The University of Manchester

MANCHESTER 1824

myGrid

myexperiment

Taverna

Scientific Workflow Management System

BioCatalogue beta

"The Life Science Web Services Registry"

Taverna, Biocatalogue, and myExperiment: a three-legged foundation for effective collaboration in E-science

A collaborative talk by Paolo Missier

Information Management Group

School of Computer Science, University of Manchester, UK

with additional material kindly shared by:

Prof. Dave DeRoure and David Newman, University of Southampton

Prof. Carole Goble and the e-Labs design group, University of Manchester

Outline

- Vision
  - Embracing collaboration in experimental science
  - Sharing is happening, let's do it right!

- Technology
  - Current architectural components
    - Taverna, myExperiment, Biocatalogue
  - Available vocabularies, formats that we use
    - SIOC, SKOS, ORE,...

- Open issues (Research??)
  - Sharing in the large: social and technical issues
  - Producing value, avoiding pollution

On sharing in science

On sharing in science

- Traditionally protective culture
  - no data sharing prior to publication
  - raw data is public, results are private

"A biologist would rather share their toothbrush than their gene name"

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

On sharing in science

- Traditionally protective culture
  - no data sharing prior to publication
  - raw data is public, results are private

"A biologist would rather share their toothbrush than their gene name"

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

- Scope
  - Collaboration in the small vs in the wild
    - control over sharing
    - trust management
  - Experimental, information-centric, computational science
    - experimental processes encoded as workflows

Momentum on sharing and collaboration

- Ongoing debate in many communities
  - Clinical trials [1]
  - Earth Sciences -- ESIP - data preservation / stewardship, 2009
  - Long established in some communities - Atmospheric sciences, 1998 [2]

- Science Commons recommendations for Open Science
  - Open Science recommendations from Science Commons (July 2008) [link]

- Publishers' requirements for data and methods disclosure
  - for archiving, experiment repeatability

[1] Iain Hrynaszkiewicz, D. G Altman, Trials. 2009 Mar 18;10:17 [link]

[2] Strebel DE, Landis DR, Huemmrich KF, Newcomer JA, Meeson BW: The FIFE Data Publication Experiment. Journal of the Atmospheric Sciences 1998, 55:1277-1283

Emerging community standards for information exchange (bioinformatics)

MIBBI: Minimum Information for Biological and Biomedical Investigations

MIAME: Minimum Information About a Microarray Experiment

MIAPE: The Minimum Information About a Proteomics Experiment (PSI - HUPO) [1]

[1] Taylor CF, Paton NW, Lilley KS, et al. The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol (2007) 25:887–93 [link]

Incentives to sharing

Special issue of Nature on Data Sharing (Sept. 2009)

The Toronto group: Toronto International Data Release Workshop Authors, Nature 461, 168–169 (2009)

Prepublication data sharing:
Nature 461, 168-170 (10 September 2009) | doi:10.1038/461168a; Published online 9 September 2009 http://www.nature.com/news/specials/datasharing/index.html

Incentives to sharing

Special issue of Nature on Data Sharing (Sept. 2009)

"The draft human genome sequence was the highest profile dataset rapidly released before publication [...] This experience ultimately demonstrated that the broad and early availability of sequence data greatly benefited life sciences research by leading to many new insights and discoveries, including new information on 30 disease genes published prior to the draft sequence."

The Toronto group: Toronto International Data Release Workshop Authors, Nature 461, 168–169 (2009)

Prepublication data sharing:
Nature 461, 168-170 (10 September 2009) | doi:10.1038/461168a; Published online 9 September 2009 http://www.nature.com/news/specials/datasharing/index.html

Incentives to sharing

Special issue of Nature on Data Sharing (Sept. 2009)

- timeliness requires rapid sharing
- repurposing
- data quality control

> "The draft human genome sequence was the highest profile dataset rapidly released before publication [...] This experience ultimately demonstrated that the broad and early availability of sequence data greatly benefited life sciences research by leading to many new insights and discoveries, including new information on 30 disease genes published prior to the draft sequence."

The Toronto group: Toronto International Data Release Workshop Authors, Nature 461, 168–169 (2009)

Prepublication data sharing:
Nature 461, 168-170 (10 September 2009) | doi:10.1038/461168a; Published online 9 September 2009 http://www.nature.com/news/specials/datasharing/index.html

Incentives to sharing

Special issue of Nature on Data Sharing (Sept. 2009)

- timeliness requires rapid sharing
- repurposing
- data quality control

One of the lessons from the Human Genome Project (HGP) was the recognition that making data broadly available prior to publication can be profoundly valuable to the scientific enterprise and lead to public benefits.

This is particularly the case when there is a community of scientists that can productively use the data quickly — beyond what the data producers could do themselves in a similar time period, and sometimes for scientific purposes outside the original goals of the project.

The Toronto group: Toronto International Data Release Workshop Authors, Nature 461, 168–169 (2009)

Prepublication data sharing:
Nature 461, 168-170 (10 September 2009) | doi:10.1038/461168a; Published online 9 September 2009 http://www.nature.com/news/specials/datasharing/index.html

Incentives to sharing

Special issue of Nature on Data Sharing (Sept. 2009)

- timeliness requires rapid sharing
- repurposing
- data quality control

One of the lessons from the Human Genome Project (HGP) was the

"The draft human genome sequence was the highest profile dataset

To date, many papers have been published by third parties reporting research findings enabled by data sets released before publication. The experiences shared in Toronto suggest that these have rarely affected subsequent publications authored by the data producer

producers could distinguish themselves in a similar time period, and sometimes for scientific purposes outside the original goals of the project.

The Toronto group: Toronto International Data Release Workshop Authors, Nature 461, 168–169 (2009)

Prepublication data sharing:
Nature 461, 168-170 (10 September 2009) | doi:10.1038/461168a; Published online 9 September 2009 http://www.nature.com/news/specials/datasharing/index.html

Incentives to sharing

Special issue of Nature on Data Sharing (Sept. 2009)

- timeliness requires rapid sharing
- repurposing
- data quality control

the data users should carefully read the source information, including any marker papers, associated with a released data set. Data analysts should pay particular attention to any caveats about data quality, because rapidly released data are often unstable, in that they may not yet have been subjected to full quality control and so may change.

project.

The Toronto group: Toronto International Data Release Workshop Authors, Nature 461, 168–169 (2009)
Prepublication data sharing: Nature 461, 168-170 (10 September 2009) | doi:10.1038/461168a; Published online 9 September 2009 http://www.nature.com/news/specials/datasharing/index.html

The social process of Science

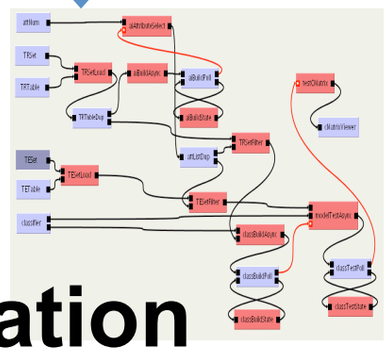
Virtual Learning Environment

Undergraduate Students

Digital Libraries

scientists

Graduate Students



Reprints

Peer-Reviewed Journal & Conference Papers

experimentation

Technical Reports

Preprints & Metadata

Repositories

Local Web

Certified Experimental Results & Analyses

Data, Metadata
Provenance
Workflows
Ontologies

The social process of Science 2.0

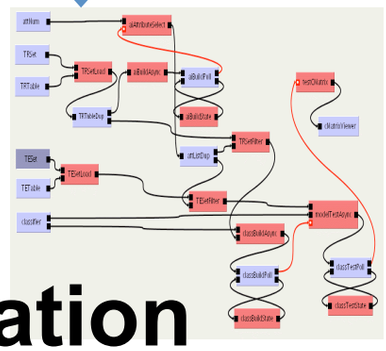
Virtual Learning Environment

Undergraduate Students

Digital Libraries

scientists

Graduate Students



Reprints

Peer-Reviewed Journal & Conference Papers

Preprints & Metadata

Technical Reports

experimentation

Repositories

Local Web

Certified Experimental Results & Analyses

Data, Metadata
Provenance
Workflows
Ontologies

The social process of Science 2.0

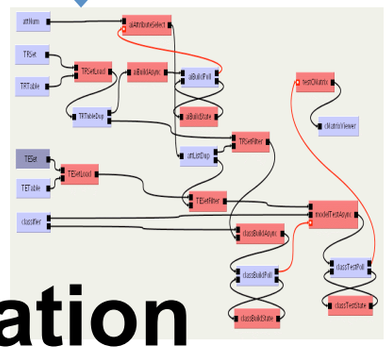
Virtual Learning Environment

Undergraduate Students

Digital Libraries

scientists

Graduate Students



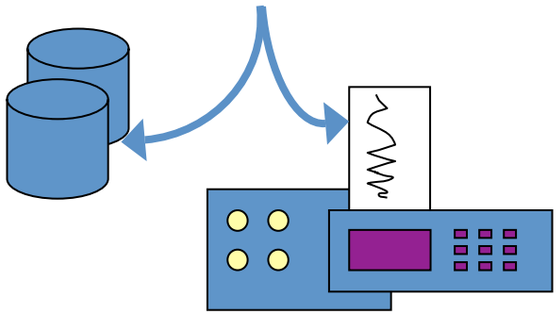
Reprints

Peer-Reviewed Journal & Conference Papers

Preprints & Metadata

Technical Reports

experimentation



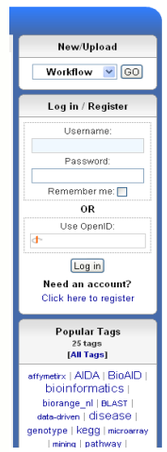
Repositories

Local Web

Certified Experimental Results & Analyses

Data, Metadata
Provenance
Workflows
Ontologies

myexperiment



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

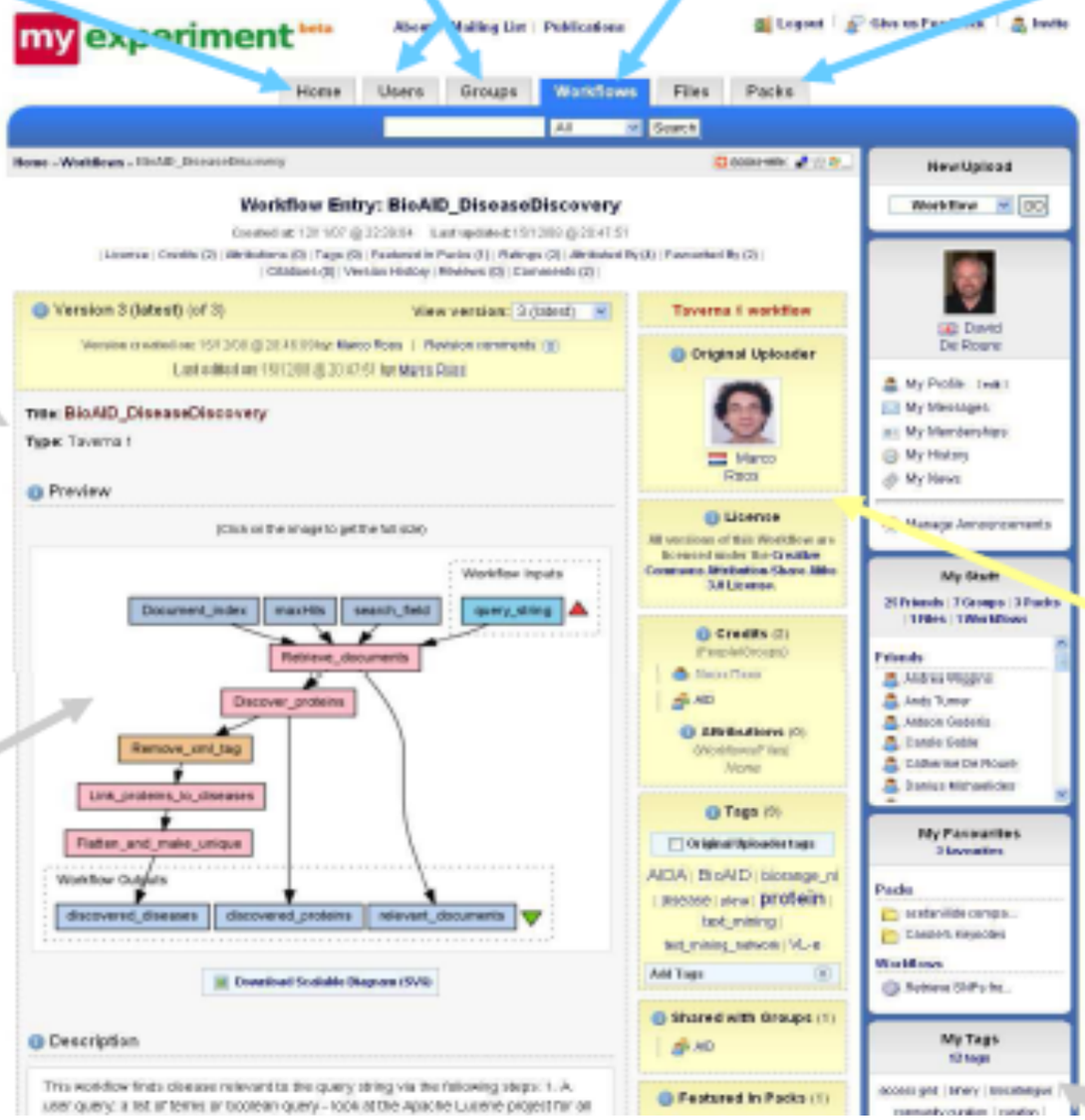
See the latest and most popular workflows: discover, view, download, run, tag and rate. Upload your workflows.

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.



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.

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 me

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

JISC EPSRC

Microsoft

my experiment

Version 1 (of 1)

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

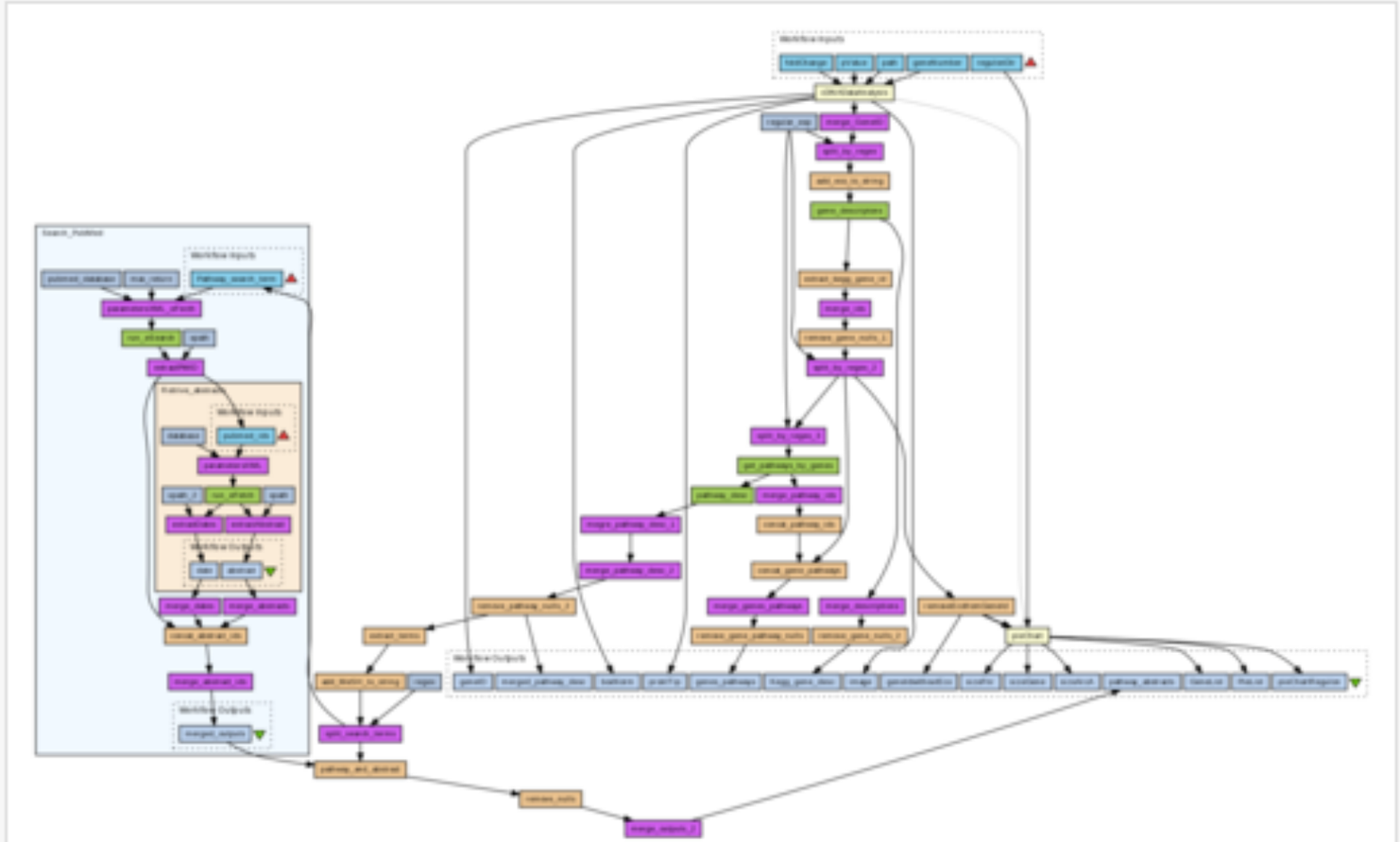
Last edited on: 12/05/08 @ 09:01:37 **by:** Saeedeh

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

All versions of this Workflow are licensed under the **Creative Commons Attribution-Share Alike 3.0 License.**

Credits (2)

(People/Groups)

- Saeedeh
- Paul Fisher

Attributions (1)

(Workflows/Files)

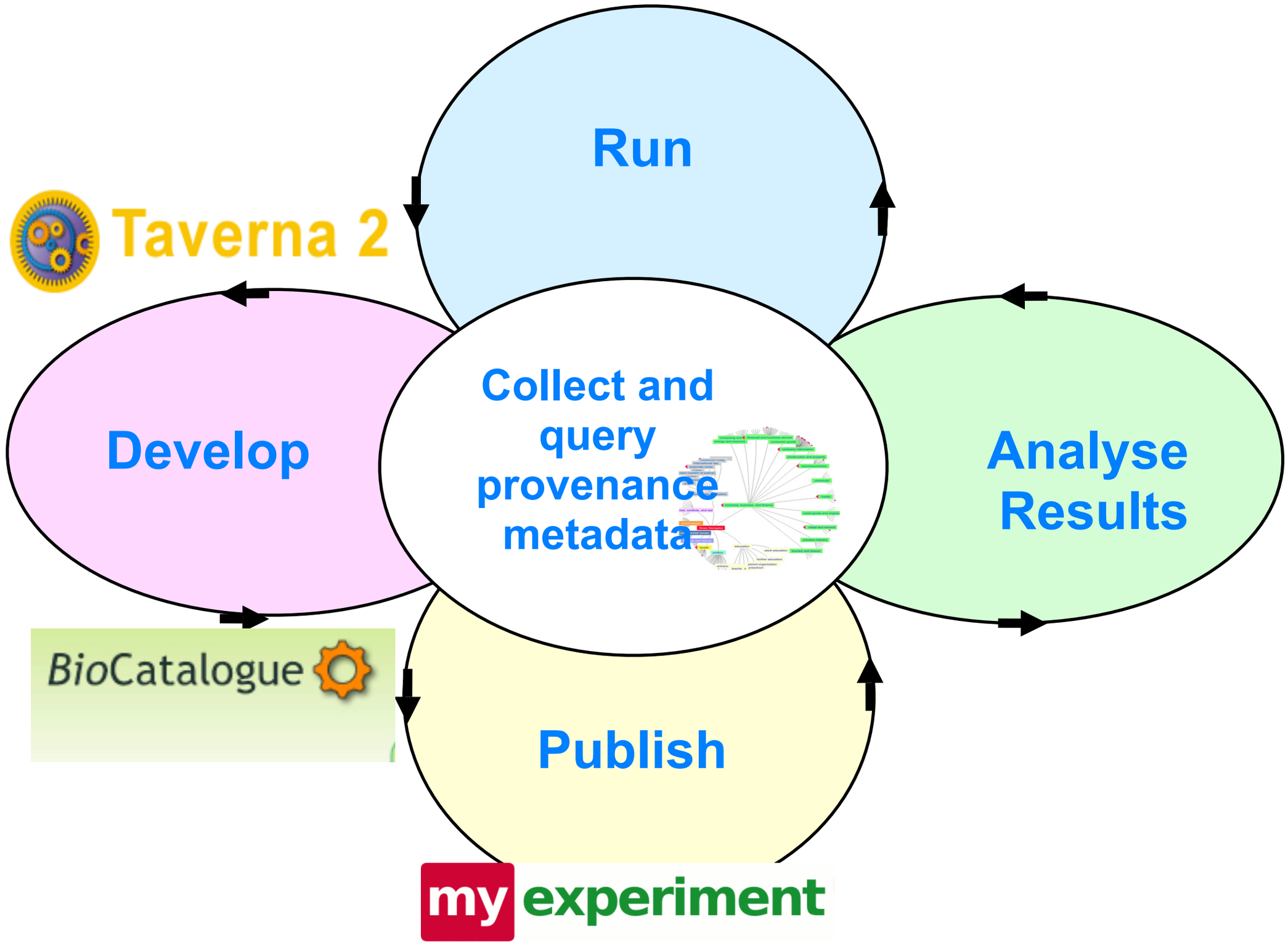
- HUMAN Microarray CEL file to candidate pathways

Tags (5)

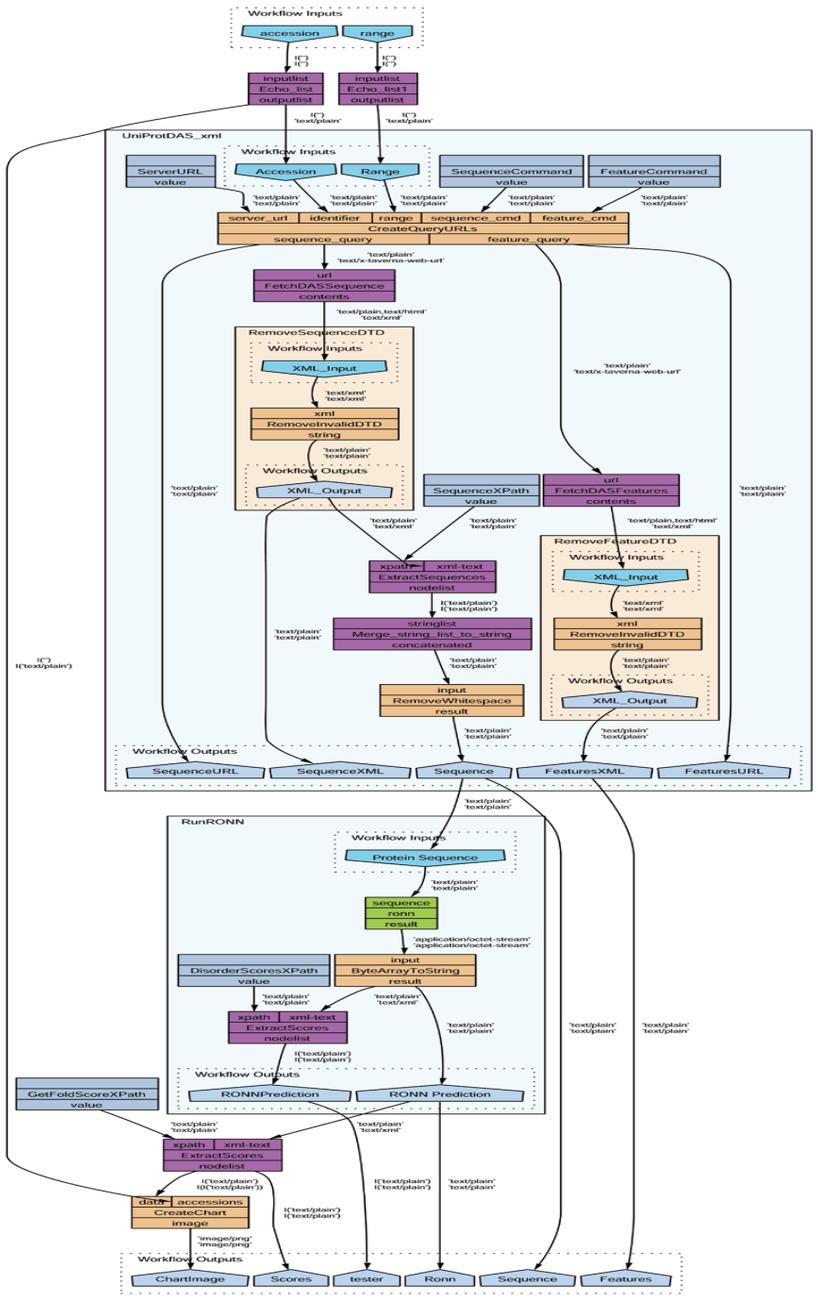
Original Uploader tags

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

Workflow-based experimentation lifecycle



Workflows: *E. Science laboris*



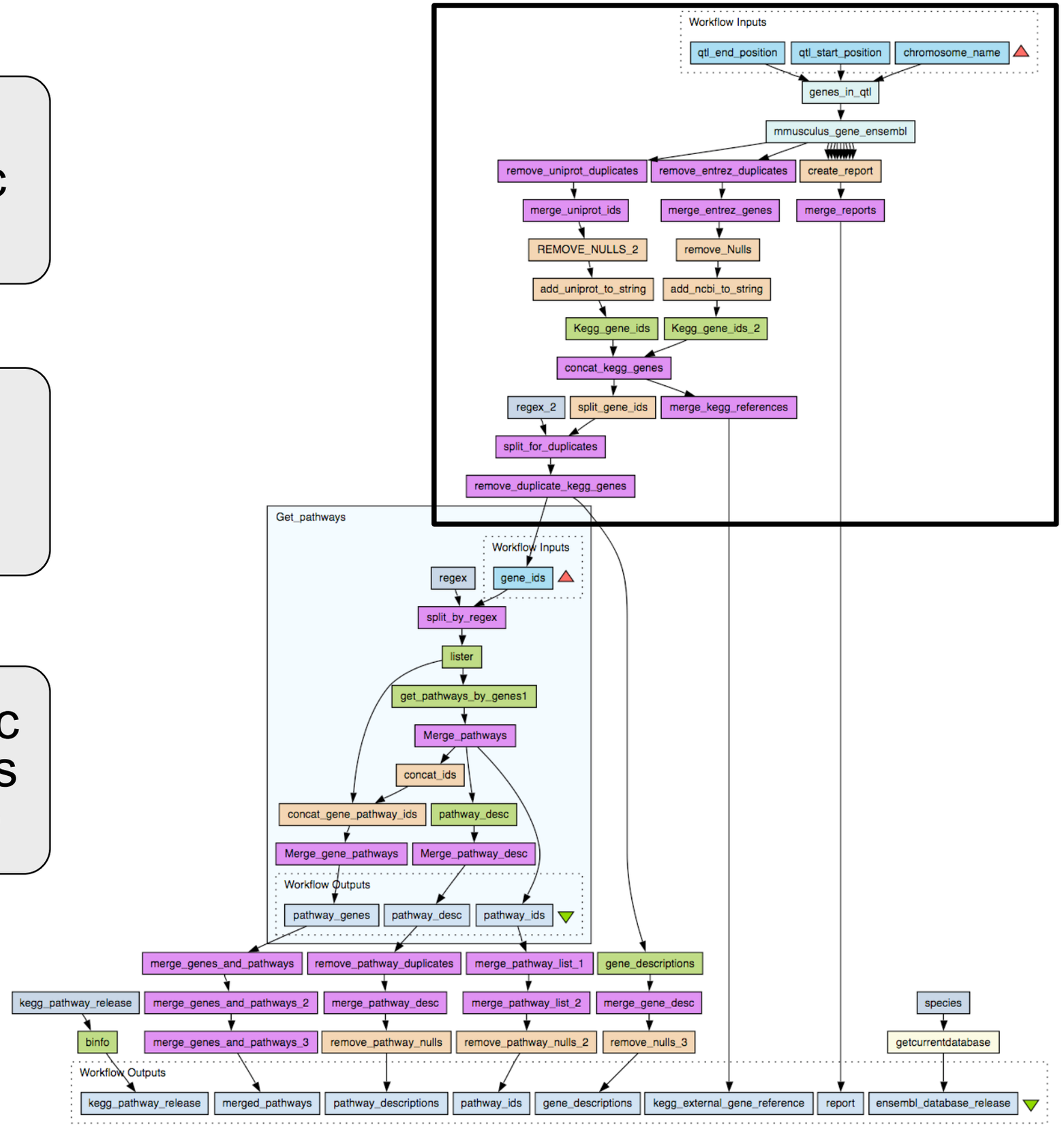
- Pipeline processing
- Automated processing
- Repetitive and mundane boring stuff made easier, reliable and adaptable.
- Shield interoperability horror
- Trackable results
- Agile software development
- Big science, small science & collaborative science

Workflow as data integrator

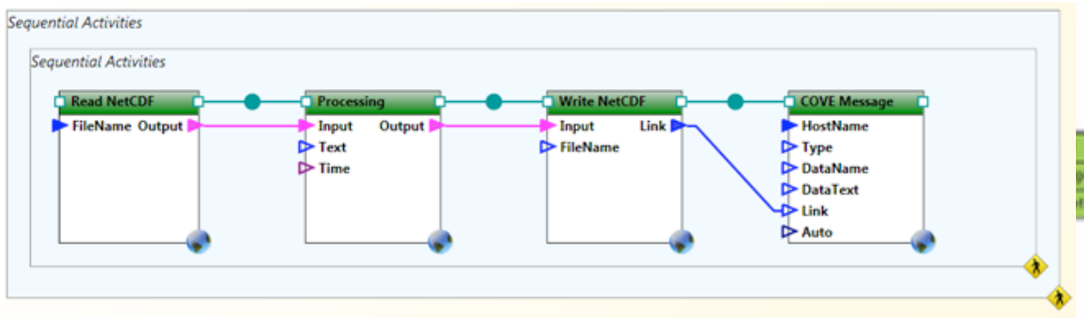
QTL genomic regions

genes in QTL

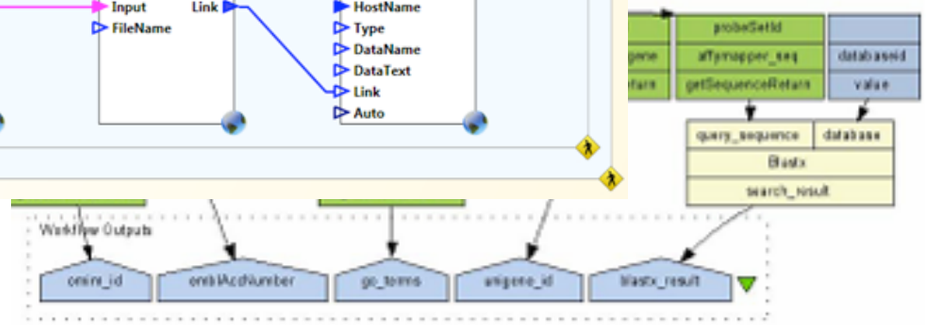
metabolic pathways (KEGG)

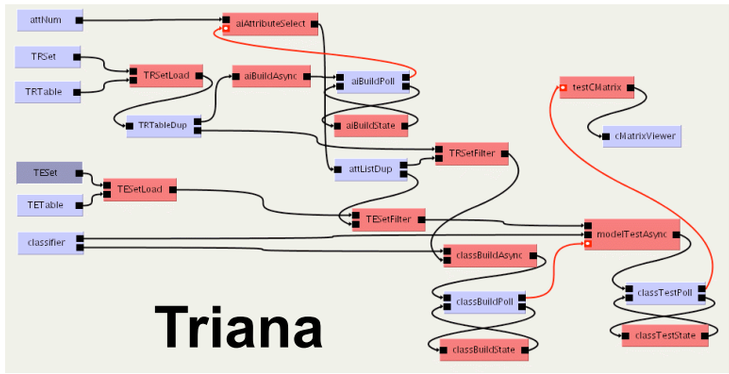


We are not alone...

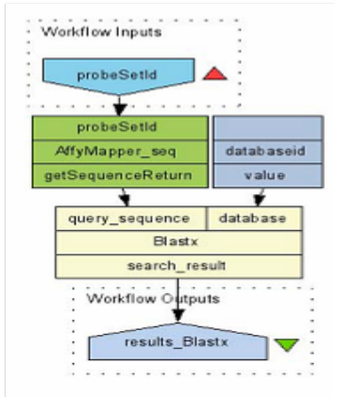


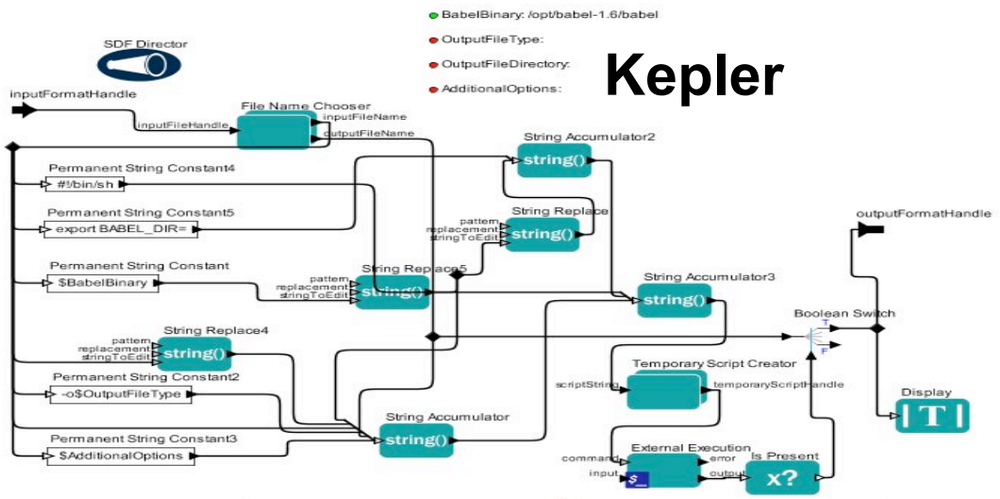
Trident



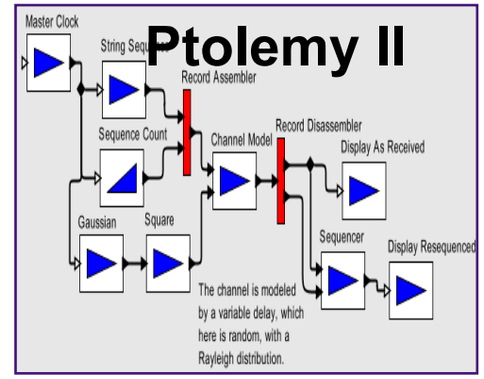


Triana



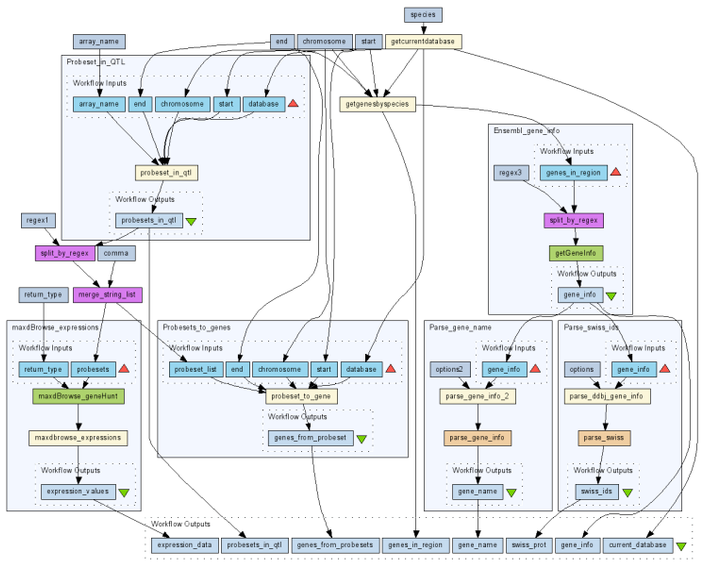


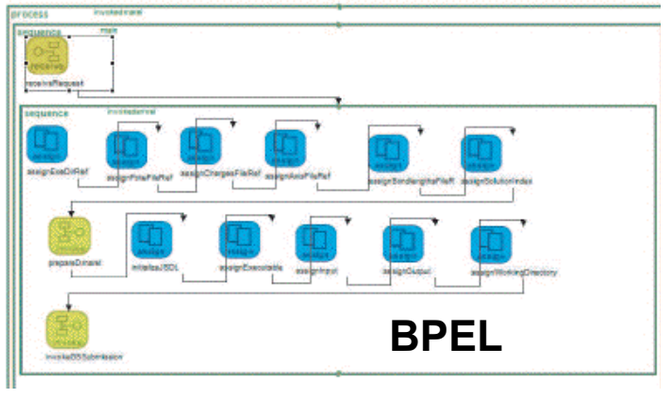
Kepler



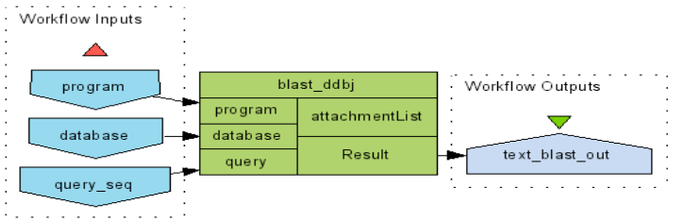
Ptolemy II

Taverna

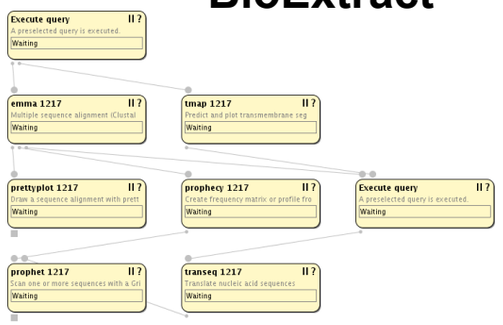




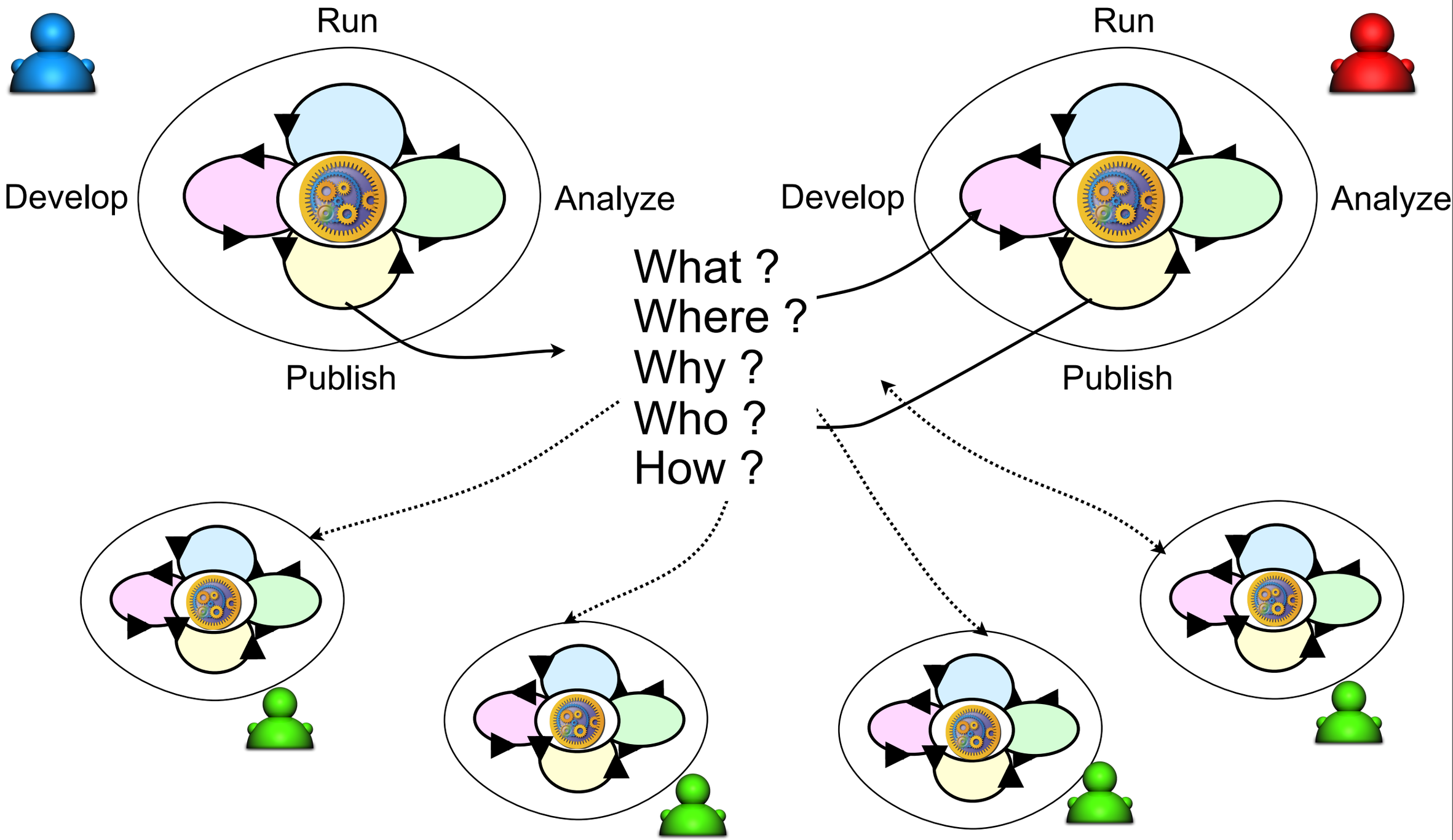
BPEL



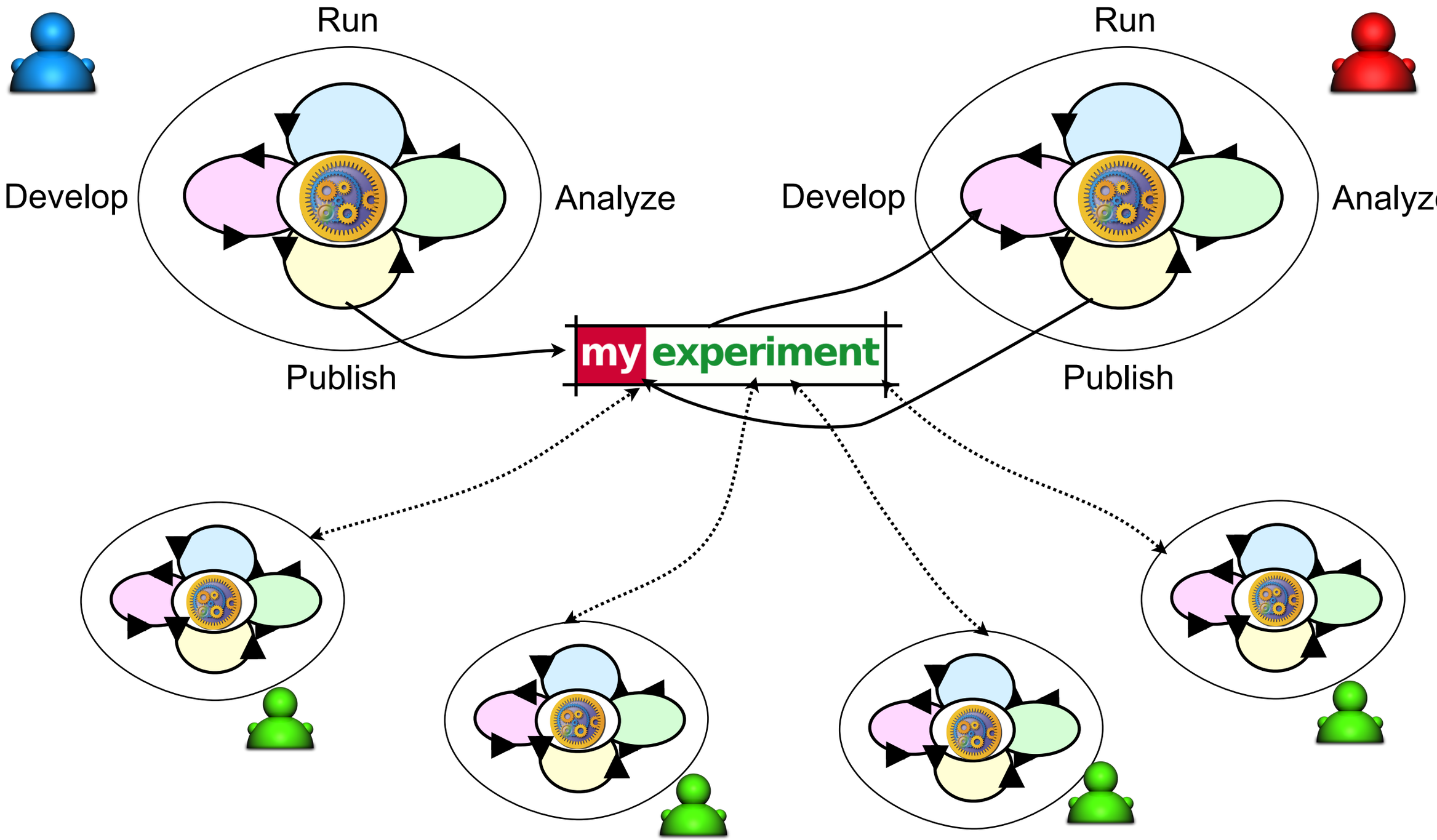
BioExtract



Scaling up along the social dimension



Scaling up along the social dimension



Reuse, Recycling, Repurposing

- Paul writes workflows for identifying biological pathways implicated in resistance to Trypanosomiasis in the mouse
- Paul meets Jo, who is from a different department.
- Jo is investigating intestinal Whipworm infection in the mouse.
- Jo reuses the same workflow, without change.
- Jo identifies biological pathways, and candidate genes, highlighting key tissue repair mechanisms linked to the ability of the mouse to expel the parasite
- A manual two year study by Jo had failed to identify these pathways, and their underlying genes

MyExperiment's open information model

- Entries described using an RDF data model
- Entries available through a REST API
- Using a combination of home-grown and standard ontologies
  - http://rdf.myexperiment.org/ontologies/

MyExperiment's open information model

- Entries described using an RDF data model
- Entries available through a REST API
- Using a combination of home-grown and standard ontologies
  - http://rdf.myexperiment.org/ontologies/

<mecontrib:Workflow rdf:about="http://rdf.myexperiment.org/Workflow/384">
<dcterms:title>What's On Next</dcterms:title>
<dcterms:description>Queries the BBC to find out what's on next...
</dcterms:description>
<sioc:has_owner rdf:resource="http://rdf.myexperiment.org/User/438" />
<dcterms:created>2008-08-28T15:23:58Z</dcterms:created>
<mebase:filename>what_s_on_next_18085.xml</mebase:filename>
<ore:isDescribedBy rdf:resource="http://rdf.myexperiment.org/ResourceMap/Workflow/384" />
<mecontrib:preview rdf:datatype="&xsd;anyURI">http://www.myexperiment.org/workflow/image/384/what_s_on_next_18085_1.png</mecontrib:preview>

MyExperiment's open information model

- Entries described using an RDF data model
- Entries available through a REST API
- Using a combination of home-grown and standard ontologies
  - http://rdf.myexperiment.org/ontologies/

myExperiment / contributions module

<mecontrib:Workflow rdf:about="http://rdf.myexperiment.org/

<dcterms:title>What's On Next</dcterms:title>

<dcterms:description>Queries the BBC to find out what's on next...

</dcterms:description>

<sioc:has_owner rdf:resource="http://rdf.myexperiment.org/User/438" />

<dcterms:created>2008-08-28T15:23:58Z</dcterms:created>

<mebase:filename>what_s_on_next_18085.xml</mebase:filename>

<ore:isDescribedBy rdf:resource="http://rdf.myexperiment.org/ResourceMap/Workflow/384" />

<mecontrib:preview rdf:datatype="&xsd;anyURI">http://www.myexperiment.org/workflow/image/384/what_s_on_next_18085_1.png</mecontrib:preview>

MyExperiment's open information model

- Entries described using an RDF data model
- Entries available through a REST API
- Using a combination of home-grown and standard ontologies
  - http://rdf.myexperiment.org/ontologies/

myExperiment / contributions module

Dublin core

<mecontrib:Workflow rdf:about="http://rdf.myexperiment.org/

<dcterms:title>What's On Next</dcterms:title>

<dcterms:description>Queries the BBC to find out what's o

</dcterms:description>

<sioc:has_owner rdf:resource="http://rdf.myexperiment.org/User/438" />

<dcterms:created>2008-08-28T15:23:58Z</dcterms:created>

<mebase:filename>what_s_on_next_18085.xml</mebase:filename>

<ore:isDescribedBy rdf:resource="http://rdf.myexperiment.org/ResourceMap/Workflow/384" />

<mecontrib:preview rdf:datatype="&xsd;anyURI">http://www.myexperiment.org/workflow/image/384/what_s_on_next_18085_1.png</mecontrib:preview>

MyExperiment's open information model

- Entries described using an RDF data model
- Entries available through a REST API
- Using a combination of home-grown and standard ontologies
  - http://rdf.myexperiment.org/ontologies/

<mecontrib:Workflow rdf:about="http://rdf.myexperiment.org/

<dcterms:title>What's On Next</dcterms:title>

<dcterms:description>Queries the BBC to find out what's o

</dcterms:description>

<sioc:has_owner rdf:resource="http://rdf.myexperiment.org/User/438" />

<dcterms:created>2008-08-28T15:23:58Z</dcterms:created>

<mebase:filename>what_s_on_next_18085.xml</mebase:filename>

<ore:isDescribedBy rdf:resource="http://rdf.myexperiment.org/ResourceMap/Workflow/384" />

<mecontrib:preview rdf:datatype="&xsd;anyURI">http://www.myexperiment.org/workflow/image/384/what_s_on_next_18085_1.png</mecontrib:preview>

myExperiment / contributions module

Dublin core

SIOC ontology

MyExperiment's open information model

- Entries described using an RDF data model
- Entries available through a REST API
- Using a combination of home-grown and standard ontologies
  - http://rdf.myexperiment.org/ontologies/

```
<mecontrib:Workflow rdf:about="http://rdf.myexperiment.org/
   <dcterms:title>What&apos;s On Next</dcterms:title>
   <dcterms:description>Queries the BBC to find out what&apos;s o
   </dcterms:description>
   <sioc:has_owner rdf:resource="http://rdf.myexperiment.org/User/438" />
   <dcterms:created>2008-08-28T15:23:58Z</dcterms:created>
   <mebase:filename>what_s_on_next_18085.xml</mebase:filename>
   <ore:isDescribedBy rdf:resource="http://rdf.myexperiment.org/
ResourceMap/Workflow/384" />
   <mecontrib:preview rdf:datatype="&xsd;anyURI">http://
www.myexperiment.org/workflow/image/384/
what_s_on_next_18085_1.png</mecontrib:preview>
```

myExperiment / contributions module

Dublin core

SIOC ontology

myExperiment / base module

MyExperiment's open information model

- Entries described using an RDF data model
- Entries available through a REST API
- Using a combination of home-grown and standard ontologies
  - http://rdf.myexperiment.org/ontologies/

```
<mecontrib:Workflow rdf:about="http://rdf.myexperiment.org/
   <dcterms:title>What&apos;s On Next</dcterms:title>
   <dcterms:description>Queries the BBC to find out what&apos;s o
   </dcterms:description>
   <sioc:has_owner rdf:resource="http://rdf.myexperiment.org/User/438" />
   <dcterms:created>2008-08-28T15:23:58Z</dcterms:created>
   <mebase:filename>what_s_on_next_18085.xml</mebase:filename>
   <ore:isDescribedBy rdf:resource="http://rdf.myexperiment.org/
ResourceMap/Workflow/384" />
   <mecontrib:preview rdf:datatype="&xsd;anyURI">http://
www.myexperiment.org/workflow/image/384/
what_s_on_next_18085_1.png</mecontrib:preview>
```

myExperiment / contributions module

Dublin core

SIOC ontology

myExperiment / base module

ORE ontology

Attribution and derivation

- Keys to controlled sharing

```
<mebase:has-policy>
  <snarm:Policy rdf:about="http://rdf.myexperiment.org/Policy/733">
    <snarm:has-access>
      <snarm:RestrictedAccess
          rdf:about="http://rdf.myexperiment.org/Policy/733/Access/ViewUser438">
        <snarm:has-accesser rdf:resource="http://rdf.myexperiment.org/User/438"/>
        <snarm:has-access-type rdf:resource="&mespec;View"/>
      </snarm:RestrictedAccess>
    </snarm:has-access>
  </snarm:Policy>
</mebase:has-policy>
</rdf:RDF>
```

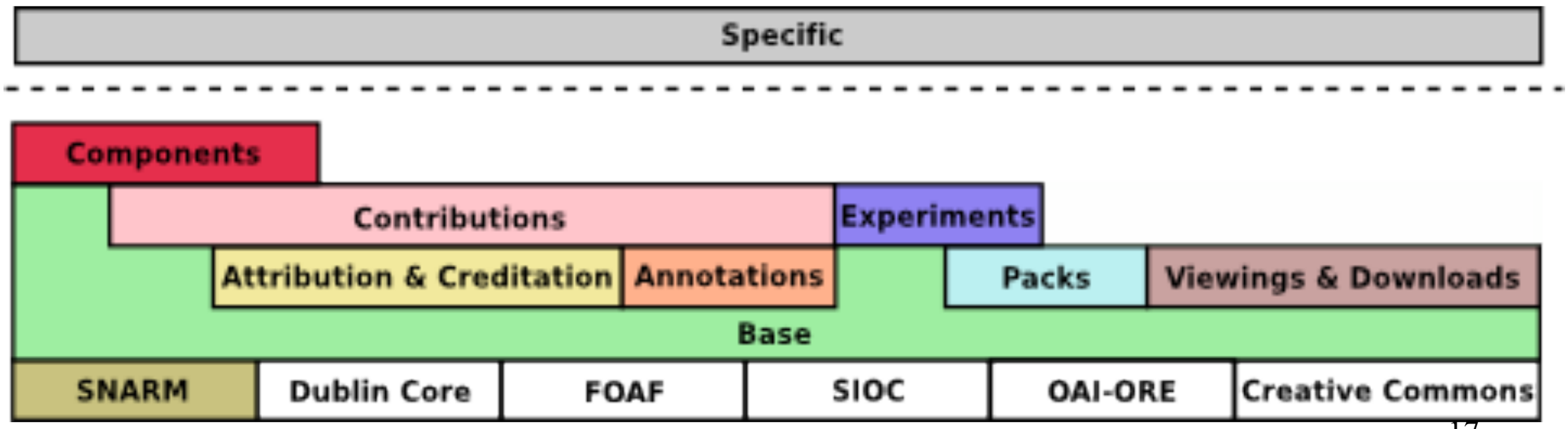
Simple Network Access Rights Management (SNARM)
Attribution: **David R. Newman**, myExperiment team, University of Southampton, UK

Attribution and derivation

- Keys to controlled sharing

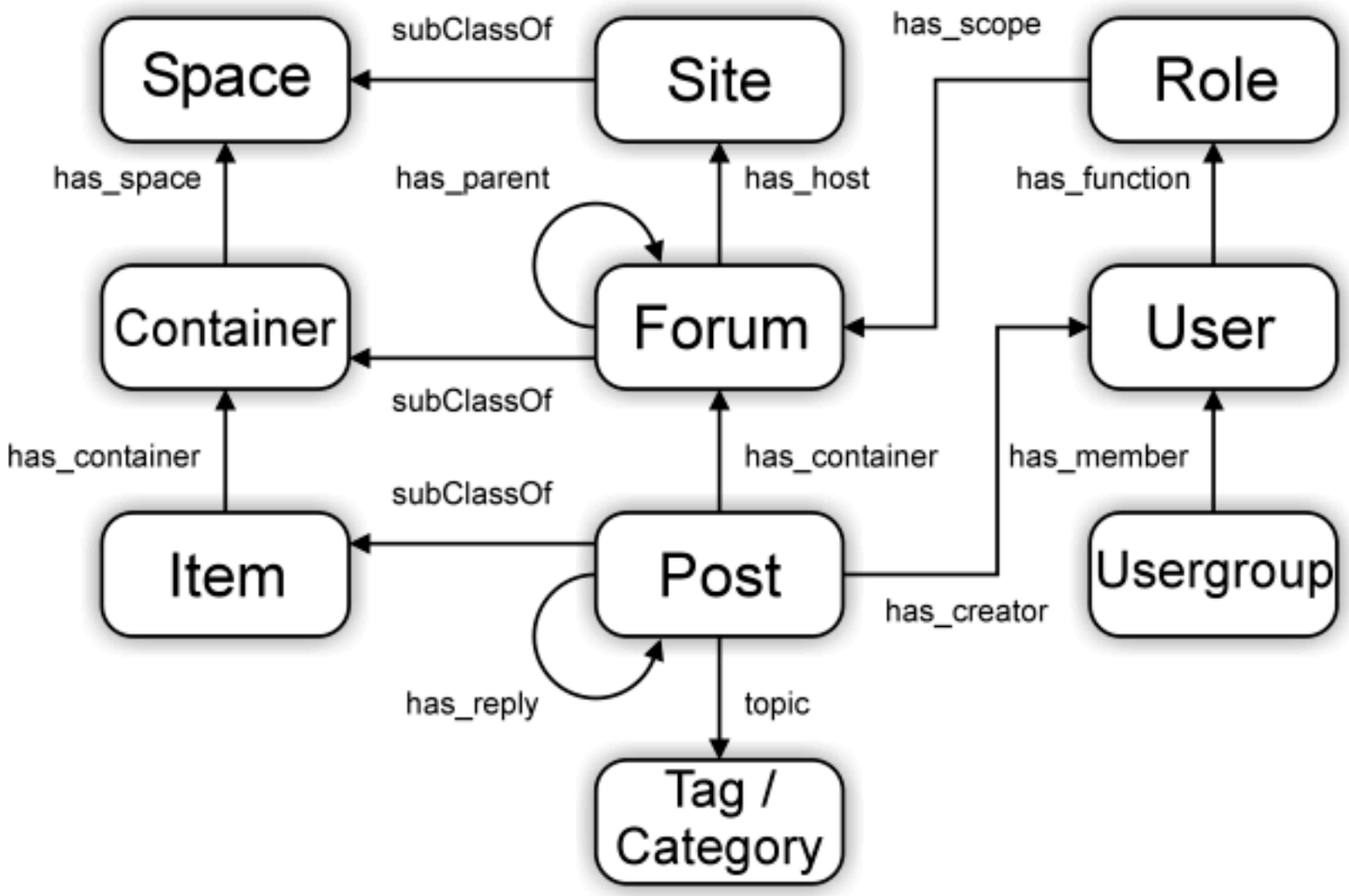
```
<mebase:has-policy>
  <snarm:Policy rdf:about="http://rdf.myexperiment.org/Policy/733">
    <snarm:has-access>
      <snarm:RestrictedAccess
          rdf:about="http://rdf.myexperiment.org/Policy/733/Access/ViewUser438">
        <snarm:has-accesser rdf:resource="http://rdf.myexperiment.org/User/438"/>
        <snarm:has-access-type rdf:resource="&mespec;View"/>
      </snarm:RestrictedAccess>
    </snarm:has-access>
  </snarm:Policy>
</mebase:has-policy>
</rdf:RDF>
```

Simple Network Access Rights Management (SNARM)
Attribution: **David R. Newman**, myExperiment team, University of Southampton, UK



SIOC: Semantically-Interlinked Online Communities

http://sioc-project.org/



Good tool support

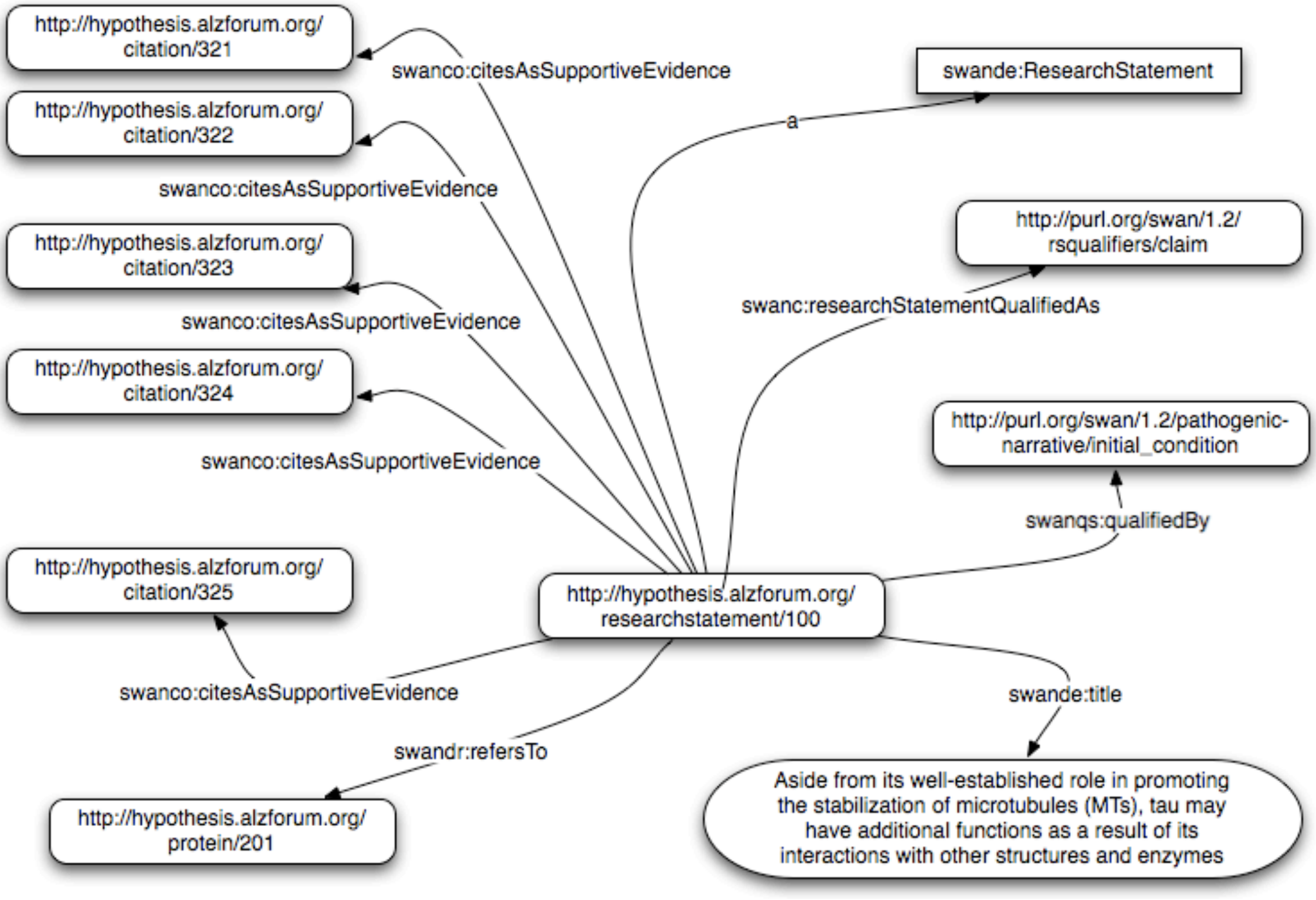
Commonly used in conjunction with the FOAF vocabulary

SWAN: Semantic Web Applications in Neuromedicine

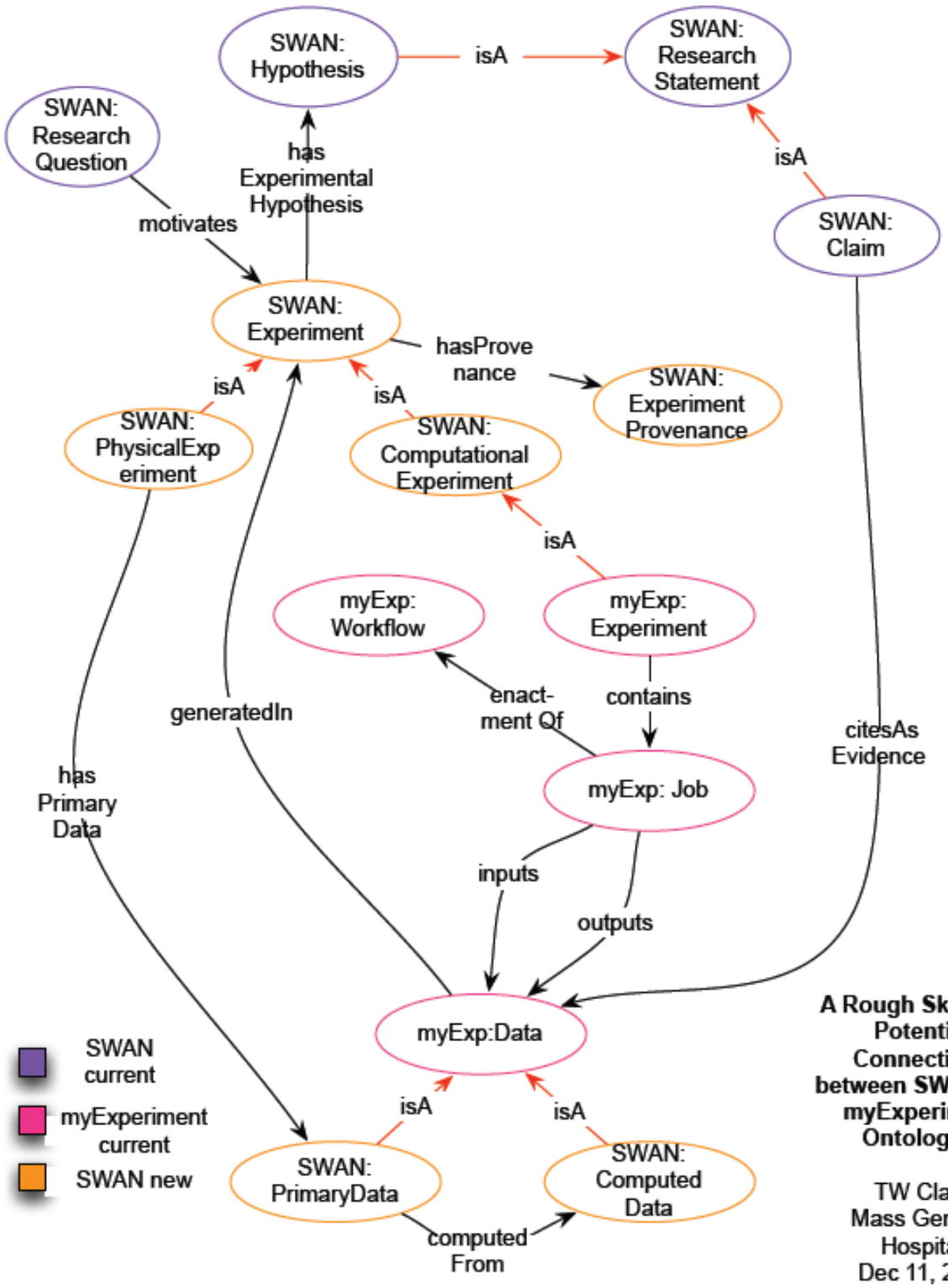
http://www.w3.org/2001/sw/hcls/notes/swan/

Goal: modelling scientific discourse

Massachusetts General Hospital / Harvard Medical School



SWAN and myExperiment



combining SWAN and the myExperiment ontology

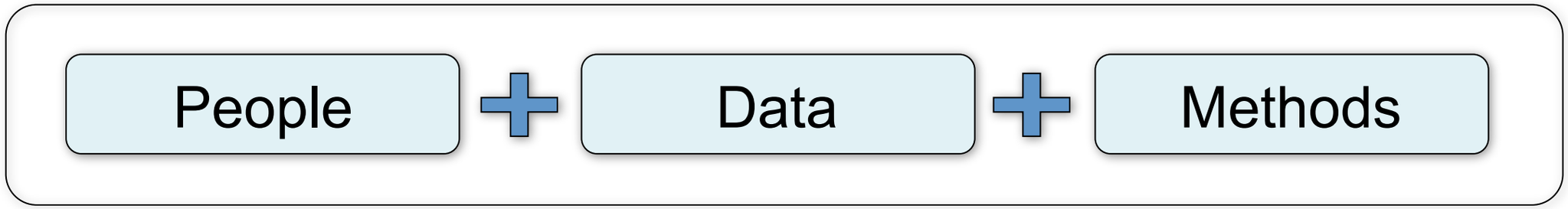
Tim Clark

Newman, D., Bechhofer, S. and De Roure, D. (2009) myExperiment: An ontology for e-Research. In: Semantic Web Applications in Scientific Discourse, 25-10-2009, Washington DC, US. (Submitted) [link]

The e-Labs collaboration model

An e-Lab consists of:

1. a community
2. research objects = data + methods + annotations
3. management infrastructure for research objects



- their abstract properties
  - life cycle
  - state
  - compound structure
  - provenance
- the rules and mechanisms for sharing
  - attribution, derivation, access control
  - aggregation / de-aggregation into components
  - introspection
  - exchange protocols

Paul's Pack

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.

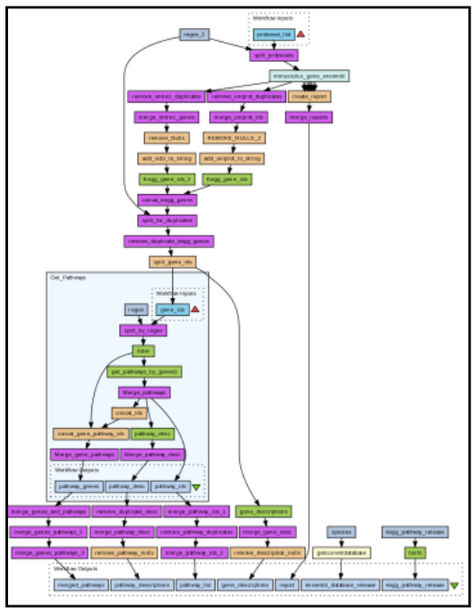
Logs

Metadata

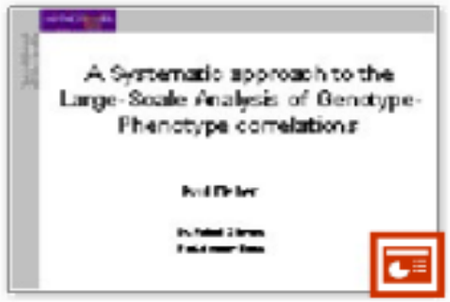
Results

```
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



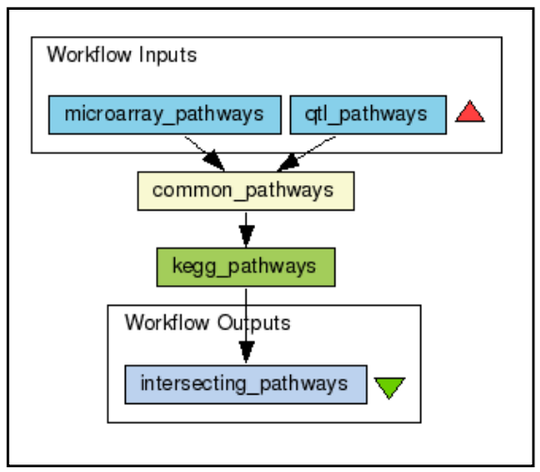
QTL



complete.ppt

Slides

Paper



Common pathways

Workflow 13

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

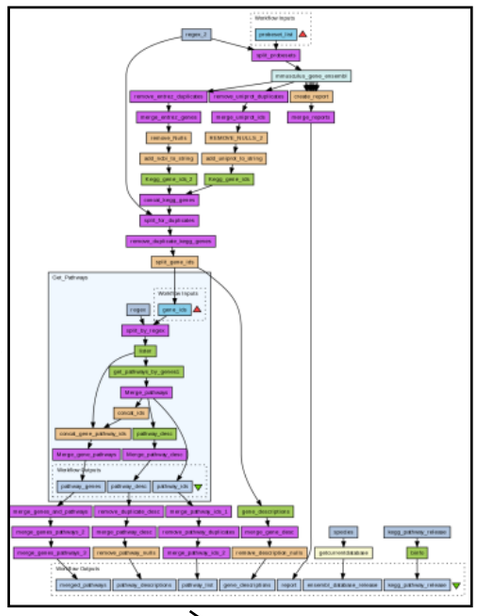
Results

Paul's Pack

Workflow 16

Results

QTL



```
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)
```

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.

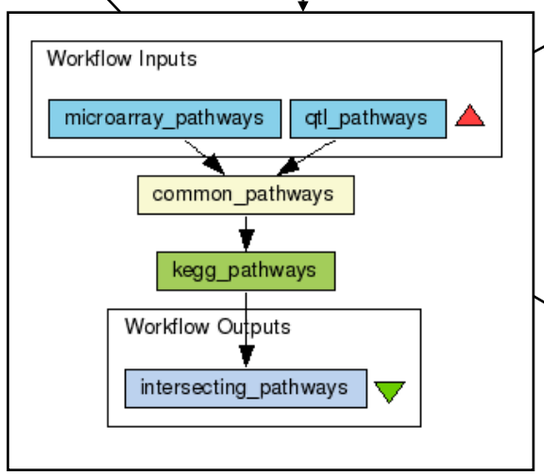
complete.ppt

Logs

Metadata

Slides

Paper



Common pathways

Workflow 13

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

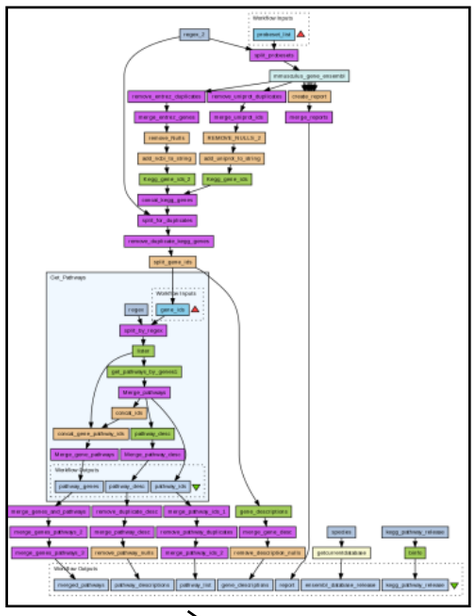
Paul's Research Object

Workflow 16

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



QTL

Published in

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.

Included in

Feeds into

Included in

complete.ppt

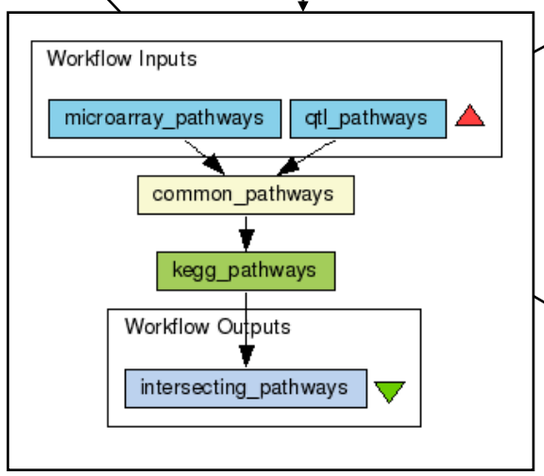
Included in

Published in

Logs

produces

Metadata



Common pathways

Workflow 13

Slides

Paper

produces

Published in

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

Current support in myExperiment: packs

Title: Taverna 2 workflow example and associated provenance

Description

test workflow along with a raw provenance dump which can be imported into mySQL

[edit]

Items (2)

Workflow: Workflow for provenance testing -- example 1 (Paolo)

[Add a comment here]

Added by **Paolo** ... 3 days ago (10/09/09 @ 14:22:33)

more

File: Provenance DB dump to be used as a reference (Paolo)

[Add a comment here]

Added by **Paolo** ... 3 days ago (10/09/09 @ 14:22:26)

more

Creator

Paolo

2 items in this pack

Tags (1)

Creator tags

provenancetesting

[edit]

Add Tags

Tags from Items (1)

provenancetesting

Shared with Groups (0)

None

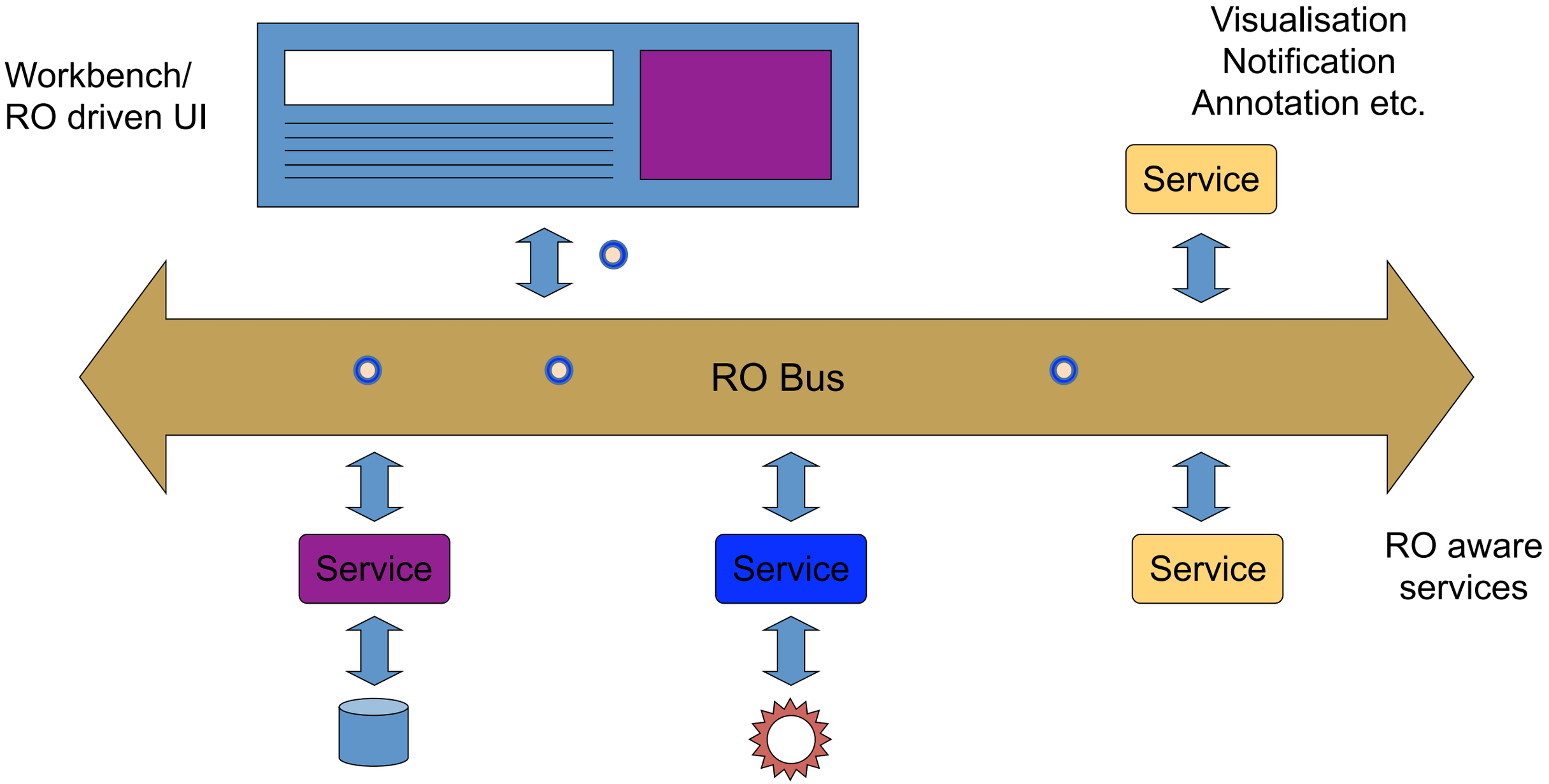
a Taverna workflow

raw provenance DB for a few runs of the workflow

Paolo uses myExperiment to interact with users whilst testing his Taverna provenance component

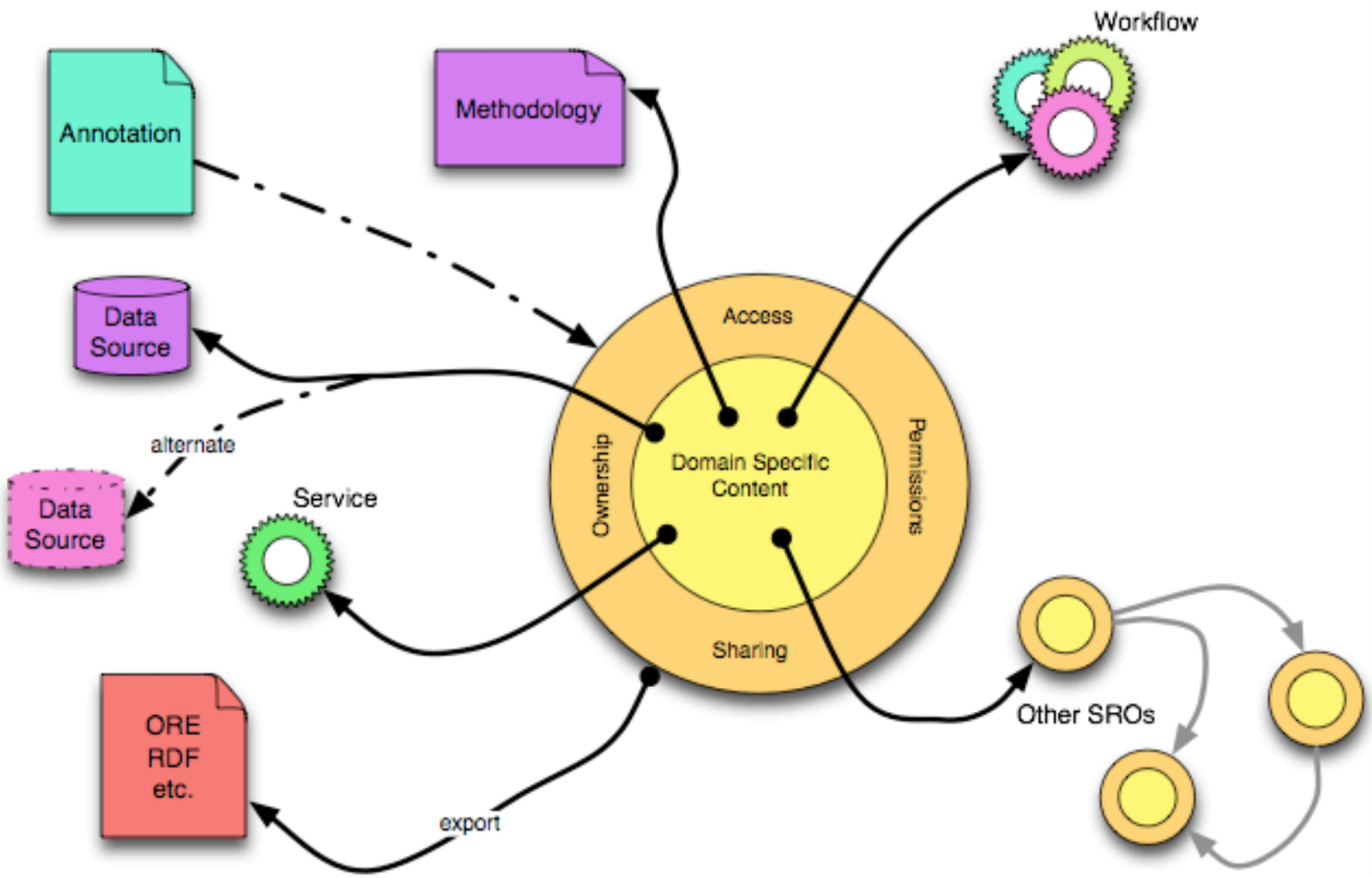
e-Labs + Research Objects

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

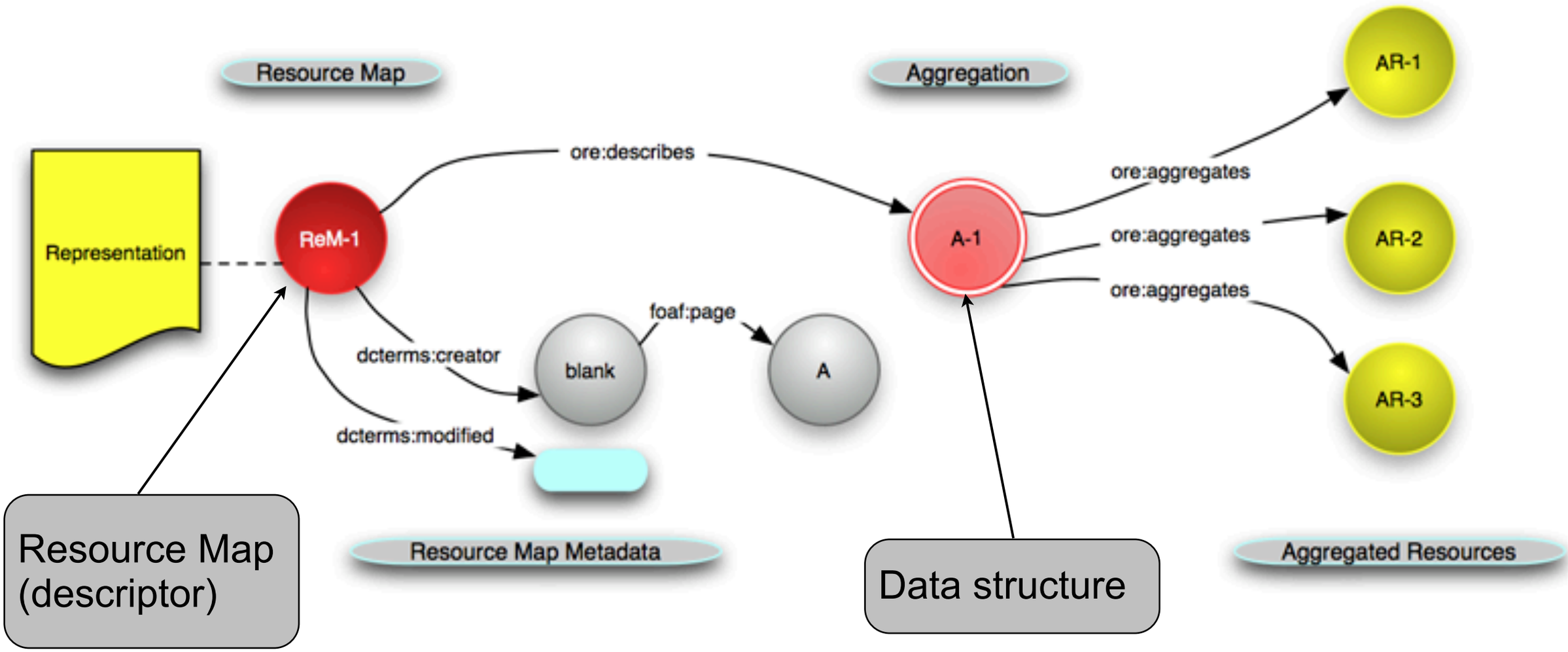


Anatomy of a research object

1. **Composite**
2. **Distributed**
3. **Annotated**
4. **Repeatable**
5. **Interoperable**



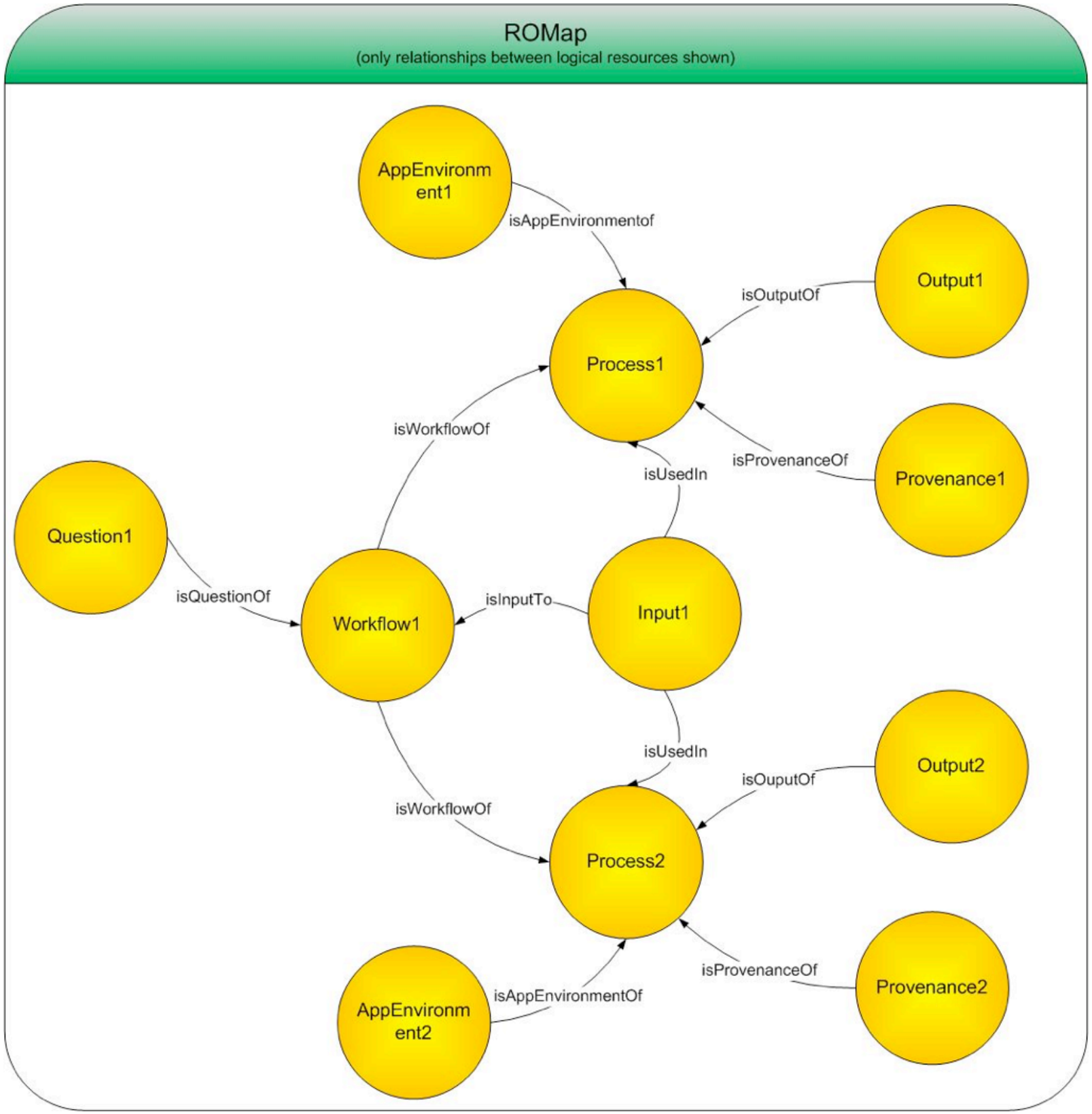
ORE: representing generic aggregations



http://www.openarchives.org/ore/1.0/primer.html section 4

Open Archives Initiative
Object Reuse and Exchange

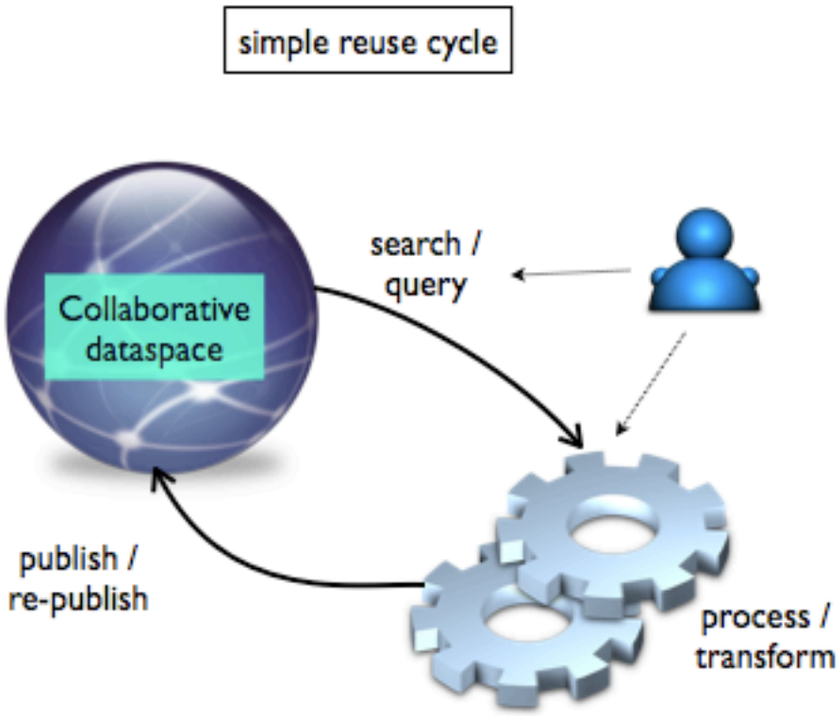
Research Objects: a concrete example

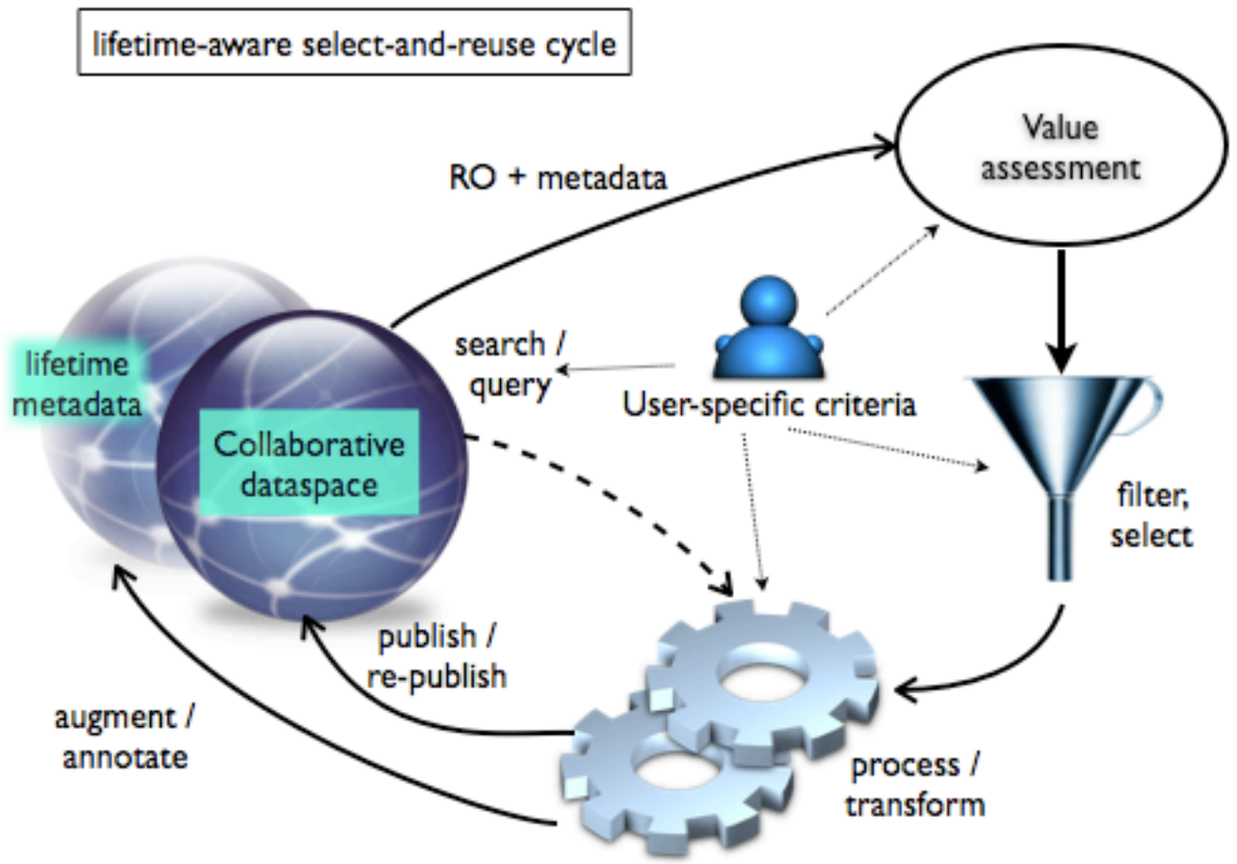


Source: e-Labs and Research Objects White Paper, School of Computer Science, University of Manchester, upcoming

Moving forward

- Research objects as the next currency in the scientific reputation market?
- From quantity to quality to value





Summary

- Collaborative / open science gaining momentum
- How do we get the collaboration model right?
- Technology:
  - many initiatives in the vocabulary / ontology space
  - it's mostly about adoption, integration
  - myExperiment moving quickly (and adventurously)
    - second round of funding secured from UK funding body
- The (potential) pitfalls:

quantity without quality = trouble...