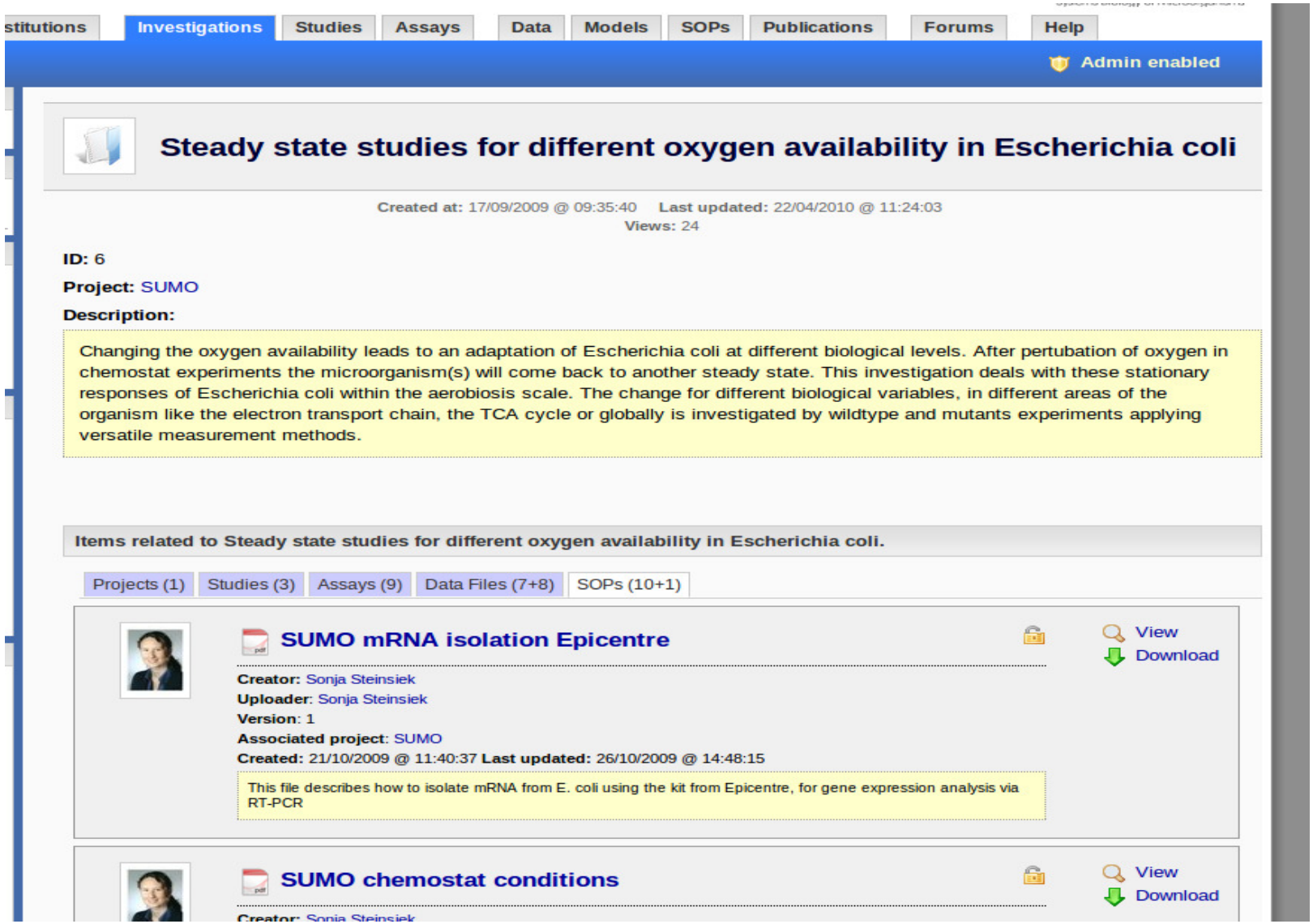
Primer Name	Left Flank Nucleotide sequence	Right Flank Nucleotide sequence
DAXX_1274_1812	CAGGAGGAATGGCGAGTG	AGCTTAGTCCTTCCCAAGCC
DAXX_140754_456_1070	CTTGTAGGATTGGGACTGGG	TCTCCTCCTCTTCCTCCTCC
DAXX_2270_2720	TGGGCAGGAGAGATGGTTC	ATGGTTCAAGGGAAGGGAAA
DAXX_2644_3187	TGTGTGATTGGCTGGTTGTT	GCAAATACGAGGAGTCTGGG
DAXX_exon5	TCCTCCTCCTACCAATCAAA	AGCAGAACTAACACCACAAGG
Daxx_Upst_479_1104	CAGGCTTCCTCATCAACACC	TGTCTCTATGGCTGTGCAGG

Published in

Results

Paul Fisher

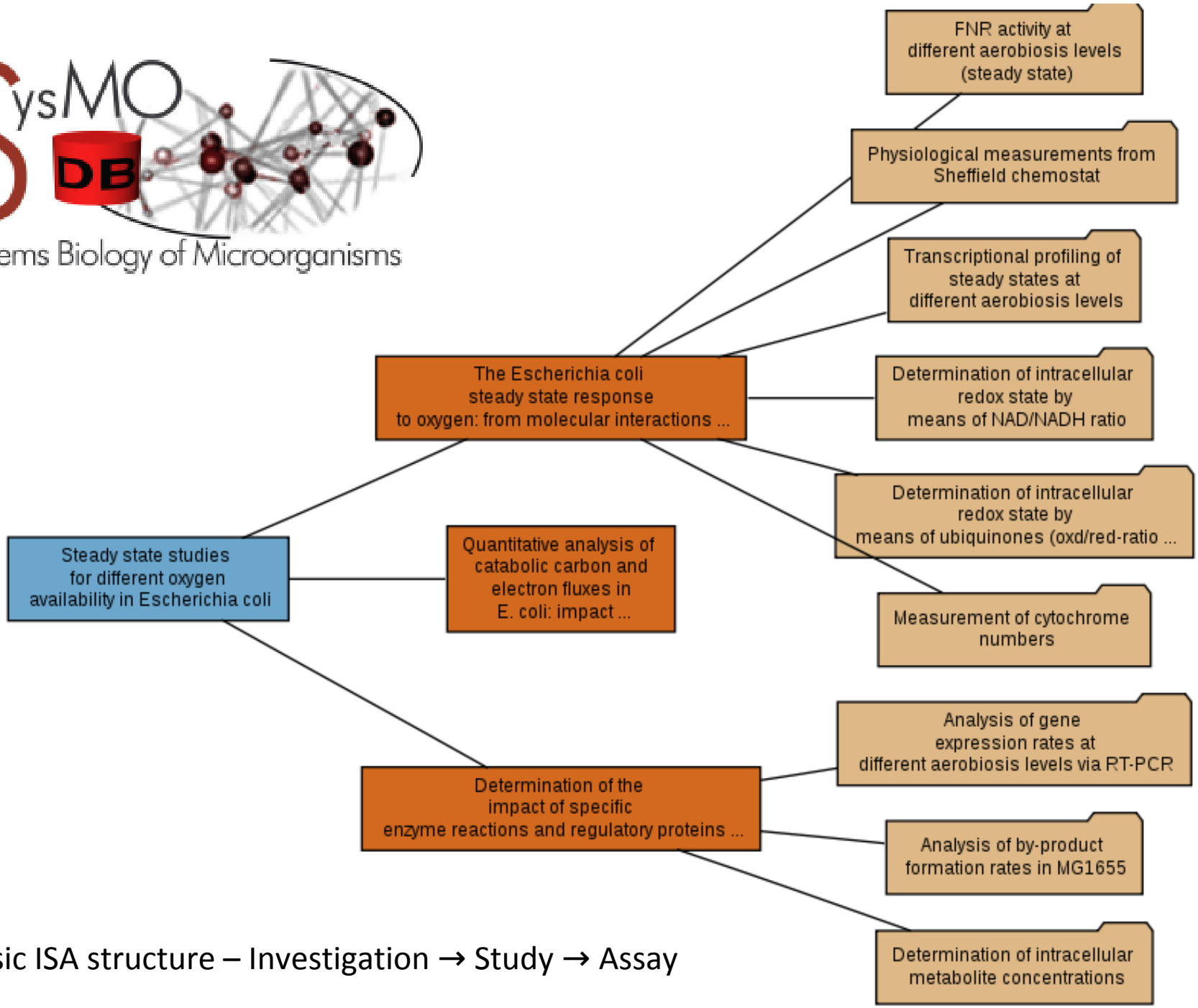
Example Investigation. Contains multiple Studies, Assays, and Assets (SOPs,Models,Datafiles)



Stuart Owen

SysMO DB

Systems Biology of Microorganisms



Basic ISA structure – Investigation → Study → Assay

The Six Rs of Research Object Behaviours

Research Objects enable data-intensive research to be:

1. Replayable – go back and see what happened
2. Repeatable – run the experiment again
3. Reproducible – independent expt to reproduce
4. Reusable – use as part of new experiments
5. Repurposeable – reuse the pieces in new expt
6. Reliable – robust under automation
7. Referenceable – citable and traceable

http://blog.openwetware.org/deroure/?p=56

Stereotypes

- Publication Object
  - Record of Activity
  - Credit/attribution
- Live Object
  - RO as work in progress
  - Up to date references to appropriate resource
- Archived Object
  - RO as a record of what happened
  - Curated, “fossilised”, immutable aggregation

Graceful Degradation

Research Object services are able to consume Research Objects without necessarily understanding or processing all of their content

- View Object
  - Named Graphs for LD
- Exposing Object
  - Standardised wrapper around data sources
- Method Object
  - RO as protocol

Sean Bechhofer

SALAMI

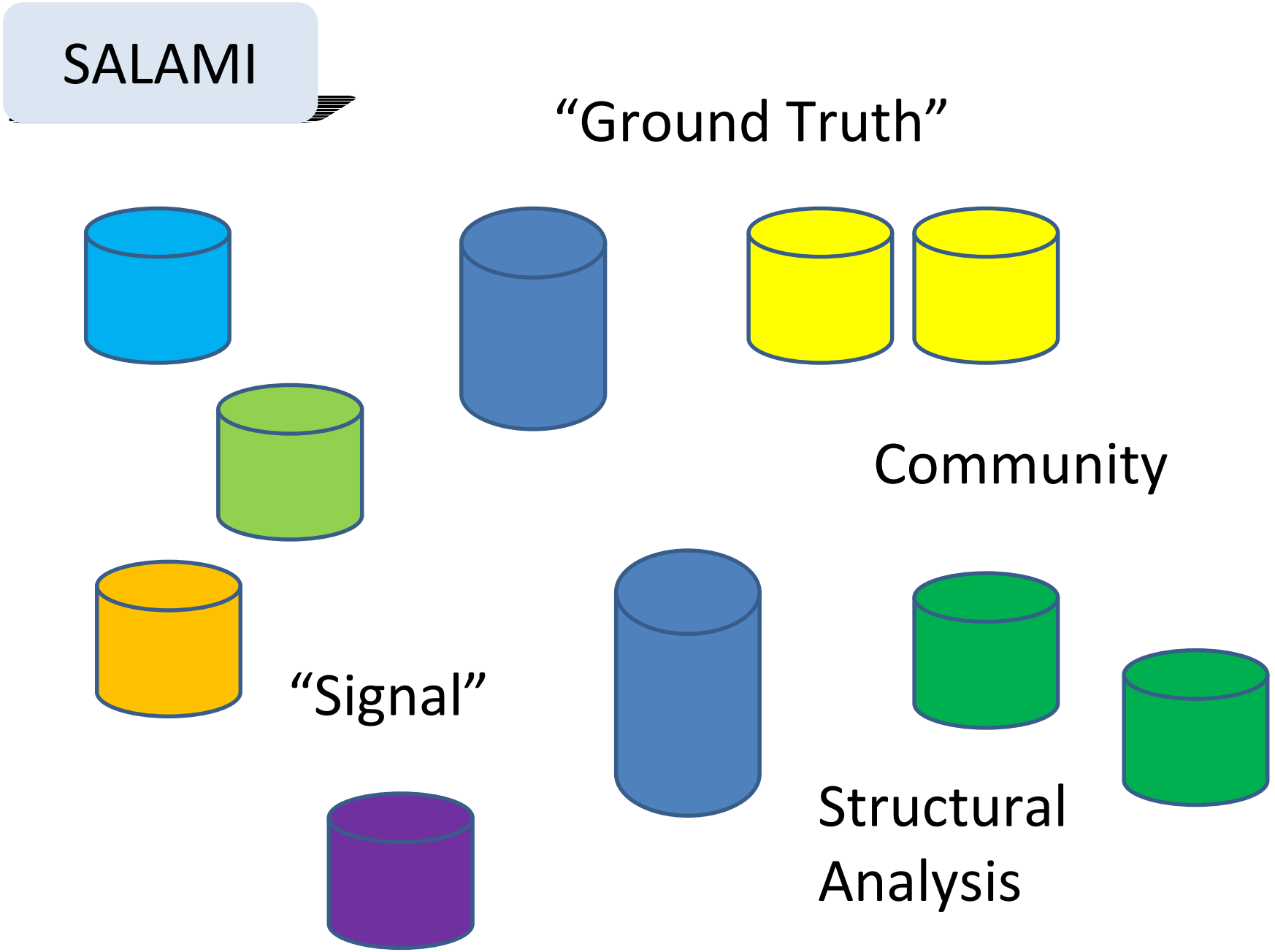
Generating a musicological resource using Internet Archive + Music Info Retrieval Algorithms + Supercomputer + Crowdsourced ground truth

Structural Analysis of Large Amounts of Music Information (SALAMI)

A Paradigm Shift in Music Structural Analysis

We are proposing an innovative and ambitious computational musicology project called Structural Analysis of Large Amounts of Data (SALAMI). To date, musical analysis has been conducted by individuals and on a small scale. Our computational approach, combined with the huge volume of data now available from such source as the Internet Archive, will a) deliver a very substantive corpus of musical analyses in a common framework for use by music scholars, students and beyond; and, b) establish a methodology and tooling which will enable others to add to this in the future and to broaden the application of the techniques we establish. A resource of SALAMI's magnitude empowers musicologists to approach their work in a new and different way, starting with the data, and to ask research questions that have not been possible before.

SALAMI will analyse 23,000 hours of digitised music to build a resource for musicologists, based on the music's underlying structure. It will use thousands of supercomputing hours, donated by the National Center for Supercomputing Applications (NCSA), and draw on a bewildering range of music from the Internet Archive –from A Capella to Zydeco, Appalachia to Zambia, and Medieval to Post-Modern.


SALAMI
“Ground Truth”
Community
“Signal”
Structural
Analysis

How Country is my Country?

1) Use SPARQL to generate a collection of signal
2) Publish that collection
3) Our local signal repository has copies of the actual signal, and publishes sub-graphs of linked data asserting what those signals are of (using the URI for that track/record etc.)
4) The workflow performing the feature extraction combines (2) and (3) when fetching the signal for feature extraction and classification, and persists the URI for the signal artefact (track/record etc.)
5) The results are published (e.g. of genre classification) and reference that URI

Kevin Page

Find all artists and show their countries

```
PREFIX geo:
   <http://www.geonames.org/ontology#>
SELECT ?name ?country
WHERE
{ ?artist
     a mo:MusicArtist;
     foaf:based_near ?place;
     foaf:name ?name.
  ?place
     geo:inCountry ?country }
ORDER BY ?name
```

Find all records by artists from France

```
PREFIX geo: <http://www.geonames.org/ontology#>
SELECT DISTINCT ?record
WHERE
{ ?artist
    a mo:MusicArtist;
    foaf:name ?name;
    foaf:based_near ?place.
  ?place
    geo:inCountry
    <http://www.geonames.org/countries/#FR>.
  ?record
    a mo:Record;
    foaf:maker ?artist }
ORDER BY ?record
```

Find all tracks from records by artists from France

```
PREFIX geo: <http://www.geonames.org/ontology#>
SELECT DISTINCT ?track
WHERE
{ ?artist
    a mo:MusicArtist;
    foaf:name ?name;
    foaf:based_near ?place.
  ?place
    geo:inCountry <http://www.geonames.org/countries/#FR>.
  ?record
    a mo:Record;
    foaf:maker ?artist;
    mo:track ?track }
ORDER BY ?track
```

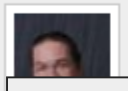
Taverna 2 beta

Text search within sparql point (v1)

View

Download (v1)

Original Uploader



Created: 19/01/09 @ 16:37:54

Credits: Francois Belleau

License: Creative Commons Attribution-Share Alike 3.0 License

This workflow make possible full text search within different sparql

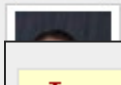
Taverna 2 beta

Sesame triplestore loader from a dereferenced URI (v2)

View

Download (v2)

Original Uploader



Created: 17/02/09 @ 05:11:34 | **Last updated:** 19/02/09 @ 05:03:09

Credits: Francois Belleau

License: Creative Commons Attribution-Share Alike 3.0 License

Taverna 2 beta

Author's collaborators according to pubmed (v1)

View

Download (v1)

Original Uploader

Francois Belleau

Created: 20/01/09 @ 02:22:32 | **Last updated:** 20/01/09 @ 03:41:26

Credits: Francois Belleau

License: Creative Commons Attribution-Share Alike 3.0 License

```
SELECT distinct ?s2 FROM <http://atlas.bio2rdf.org/sparql>
WHERE {  ?s1 ?p1 ?o1 .  ?o1 bif:contains "author" .  ?s2 ?p2 ?s1 .
 FILTER( regex(?s1, "pubmed") ) } followed by SELECT ?creator,
count() FROM <http://localhost:8890/sparql> WHERE {  ?s1 ?p1 .
?s1 ?p2 ?o2 .  FILTER( regex(?o2, "author"))  ?s1 ?creator . }
ORDER BY DESC(count())
```

Rating: 3.0 / 5 (1 rating) | **Versions:** 1 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 0

Viewed internally: 40 times | **Downloaded internally:** 13 times

Tags (4):

bio2rdf | pubmed | sparql | text_mining

Francois Belleau

Evolution of our research environment

1st Generation

Current practices of early adoptors of tools.

Characteris
their partic
tools, data

Traditional
publication
and links to

Provenance

Science is a
shift to emp

2nd Generation

Projects delivering now

Some inst

Key chara
of tools, c

Contain s
reproduci

Provenan

New scier
opportun
investigat

Some exp

3rd Generation

The solutions we'll be delivering in 5 years

Characterised by global reuse of tools, data and methods across any discipline, and surfacing the right levels of complexity for the researcher.

Routine use.

Key characteristic is **radical sharing .**

Research is significantly data driven - plundering the backlog of data, results and methods.

Increasing automation and decision-support for the researcher - the VRE becomes assistive.

Provenance assists design.

Curation is autonomic and social.

Though this be madness, yet there is method in it*

Deluge of data => Deluge of methods to process it?

Recording, re-using and sharing methods:

- Supports reproducible science
- Enables interpretation & trust of results
- Supports re-use and re-purposing
- Shares know-how
- Builds capability to understand data

Methods should be first class citizens!

* Polonius in Hamlet

Linked Open Methods*

- How we share
  - We are co-evolving a social infrastructure for sharing
- What we share
  - In the future we'll be saying "Could I have a copy of your Research Object please?" (if we didn't pick it up from the tweet...)
- Current work
  - Comunity curation, expert curation, assisted curation
  - Emerging practice in automation over linked data
  - Boundaries and guarantees: "the Web – particle duality"

* Sean Bechhofer

Repositories & Linked Data

- Linked Data community has guidelines and tooling for *production*
- Production practice will improve as consumption increases
  - e.g. Discovery
  - e.g. Versioning
- Issues of authority, licence, governance and curation are perhaps best addressed by the open repository community
- Balancing freshness with persistence

Contact

David De Roure

dder@ecs.soton.ac.uk

Carole Goble

carole.goble@manchester.ac.uk

Visit **wiki.myexperiment.org**

The Team

Sergejs Aleksejevs Mark Borkum Sean Bechhofer Jiten Bhagat Simon Coles Don Cruickshank Cat De Roure Paul Fisher Jeremy Frey Matt Gamble Duncan Hull Kumar Kollara Peter Li Ravi Madduri Danius Michaelides Paolo Missier David Newman Cameron Neylon Stuart Owen Kevin Page Rob Procter Marco Roos Stian Soiland Shoaib Sufi Mannie Tagarira Andrea Wiggins Alan Williams Katy Wolstencroft Tom Eveleigh June Finch Antoon Goderis Andrew Harrison Matt Lee Yuwei Lin Kurt Mueller Savas Parastatidis Meik Poschen Marcus Ramsden Ian Taylor Alexander Voss David Withers Ed Zaluska

Funders

- JISC Virtual Research Environments and Repositories programmes
- EPSRC myGrid and e-Research South platform awards
- Microsoft Research Technical Computing Initiative
- Andrew W. Mellon Foundation

Publications

http://wiki.myexperiment.org/index.php/Papers

- De Roure, D., Goble, C. and Stevens, R. (2009) “The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflows,” Future Generation Computer Systems 25, pp. 561-567.
- Goble, C.A., Bhagat, J., Aleksejevs, S., Cruickshank, D., Michaelides, D., Newman, D., Borkum, M., Bechhofer, S., Roos, M., Li, P., and De Roure, D.: myExperiment: a repository and social network for the sharing of bioinformatics workflows, Nucl. Acids Res., 2010. doi:10.1093/nar/gkq429
- De Roure, D. and Goble, C. (2009) "Software Design for Empowering Scientists," IEEE Software, vol. 26, no. 1, pp. 88-95, January/February 2009.
- Newman, D.R., Bechhofer, S. and De Roure, D. (2009) “myExperiment: An ontology for e-Research,” Workshop on Semantic Web Applications in Scientific Discourse at 8th International Semantic Web Conference (ISWC 2009), Washington DC, October 2009.
- Bechhofer, S., De Roure, D., Gamble, M., Goble, C. and Buchan, I. (2010) Research Objects: Towards Exchange and Reuse of Digital Knowledge. In: *The Future of the Web for Collaborative Science (FWCS 2010)*, April 2010, Raleigh, NC, USA.


myexperiment


UNIVERSITY OF
Southampton

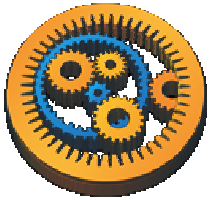

MANCHESTER
1824


JISC


myGrid


e-ResearcH
SOUTH


EPSRC

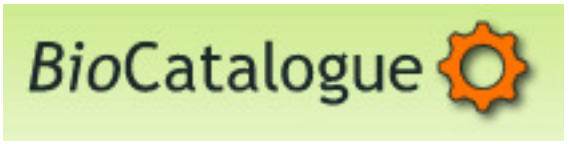


Taverna


Microsoft
Research


open middleware
infrastructure
institute uk


SysMO
DB
Systems Biology of Microorganisms


BioCatalogue


MethodBox