The myExperiment Open Repository for Scientific Workflows

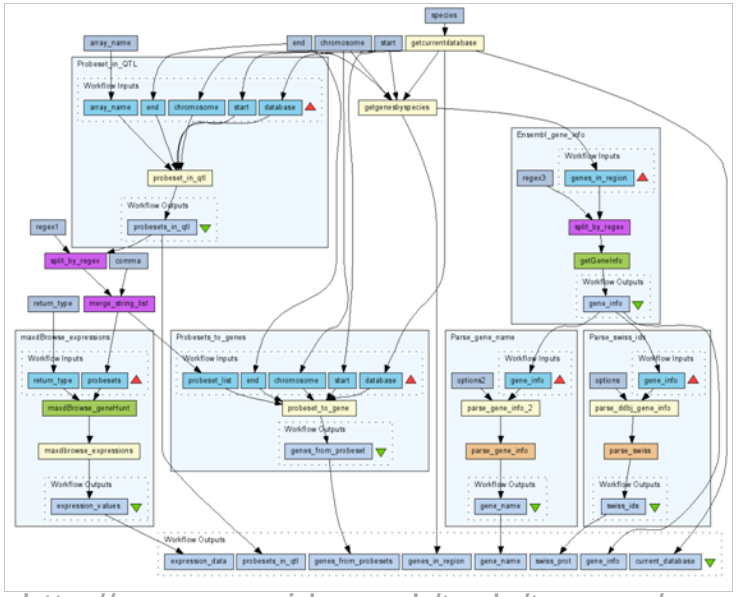
Don Cruickshank

University of Southampton

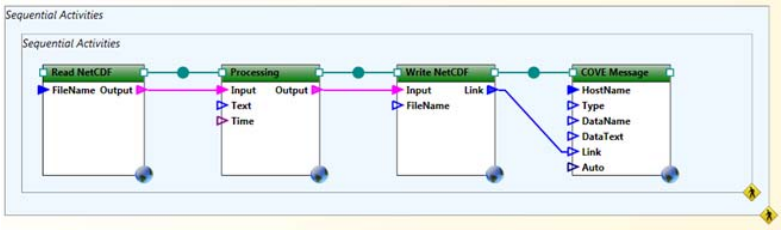
Open Repositories 2009, Atlanta, Georgia, USA
May 2009

What is a workflow?

Pieces of process



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

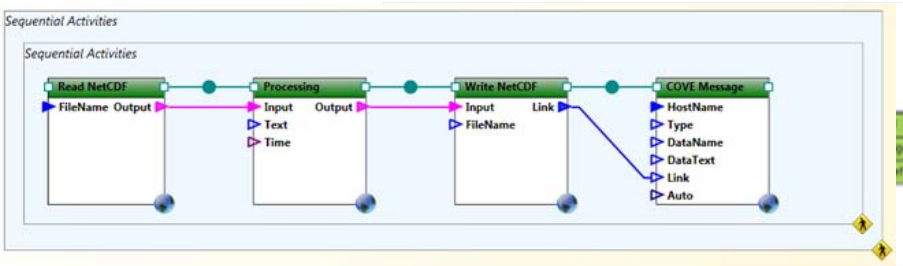


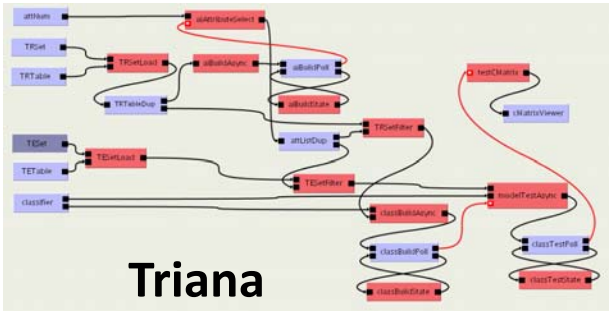
http://www.microsoft.com/mscorp/tc/trident.mspx

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

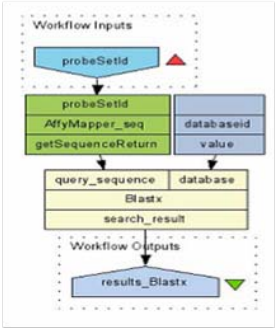
1. Make up separate 1 mL of 1M solutions of piperonal and 5-methylfurfurylamine in CDCl3.
2. Take HMRs and CMRs of the aldehyde and amine. Use 5 sec relaxation time and acquire for about 15 mins for the CMR. This should be good enough based on James' results at 1M in methanol.
3. Combine the two solutions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

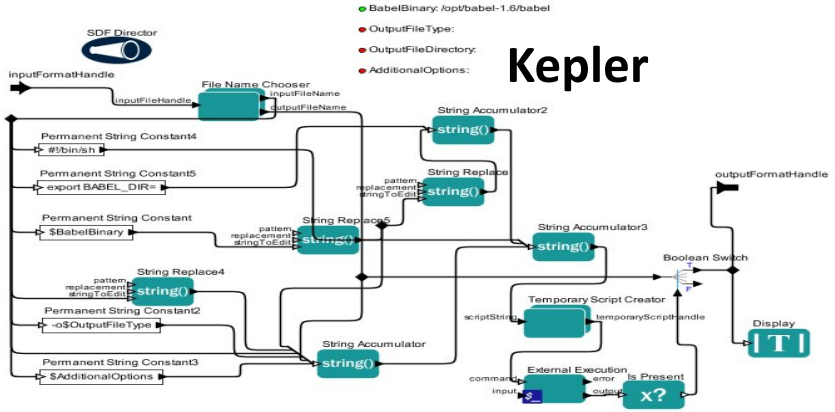
http://usefulchem.wikispaces.com/page/code/EXPLAN001

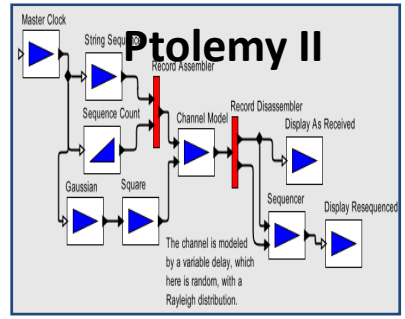




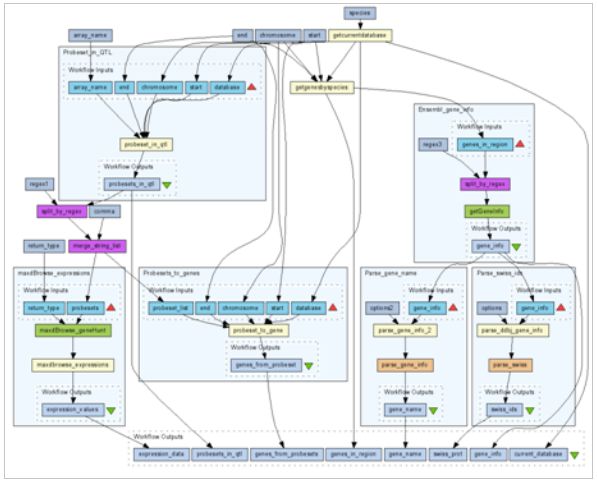
Trident

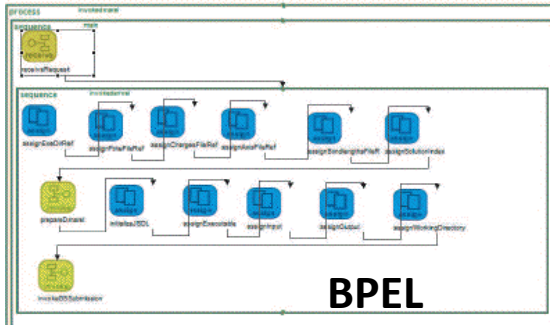


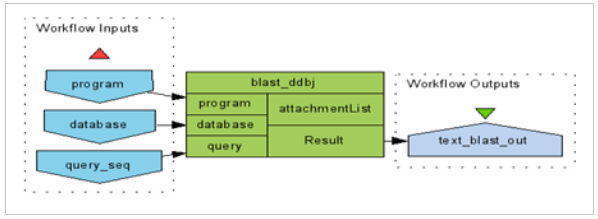


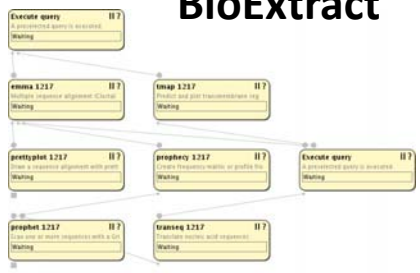


Taverna









Workflow Entry: FLOSS Communication Centralization Plot, Unit Weighted

Created at: 07/03/08 @ 02:00:42 Last updated: 07/02/09 @ 18:49:48

Version 2 (latest) (of 2)

View version: 2 (latest)

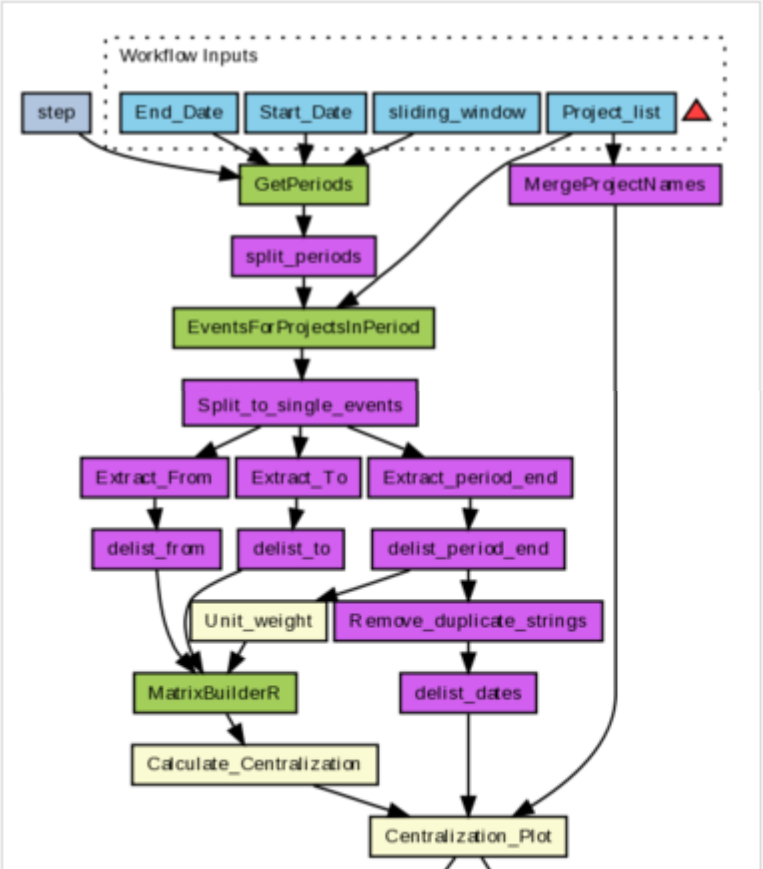
Version created on: 07/02/09 @ 18:49:47 by: Andrea Wiggins | Revision comments

Title: FLOSS Communication Centralization Plot, Unit Weighted

Type: Taverna 1

Preview

(Click on the image to get the full size)



Taverna 1 workflow

Original Uploader

Andrea Wiggins

License

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

Credits (3)

(People/Groups)

- Andrea Wiggins
- Crowston
- James Howison

Attributions (0)

(Workflows/Files)

None

Tags (8)

Original Uploader tags

centralization | communication network | floss | open source software | oss | sna | social

Carole Goble

User: Marco Roos

Name: Marco Roos

Joined: Saturday 21 July 2007 @ 10:43:23 (BST)

Last seen: Friday 15 May 2009 @ 15:50:51 (BST)

Email (public): roos [at] science.uva.nl

Website: http://home.medewerker.uva.nl/m.roos1

Location: Amsterdam, Netherlands

44 Friends
10 Groups (admin)
7 Groups (member)
2 Blogs
7 Packs
2 Files
26 Workflows
0 Favourites

Marco Roos has been credited 28 times

Marco Roos has an average rating of:

3.8 / 5

(6 ratings in total)

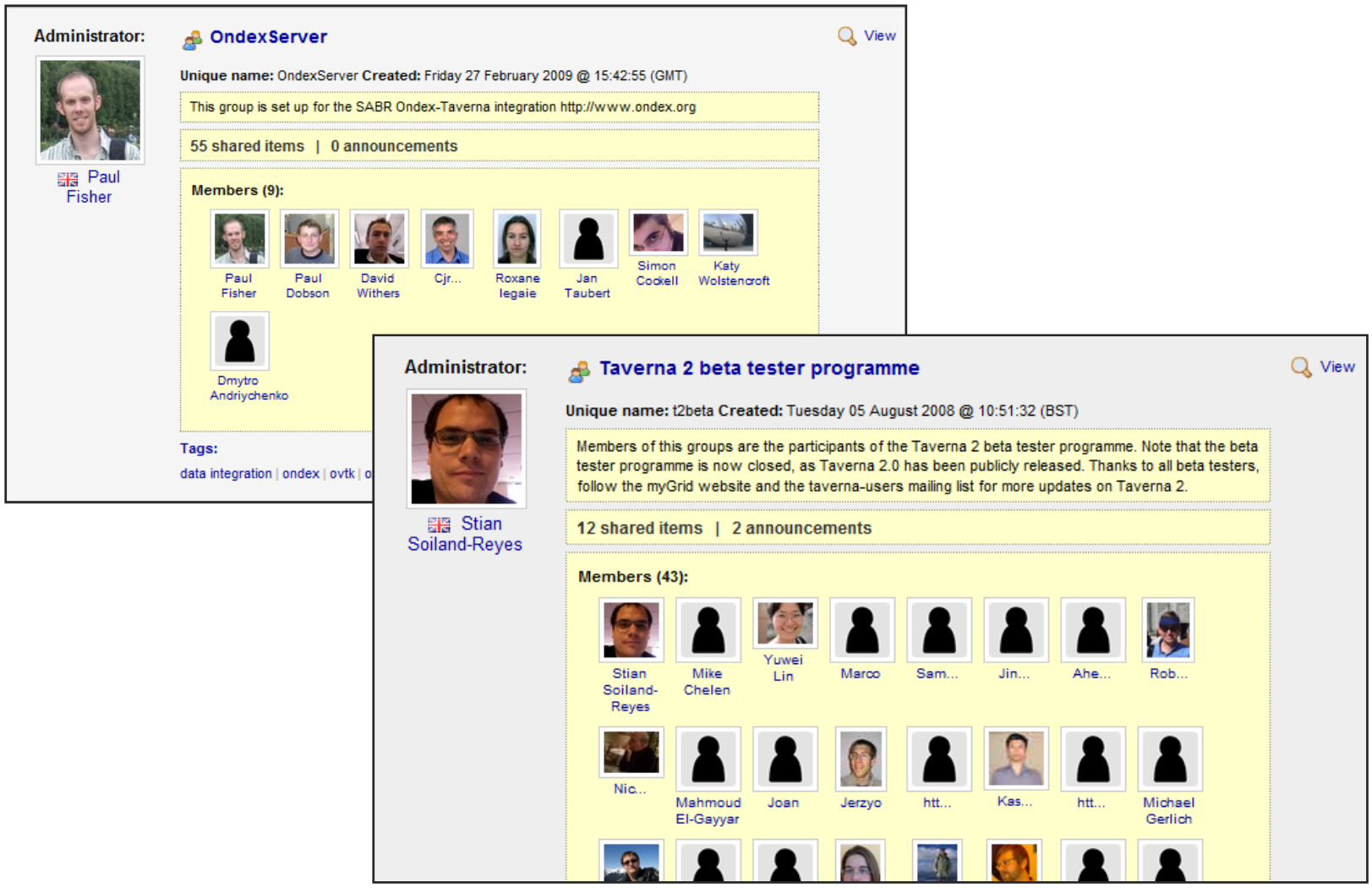
for their items

My role as a biologist and bioinformatician in e-science is to help increase the usefulness of emerging information technologies for biology, while experimenting with new ways to increase insight into mechanisms related to structure and function of DNA in the cell. I experiment with technologies such as workflow, knowledge extraction from text, semantic web and virtual research environments such as myExperiment.

More information on the blog below (originally uploaded as an example for the 'NBIC on workflows' workshop in Lunteren, the Netherlands, March 2008).

http://www.myexperiment.org/blogs/15

Groups in myExperiment



Workflow types in myExperiment

- Not *all* workflows are stored in XML
  - Many are stored in XML
  - Some are XML but wrapped in other containers, e.g. ZIP
  - Some are just plain text files
- Many types of workflows out there
  - We cannot hope to provide specialised support for all!
  - XML makes it easier, but thumbnail generation is the common sticking point

| 587 Taverna 1 workflows | 28 Taverna 2 beta workflows | 24 GWorkflowDL workflows | 12 Chemistry Plan workflows | 6 Trident (Package) workflows | 3 Trident (XOML) workflows | 3 WSVLAM workflows | 2 Nimrod Parameter Sweep workflows | 2 Excel 2007 Macro-Enabled Workbook workflows | 2 Kepler workflows | 2 SimileXMLv3 workflows | 1 Stella workflow | 1 Pipeline Pilot workflow | 1 Makefile workflow | 1 Manual Web Service workflow |

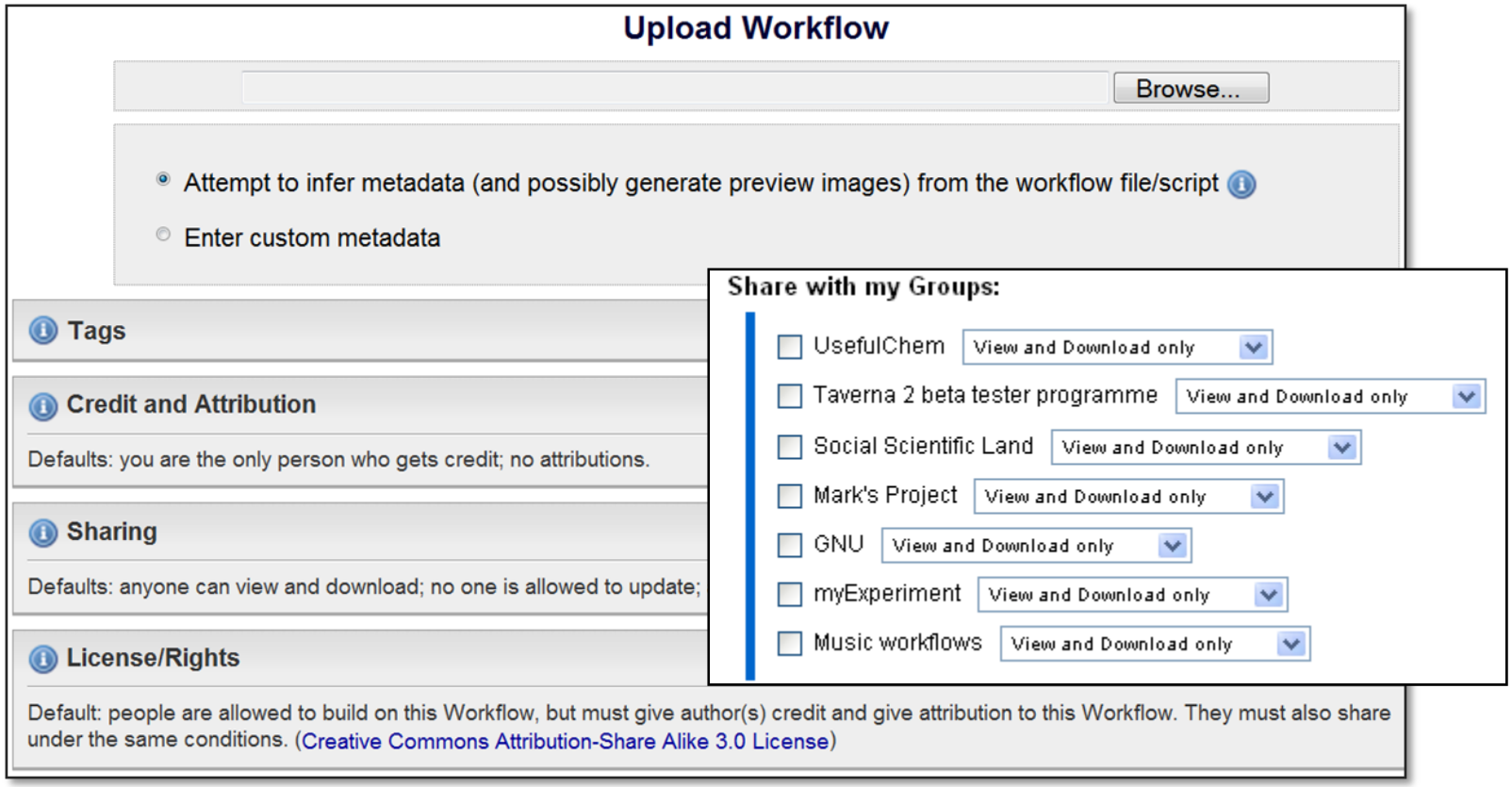
Upload New Workflow | View All Workflows

Top 50 tags for Workflows [See All Tags]

affymetrix | AIDA | alignment | benchmarks | bio2rdf | BioAID | bioassist_nl | bioinformatics | biomoby | biorange_nl | BLAST | cdk-taverna | d-grid | data integration | demo | design pattern | disease | e-science | ebi | example | gene | genotype | graph | gwes | gworkflowdl | kegg | Kegg Pathways | localworker | microarray | multiple sequence alignment | mygrid | ondex | pathway | pathway-driven | pathways | phenotype | protein | protein annotation | pubmed | sequence | sequence similarity search | shim | social sciences | taverna | test | text mining | text_mining | uniprot | VL-e | workflow pattern

Latest | Last Updated | Most Viewed | Most Downloaded | Most Favourited

“Just enough” user controlled sharing



The most important aspect of myExperiment

Designed by scientists

Packs

- Simple aggregation of resources
- Started as a “just enough” piece for scientists
- Maps onto ORE

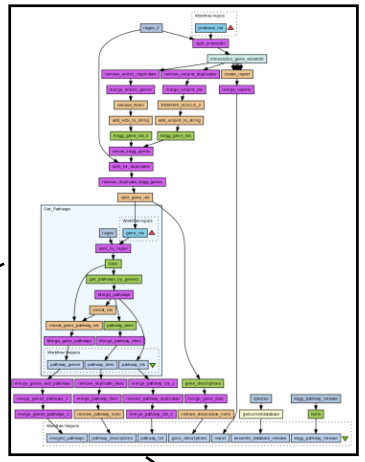
Paul Fisher

Workflow 16

QTL

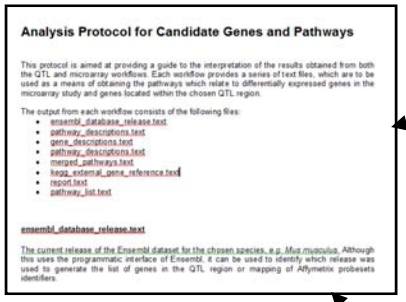
Logs

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



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

Published in

Included in

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

complete.ppt

Included in

Included in

produces

Metadata

Slides

Paper

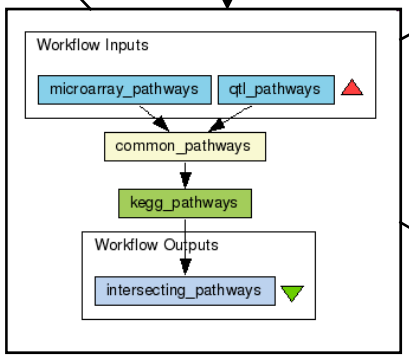
Tags (19)

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

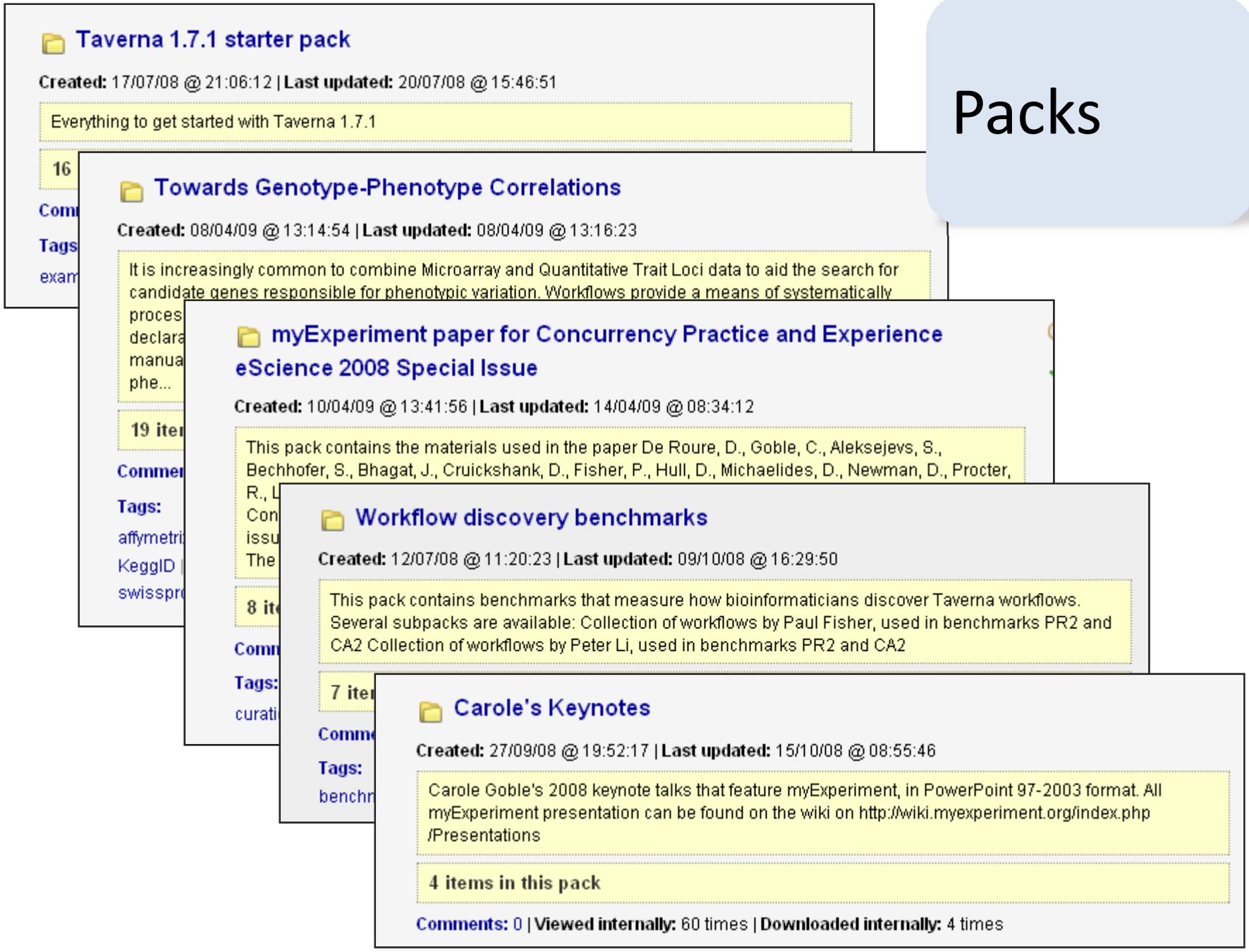
Workflow 13

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

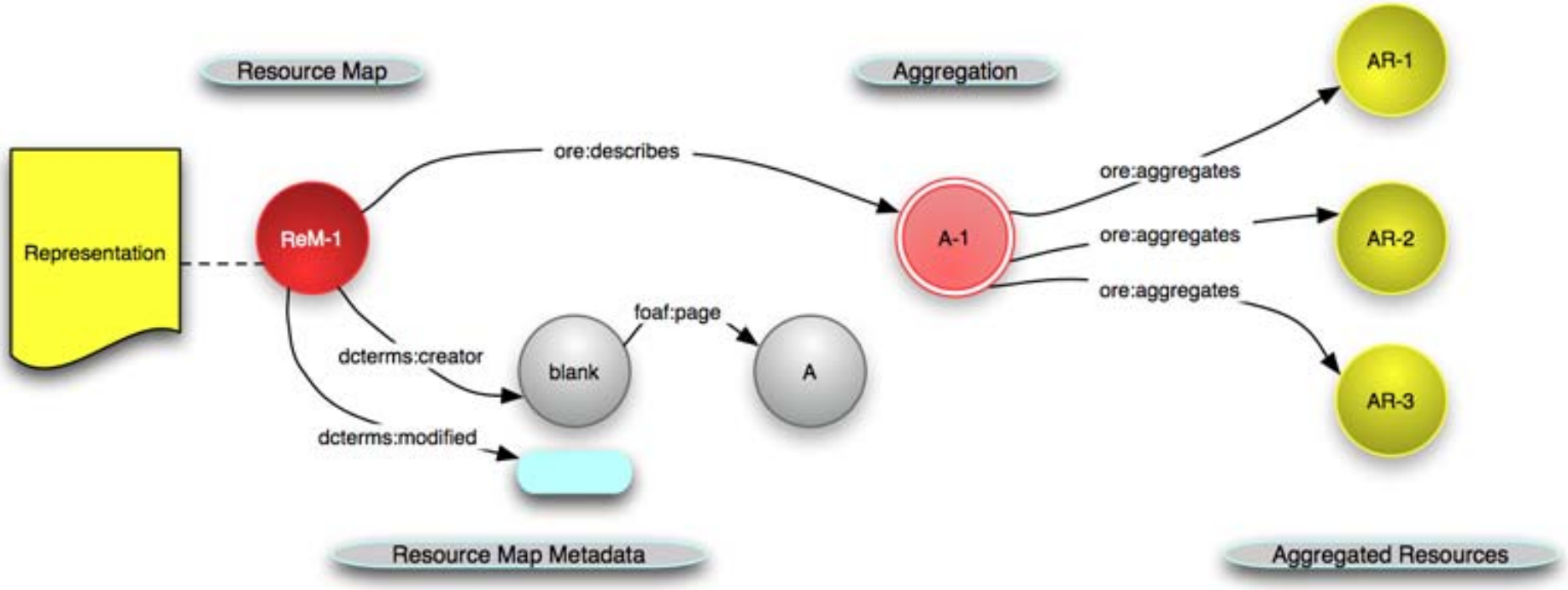
Published in

Results

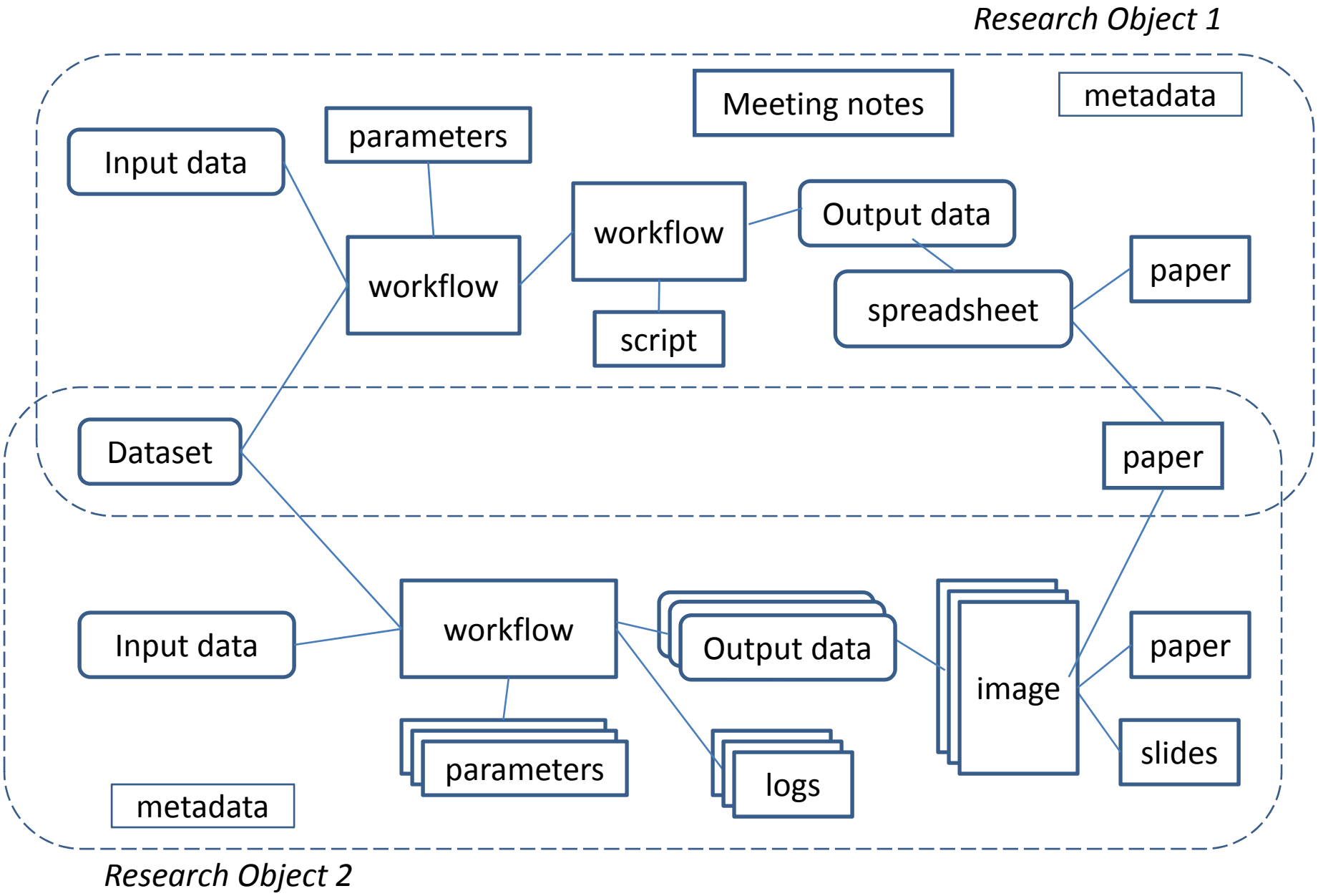

Taverna 1.7.1 starter pack
Created: 17/07/08 @ 21:06:12 | Last updated: 20/07/08 @ 15:46:51
Everything to get started with Taverna 1.7.1
Towards Genotype-Phenotype Correlations
Created: 08/04/09 @ 13:14:54 | Last updated: 08/04/09 @ 13:16:23
myExperiment paper for Concurrency Practice and Experience 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.,
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 CA2 Collection of workflows by Peter Li, used in benchmarks PR2 and CA2
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
4 items in this pack
Comments: 0 | Viewed internally: 60 times | Downloaded internally: 4 times

Packs

Exporting packs



Open Archives Initiative
Object Reuse and Exchange


Research Object 1
Meeting notes
metadata
Input data
parameters
workflow
workflow
Output data
spreadsheet
paper
script
Dataset
paper
Input data
workflow
Output data
image
paper
slides
parameters
logs
metadata
Research Object 2

Developer resources

- REST API
  - Import and export
- SPARQL query engine
  - Export only
- ORE
  - Export only
- Documentation on the wiki (in progress)

For Developers

myexperiment

XML

facebook

iGoogle

android

HTML

API config

ORE
FOAF
SIOC

Managed REST API

Search API

Search Engine

SPARQL endpoint

RDF Store

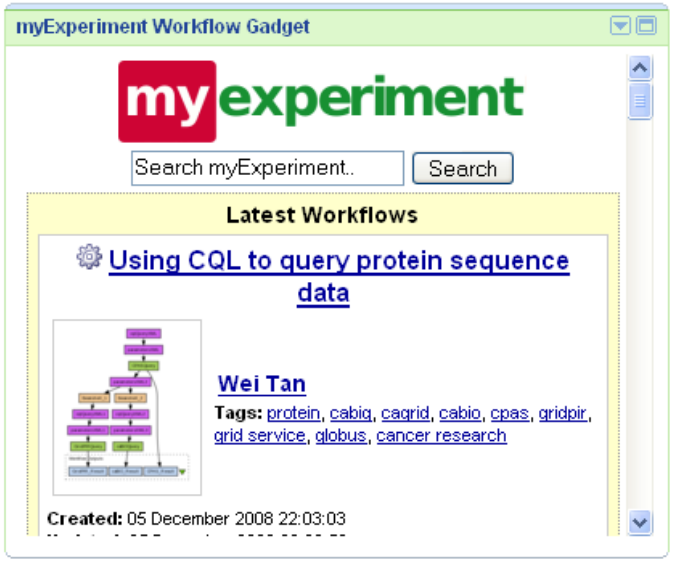
tags ratings reviews profiles
workflows credits groups
files packs friendships

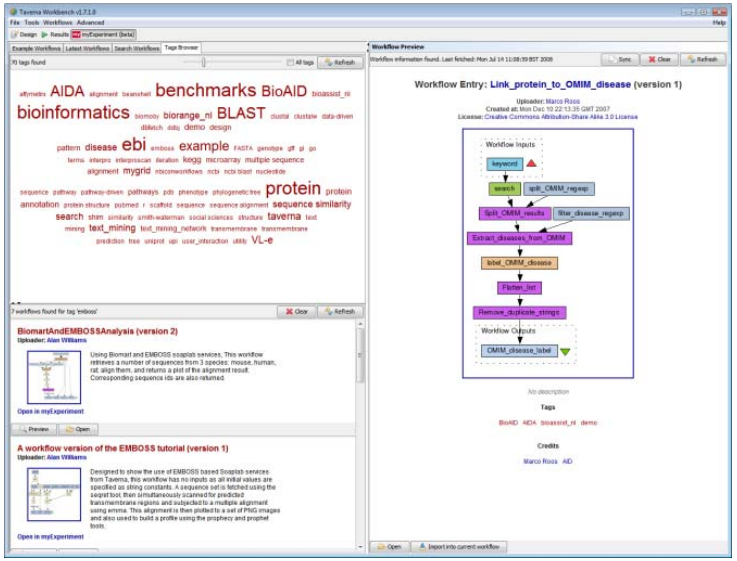
EPrints
DSpace
Fedora
S3
SRB

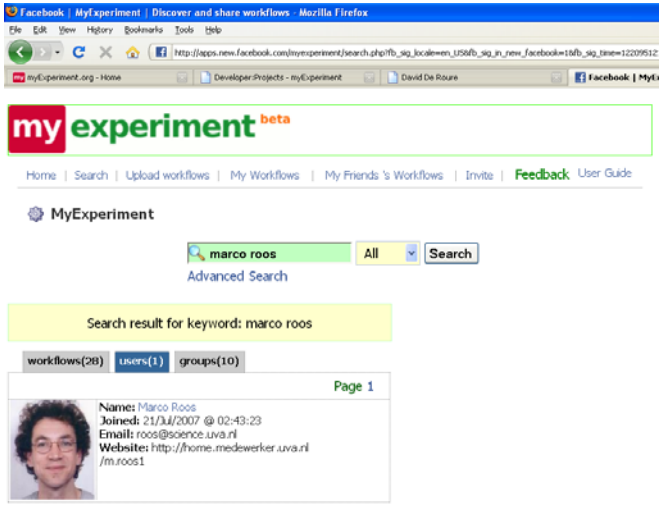
mySQL

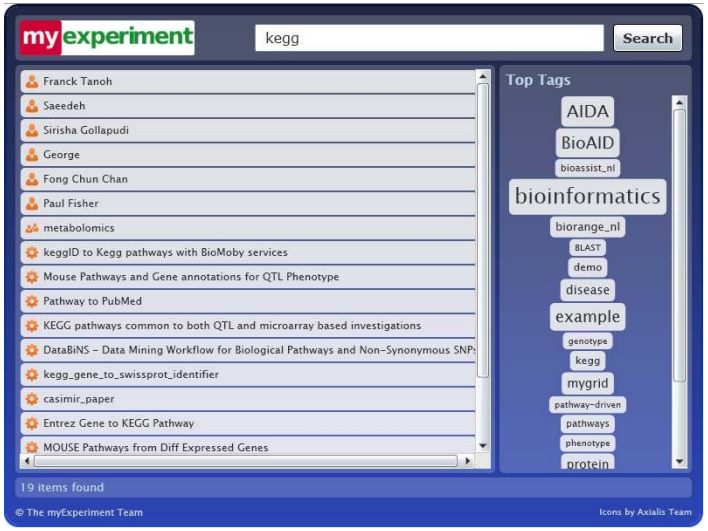
Enactor API

Enactor


myExperiment Workflow Gadget
myexperiment
Search myExperiment..
Search
Latest Workflows
Using CQL to query protein sequence data
Wei Tan
Tags: protein, cabig, caGrid, cabio, cpas, gridpir, grid service, globus, cancer research
Created: 05 December 2008 22:03:03


Workflow Entry: Link_protein_to_OMIM_disease (version 1)
BiomartAndEMBOSSAnalysis (version 2)
A workflow version of the EMBOSS tutorial (version 1)


Facebook | MyExperiment | Discover and share workflows - Mozilla Firefox
myexperiment 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

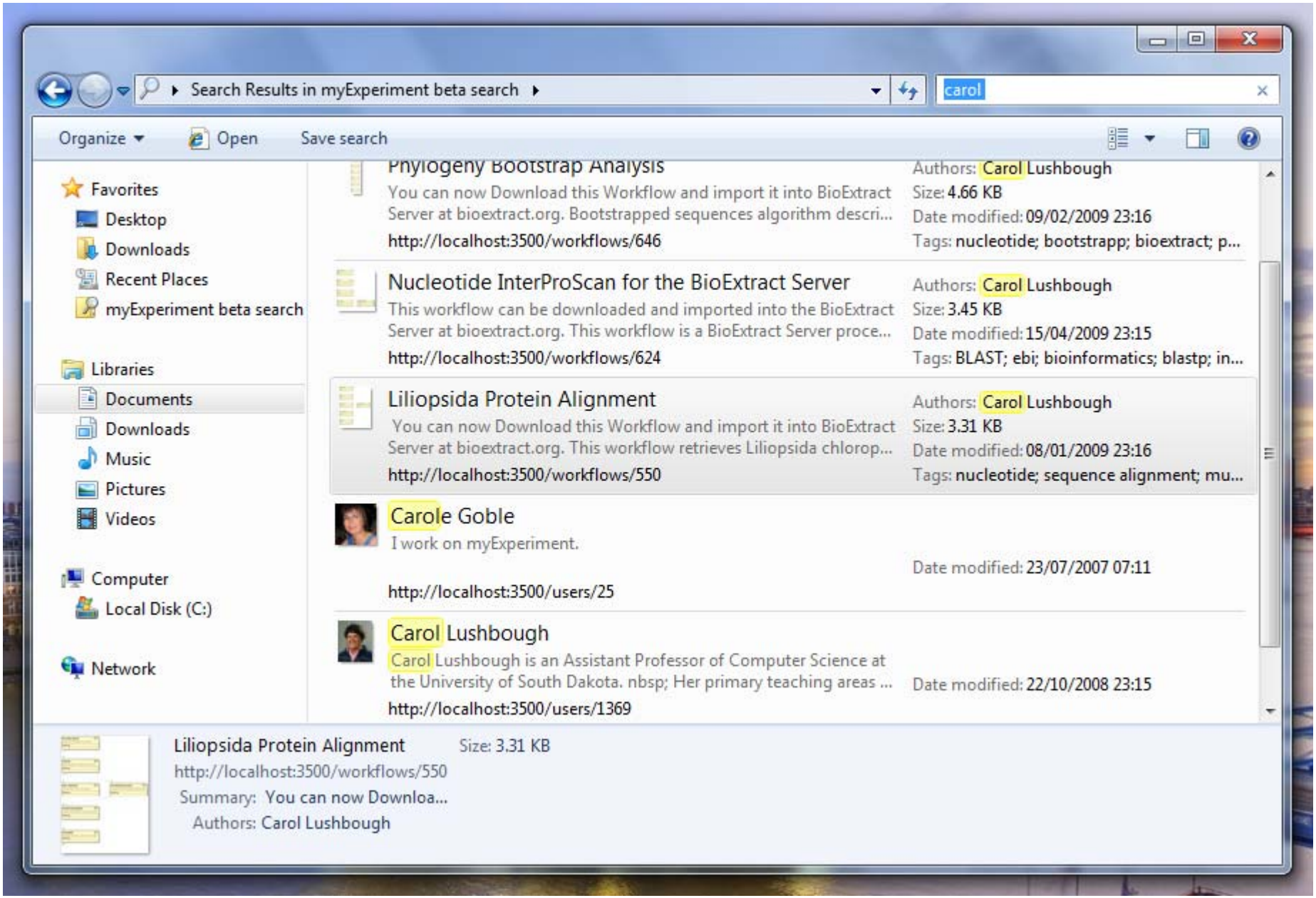

myexperiment
kegg
Search
Franck Tanoh
Saeedeh
Sirisha Gollapudi
George
Fong Chun Chan
Paul Fisher
metabolomics
keggID to Kegg pathways with BioMoby services
Mouse Pathways and Gene annotations for QTL Phenotype
Pathway to PubMed
KEGG pathways common to both QTL and microarray based investigations
DataBiNS – Data Mining Workflow for Biological Pathways and Non-Synonymous SNP
kegg_gene_to_swissprot_identifier
casimir_paper
Entrez Gene to KEGG Pathway
MOUSE Pathways from Diff Expressed Genes
Top Tags
AIDA
BioAID
bioassist_nl
bioinformatics
biorange_nl
BLAST
demo
disease
example
genotype
kegg
mygrid
pathway-driven
pathways
phenotype
protein
19 items found
© The myExperiment Team
Icons by Axialis Team

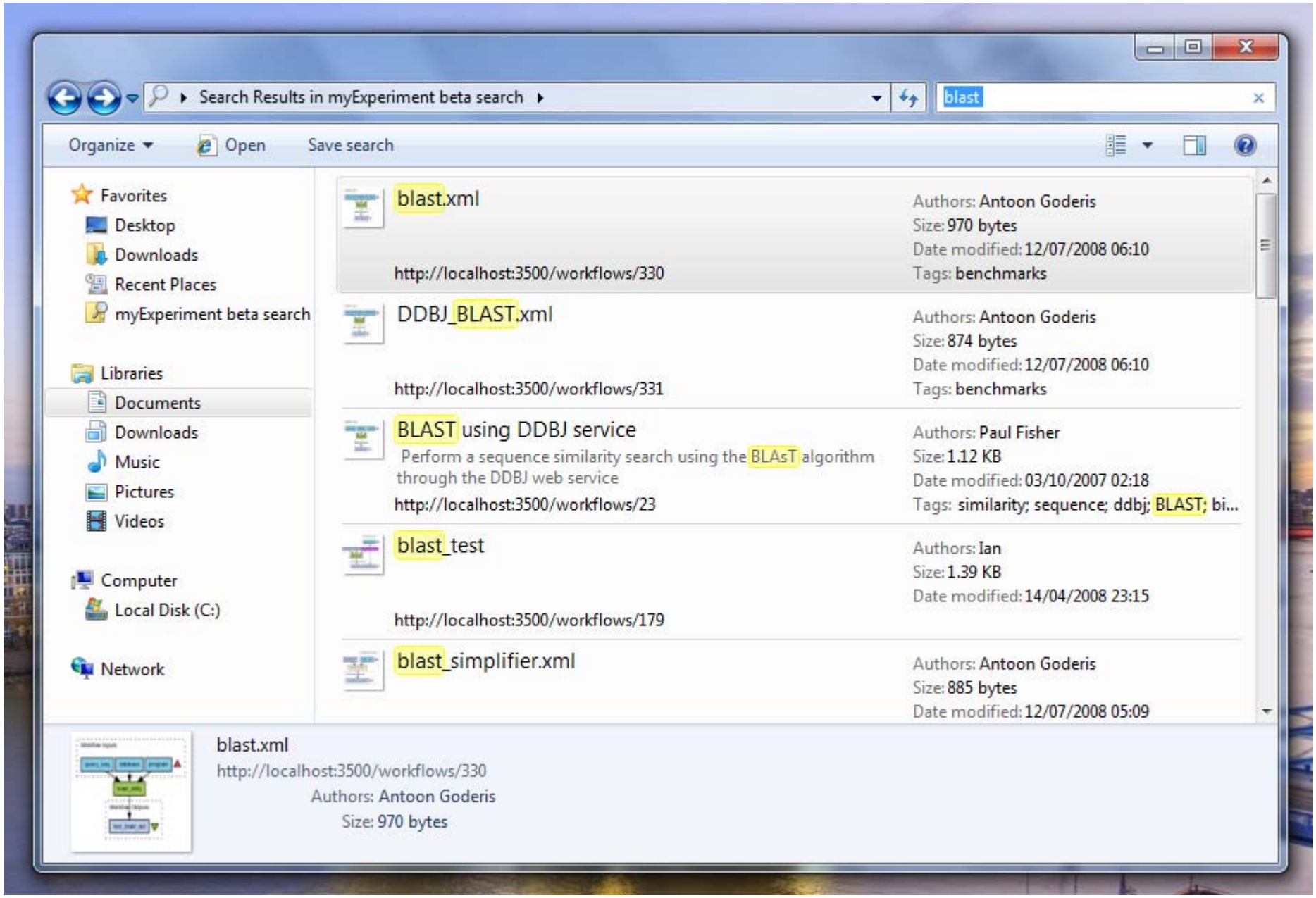
Authentication

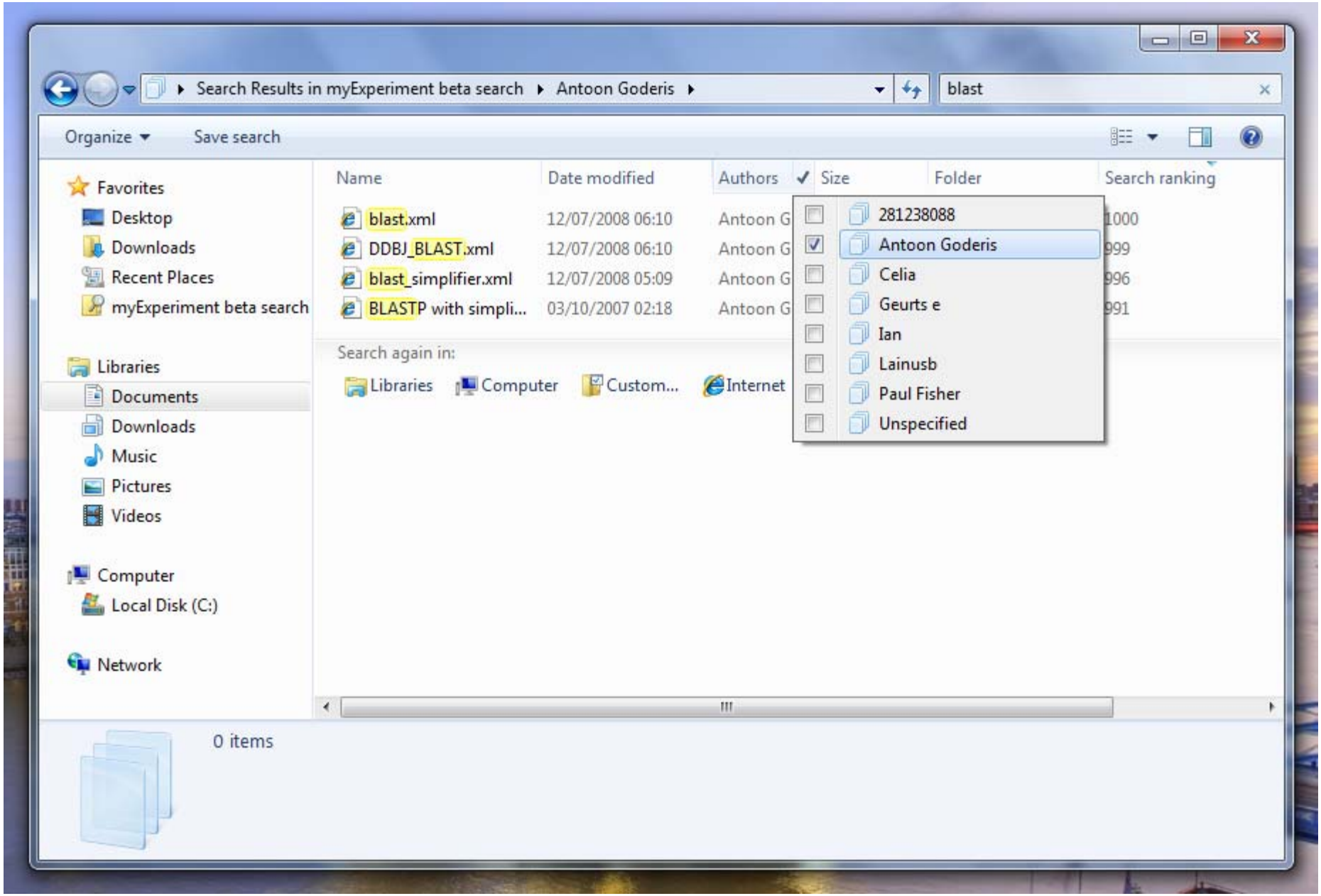
- OpenID for single sign-on
- Oauth for secure mashups
  - User keys
  - System keys
- Username/password for HTTP access

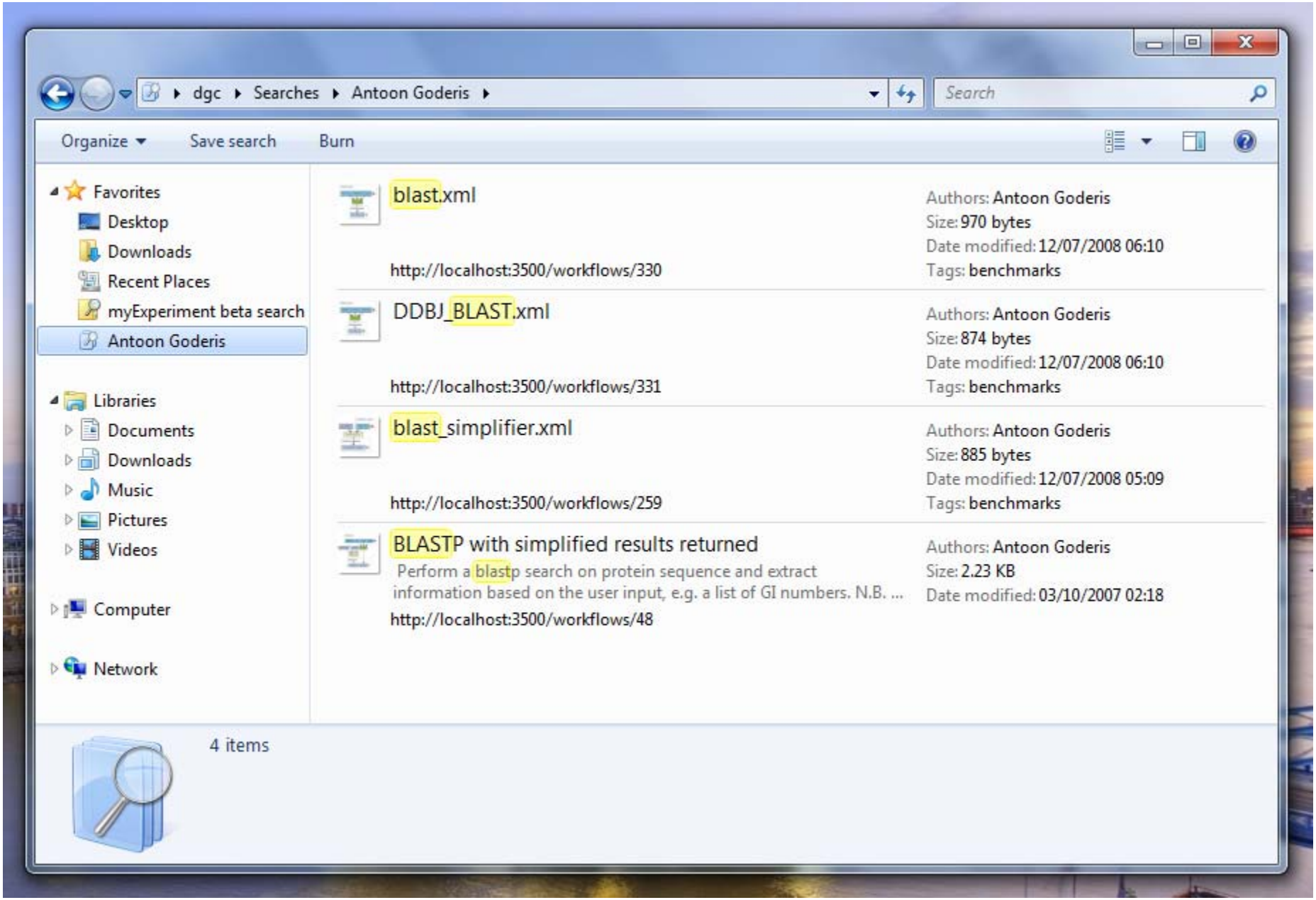
Windows 7 Desktop Integration

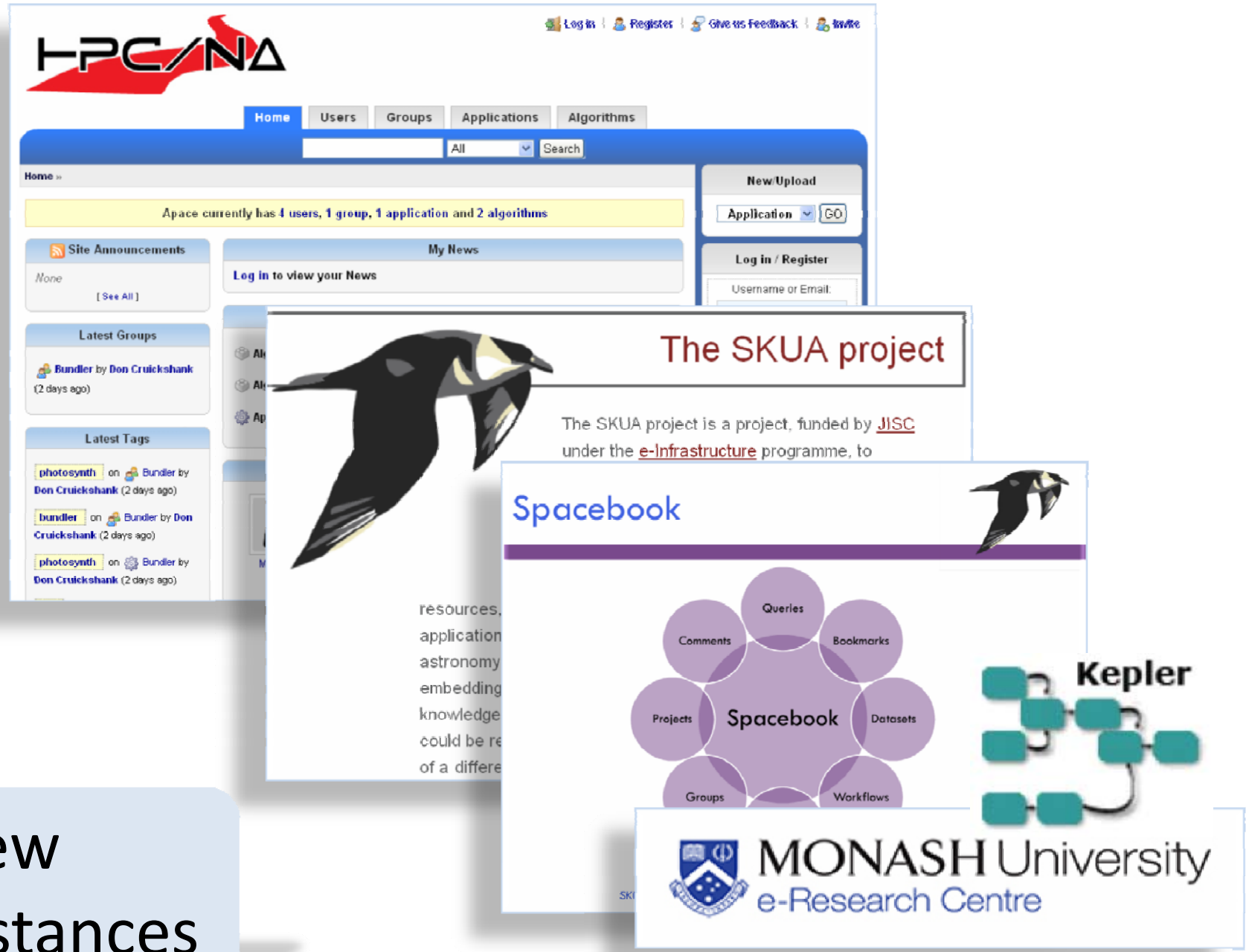
- This is still in development
- Uses the OpenSearch protocol as offered by Amazon
- Does not use any proprietary extensions from Microsoft (yet)


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


Search Results in myExperiment beta search
blast
Organize
Open
Save search
Favorites
Desktop
Downloads
Recent Places
myExperiment beta search
Libraries
Documents
Downloads
Music
Pictures
Videos
Computer
Local Disk (C:)
Network
blast.xml
http://localhost:3500/workflows/330
Authors: Antoon Goderis
Size: 970 bytes
Date modified: 12/07/2008 06:10
Tags: benchmarks
DDBJ_BLAST.xml
http://localhost:3500/workflows/331
Authors: Antoon Goderis
Size: 874 bytes
Date modified: 12/07/2008 06:10
Tags: benchmarks
BLAST using DDBJ service
Perform a sequence similarity search using the BLAsT algorithm through the DDBJ web service
http://localhost:3500/workflows/23
Authors: Paul Fisher
Size: 1.12 KB
Date modified: 03/10/2007 02:18
Tags: similarity; sequence; ddbj; BLAST; bi...
blast_test
http://localhost:3500/workflows/179
Authors: Ian
Size: 1.39 KB
Date modified: 14/04/2008 23:15
blast_simplifier.xml
Authors: Antoon Goderis
Size: 885 bytes
Date modified: 12/07/2008 05:09
blast.xml
http://localhost:3500/workflows/330
Authors: Antoon Goderis
Size: 970 bytes


Search Results in myExperiment beta search
Antoon Goderis
blast
Organize
Save search
Favorites
Desktop
Downloads
Recent Places
myExperiment beta search
Libraries
Documents
Downloads
Music
Pictures
Videos
Computer
Local Disk (C:)
Network
Name
Date modified
Authors
Size
Folder
Search ranking
blast.xml
12/07/2008 06:10
Antoon G
1000
DDBJ_BLAST.xml
12/07/2008 06:10
Antoon G
999
blast_simplifier.xml
12/07/2008 05:09
Antoon G
996
BLASTP with simpli...
03/10/2007 02:18
Antoon G
991
281238088
Antoon Goderis
Celia
Geurts e
Ian
Lainusb
Paul Fisher
Unspecified
Search again in:
Libraries
Computer
Custom...
Internet
0 items

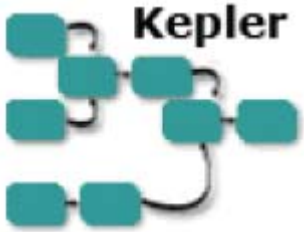

dgc ▸ Searches ▸ Antoon Goderis ▸
Search
Organize
Save search
Burn
Favorites
Desktop
Downloads
Recent Places
myExperiment beta search
Antoon Goderis
Libraries
Documents
Downloads
Music
Pictures
Videos
Computer
Network
blast.xml
http://localhost:3500/workflows/330
Authors: Antoon Goderis
Size: 970 bytes
Date modified: 12/07/2008 06:10
Tags: benchmarks
DDBJ_BLAST.xml
http://localhost:3500/workflows/331
Authors: Antoon Goderis
Size: 874 bytes
Date modified: 12/07/2008 06:10
Tags: benchmarks
blast_simplifier.xml
http://localhost:3500/workflows/259
Authors: Antoon Goderis
Size: 885 bytes
Date modified: 12/07/2008 05:09
Tags: benchmarks
BLASTP with simplified results returned
Perform a blastp search on protein sequence and extract information based on the user input, e.g. a list of GI numbers. N.B. ...
http://localhost:3500/workflows/48
Authors: Antoon Goderis
Size: 2.23 KB
Date modified: 03/10/2007 02:18
4 items

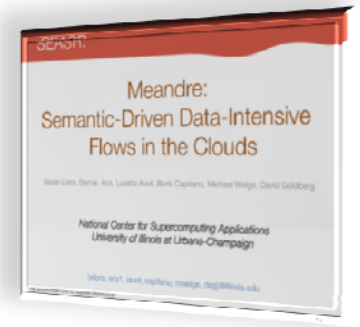


New Instances

Phase 2

- Notifications
- Taverna 2 support
- Support for expert curators
- Controlled vocabularies
- Faceted browsing
- New contribution types (Scripts, Meandre, Kepler, e-books)
- Biocatalogue integration
- Relationships between items (in and between packs)
- Indexing of packs
- Further blog / wiki integration
- Repository integration (EPrints, Fedora)
- Recommendations



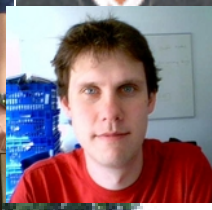


myGrid Team















David De Roure

More information

- The main site: www.myexperiment.org
- For more information, see the wiki at wiki.myexperiment.org