



Netherlands
Bioinformatics
Centre

Opening new gateways to workflows for life scientists

Kostas Karasavvas

Marco Roos

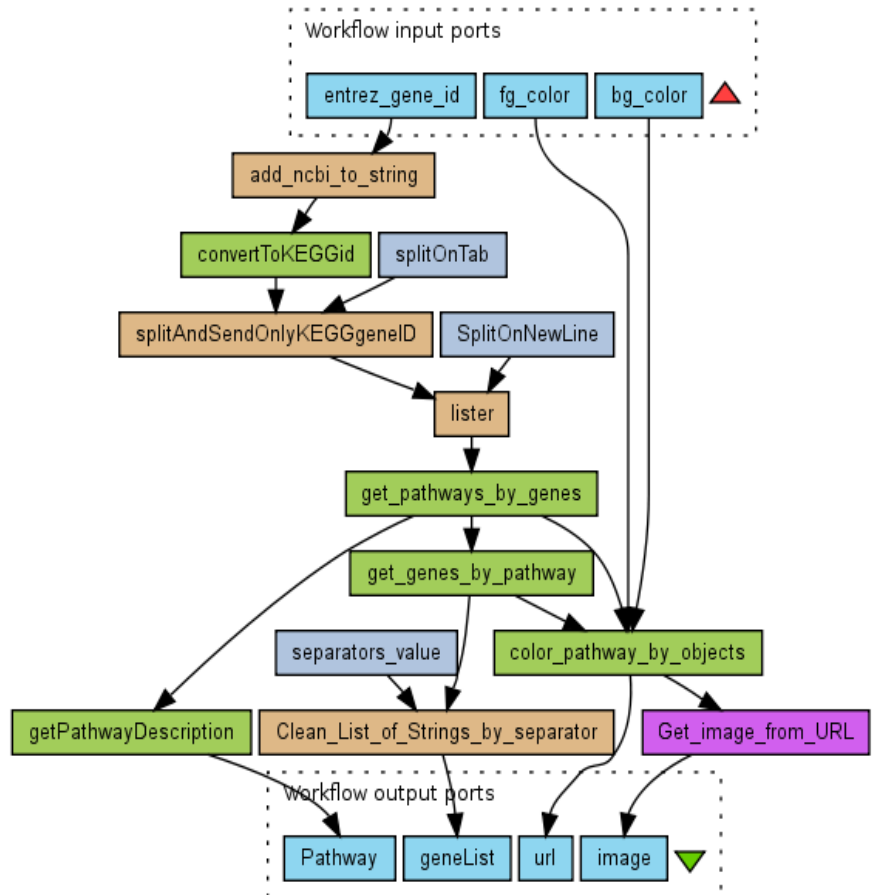


Leiden University Medical Center



Workflow approach

- Structured approach to performing bioinformatics experiments
- Steps are exposed
- Helps evaluation by supervisors and peers
- Compares to 'Materials and Methods', but you can run them
- Easier to reuse & extend (manage complexity)



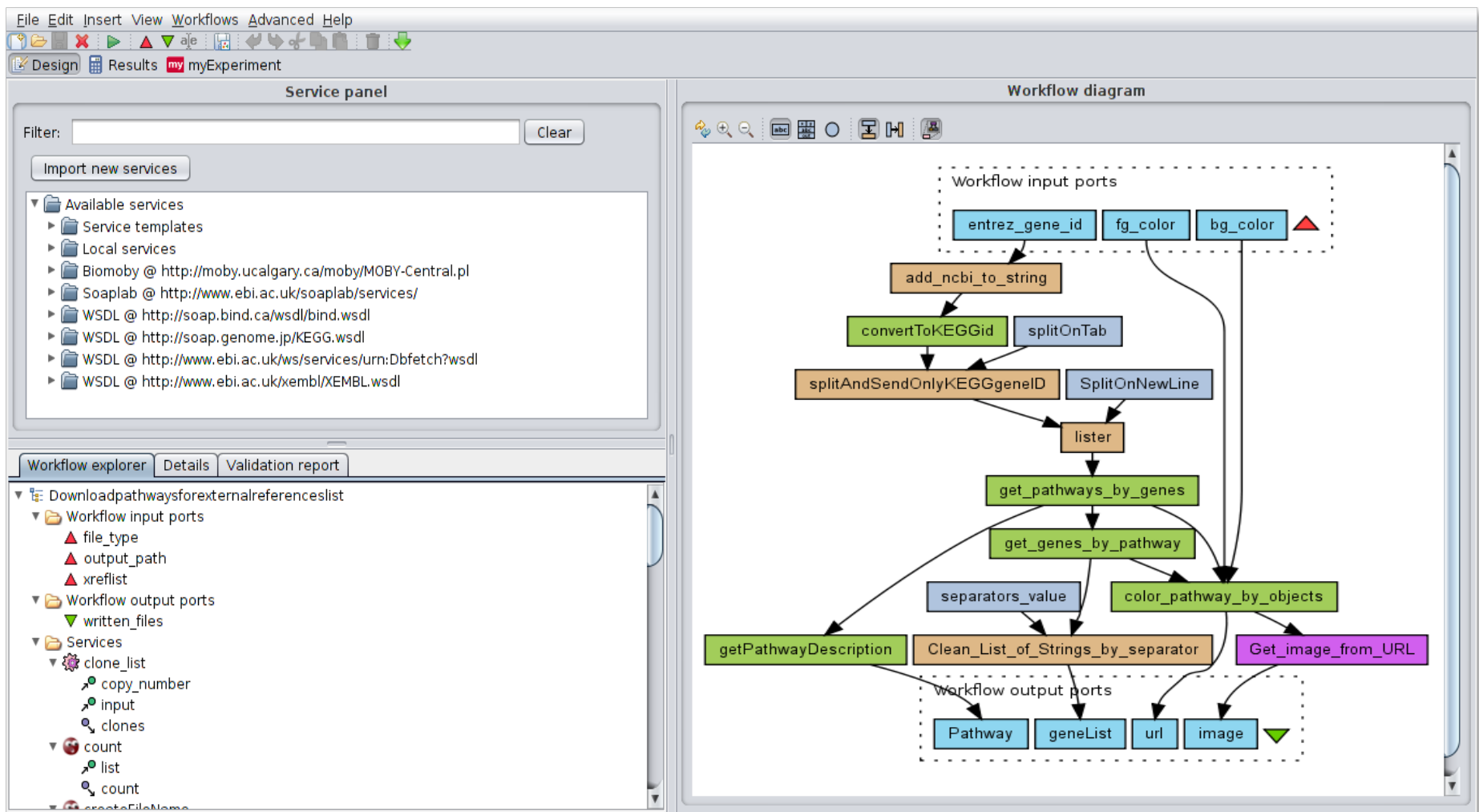
Observations

- Two widely-used systems: Galaxy & Taverna
- Typical usage scenario's
 - Galaxy:
 - 'Scripting bioinformatician' adds command line tools to local Galaxy server for their biological colleagues
 - Biologists interactively 'play with data'
 - Save interactive analysis as a workflow
 - Strength: simplicity
 - Taverna
 - Bioinformatician designs a workflow to perform complicated data analysis / data integration
 - Strength: sophisticated workflows

Why interoperate?

- Use each other's tools, use each other's workflows
 - Example:
 - calculate genome region overlaps in Galaxy
 - annotate genes by text mining Web Services in Taverna
- Exploit strengths
 - Galaxy: simplicity
 - Taverna: sophistication
- Allow user communities to benefit from each other's work
 - greater accessibility

Taverna Workbench



Galaxy

The screenshot shows the Galaxy web interface with the following components:

- Browser Tabs:** Includes tabs for 'De best...', 'Taverna...', 'UCSC G...', 'Custom...', 'Book S...', 'O διαβέ...', 'Proble...', 'Galaxy', 'Results:', 'W E-comm...', 'The Rub...', 'NBIC: A...', 'myExpe...', 'ruby co...', 'Connect...', 'Galaxy', 'Taverna...', 'Inbox | ...'.
- Address Bar:** Shows 'localhost:8080' and a search bar with 'taverna server setting the timeout'.
- Galaxy Header:** Features the 'Galaxy' logo and navigation links: 'Analyze Data', 'Workflow', 'Shared Data', 'Help', 'User'.
- Tools Panel (Left):** Contains a list of tools under 'Taverna Workflows', including 'Fetch PDB flatfile from RCSB server', 'Demonstration of configurable iteration', 'EBI InterProScan NewServices', 'BiomartAndEMBOSSAnalysis', 'Fetch today's xkcd comic', 'NCBI Gi to Kegg Pathways', 'NCBI Gi to Kegg Pathway Descriptions', 'BioAID ProteinDiscovery', and 'Get Pathway-Genes by Entrez gene id' (which is selected).
- Tool Configuration (Center):**
 - Get Pathway-Genes by Entrez gene id**
 - Select source for entrez_gene_id:** 'Type manually' (dropdown).
 - Enter entrez_gene_id:** Text input field containing '3064'.
 - Would you also like the raw results as a zip file:** 'No' (dropdown).
 - Execute** button.
- What it does:** 'Given a specific entrez gene id, returns the pathways that this gene participates in and for each of those pathways which genes are associated with.'
- Inputs:**
 - entrez_gene_id** Give an entrez gene id Examples include:
 - o 3064
- Outputs:**
 - Pathway
 - geneList
- Warnings:**
 - ⚠ **Please note that some workflows are not up-to-date or have dependencies** that cannot be met by the specific Taverna server that you specified during generation of this tool. You can make sure that the workflow is valid by running it in the Taverna Workbench first to confirm that it works before running it via Galaxy.
 - ⚠ **Please note that there might be some repetitions in the workflow description** in some of the generated workflows. This is due to a backwards compatibility issue on the myExperiment repository which keeps the old descriptions to make sure that no information is lost.
- Information:** For more information on that workflow please visit <http://www.myexperiment.org/workflows/2805>.

History Panel (Right): Shows 'Your history is empty. Click 'Get Data' on the left pane to start'.

Footer: Includes a search bar with 'Find: time', navigation links '< Previous > Next', and checkboxes for 'Highlight all' and 'Match case'.

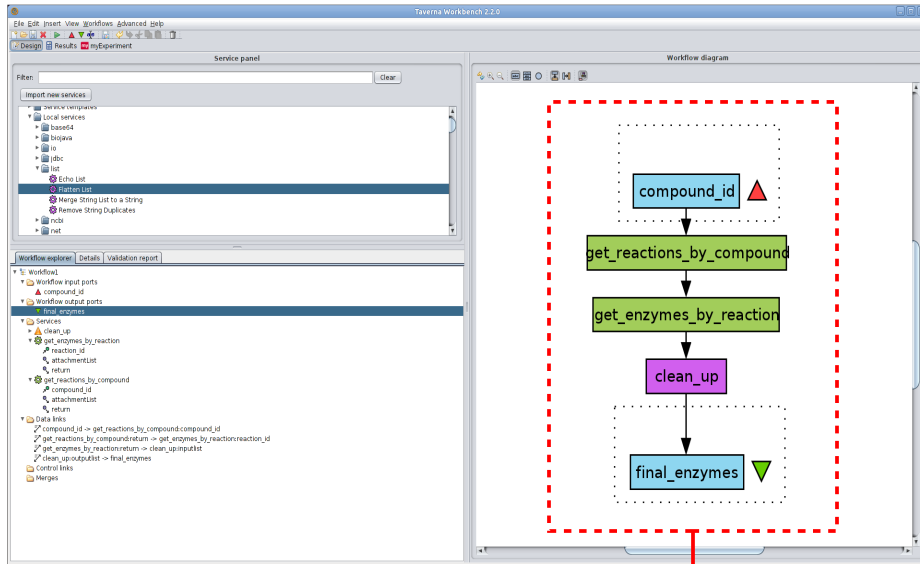
myExperiment

The screenshot shows the myExperiment website in a Mozilla Firefox browser window. The browser's address bar displays the URL <http://www.myexperiment.org/>. The website's header features the myExperiment logo on the left and a yellow banner on the right that reads: "myExperiment makes it easy to **find**, **use** and **share scientific workflows** and other **Research Objects**, and to build **communities**." Below the banner is a search bar with a dropdown menu set to "All" and a "Search" button. The main content area is divided into three columns. The left column, titled "First time visitor? Try these videos:", lists "Project Introduction" and "Bioinformatics Case Study". Below this, under "Use myExperiment to...", are links for "Find Workflows", "Share Your Workflows and Files", "Create and Find Packs of Items", "Find People and Make Friends", and "Create and Join Groups". The middle column features a large "Explore" button, a workflow diagram, and a "Find Workflows" button. The workflow diagram shows a process starting with "Workflow inputs" (query, program, database) leading to "species_title", "chromosome_filter", and "blast_query". These lead to a "blastfilecomposer" box, which then produces "Workflow outputs" (compared_output and blast_output). The right column contains a "Register" button, a "or Login:" section with fields for "Username or Email:" and "Password:", a "Remember me:" checkbox, and an "Or use OpenID:" section with a text field and the example "(eq: name.myopenid.com)".

Our work

- Taverna workflows available in Galaxy
- Taverna workflows available on a web browser
- Demonstration server with Galaxy & Taverna servers, T2Web
 - Galaxy tools/workflows available in Taverna
 - Virtual Image (for production as well)

Taverna Workflows in Galaxy (1)



Galaxy - Mozilla Firefox

File Edit View History Bookmarks Tools Help

~NBSIOAS... Spark | Workflow... | BullonVaul... | Anosixtu... | Problem lo... | Rich Ut Wid... | Outlook W... | Firefox e... | Screencast... | GalaxyTav... | eGalaxy-e... | Galaxy | start_top... | myExperim...

http://localhost:8080/

Galaxy

Analyze Data Workflow Shared Data Help User

Tools Options

manipulation

NGS: Mapping

NGS: Indel Analysis

NGS: RNA Analysis

NGS: SAM Tools

NGS: Peak Calling

NGS: Simulation

SNP/WGA: Data; Filters

SNP/WGA: QC; LD; Plots

SNP/WGA: Statistical Models

Human Genome Variation

VCF Tools

Taverna Workflows

■ FBI_InterProScan for Taverna 2

■ Workflow1

■ Workflow2

■ BioAID_ProteinDiscovery

■ Get enzyme classifications of a compound

Get enzyme classifications of a compound

Select source for compound_id:

Type manually

Enter compound_id:

C15973

Would you also like the raw results as a zip file:

No

Execute

What it does

Given a compound we want to know all the reactions that it participates in so that we get all the enzymes that drive those reactions. It uses KEGG services.

Inputs

- compound_id The compound id (from KEGG). Examples include:
 - C15973

Outputs

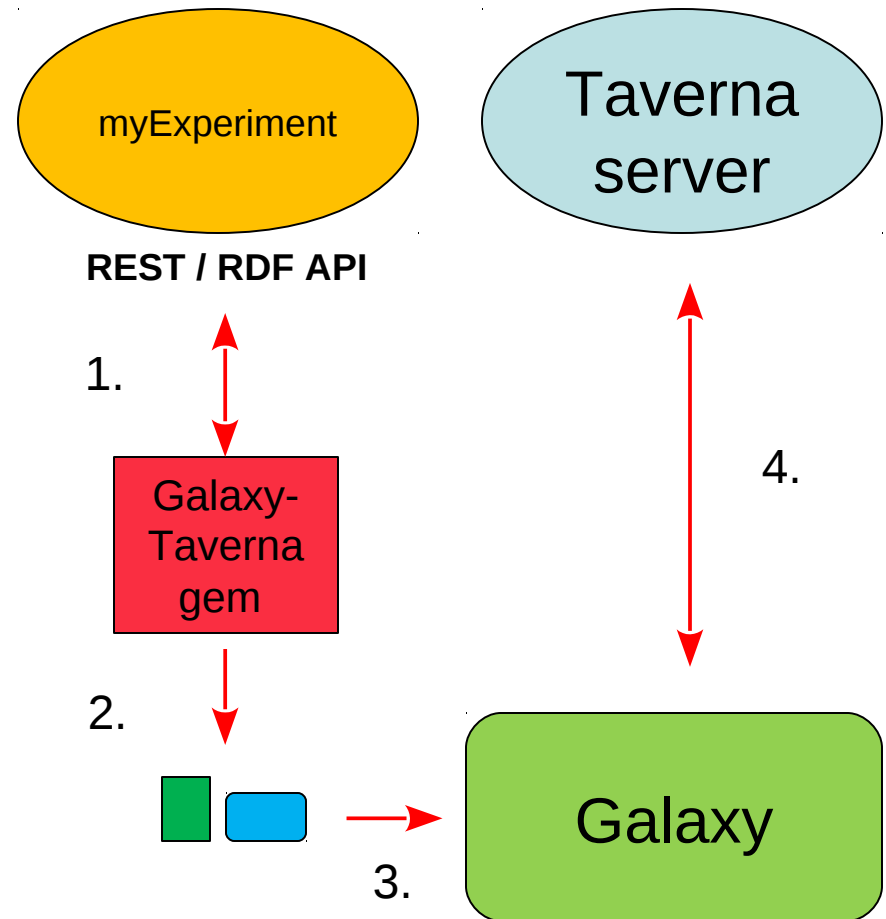
- enzymeClassificationList Enzyme classifications Examples include:
 - ec:2.3.1.168

History Options

Your history is empty. Click 'Get Data' on the left pane to start

Taverna Workflows in Galaxy (2)

- Galaxy-Taverna component
 - ruby gem
 - behind the scenes
 - generates a Galaxy tool
 - requires a workflow description
- Workflow description
 - myExperiment
 - workflow file
- Galaxy
 - tool needs to be manually installed



Taverna Workflows in Galaxy (3)

myExperiment - Workflows - EBI_InterProScan for Taverna 2 (Stian Soiland-Reyes) [Taverna 2 Workflow] - Mozilla Firefox

File Edit View History Bookmarks Tools Help

NBIC Bio... Main Pag... Spark Workflo... BullionVa... Διαδικτυ... Problem ... Rich UI W... Outlook ... Screena... Galaxy-T... eGalaxy... myExp... myGrid... Taverna... The Gala... huntingt...

http://www.myexperiment.org/workflows/820.html

myexperiment About | Mailing List | Publications

Logout | Give us Feedback | Invite

Home Users Groups Workflows Files Packs Services Topics

Home » Workflows » EBI_InterProScan for Taverna 2

Workflow Entry: EBI_InterProScan for Taverna 2

Created at: 26/01/10 @ 14:45:46 Last updated: 24/11/10 @ 10:04:09

| License | Credits (4) | Attributions (5) | Tags (5) | Featured in Packs (1) | Ratings (0) | Attributed By (2) | Favourited By (0) | Citations (0) | Version History | Reviews (0) | Comments (0) |

Version 2 (latest) (of 2) View version: 2 (latest)

Version created on: 26/01/10 @ 14:45:46 by: Stian Soiland-Reyes | Revision comments

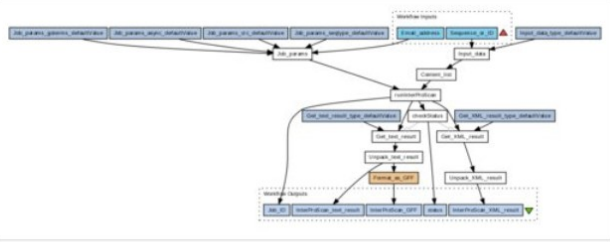
Last edited on: 24/11/10 @ 10:04:09 by: Alan Williams

Title: EBI_InterProScan for Taverna 2

Type: Taverna 2

Preview

(Click on the image to get the full size)



Workflow Type: Taverna 2

Original Uploader: Stian Soiland-Reyes

License: All versions of this Workflow are licensed under: CC BY SA

Credits (4) (People/Groups): Stian Soiland-Reyes, Katy Wolstencroft

New/Upload: Workflow GO

Kostas

My Profile [edit] | My Messages | My Memberships | My History | My News

My Stuff: 1 Friend | 1 Group | 1 Workflows

Friends: Marco Roos

Groups: BioSemantics

Workflows: My Simple get enz...

My Favourites: 0 favourites

Taverna Workflows in Galaxy (4)

The screenshot shows the myExperiment website interface in a Mozilla Firefox browser. The address bar displays <http://www.myexperiment.org/workflows/820.html>. The page title is "myExperiment - Workflows - EBI_InterProScan for Taverna 2 (Stian Soiland-Reyes) [Taverna 2 Workflow]".

The main content area is divided into sections:

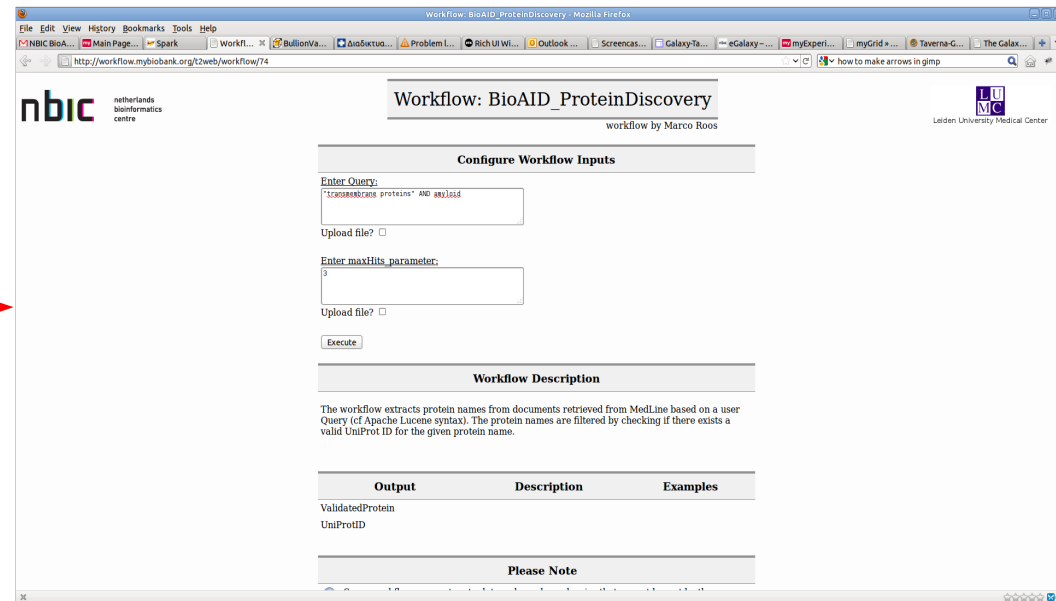
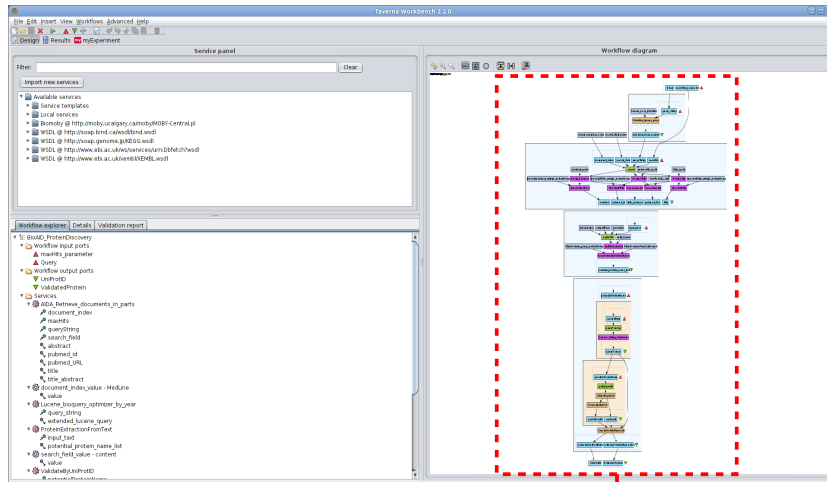
- Description:** Perform an InterProScan analysis of a protein sequence using the EBI's WSInterProScan service (see <http://www.ebi.ac.uk/Tools/webservices/services/interproscan>). The input sequence to use and the user e-mail address are inputs, the other parameters for the analysis (see Job_params) are allowed to default. InterProScan searches a protein sequence against the protein family and domain signature databases integrated into InterPro (see <http://www.ebi.ac.uk/interpro/>). A set of matches to the signatures are returned, which are annotated with the corresponding InterPro and GO term assignments for these signature matches. checkStatus is executed repeatedly as long as the status is equal to RUNNING - check Details -> Advanced for loop condition.
- Download:** Contains two buttons: "Download Workflow File/Package (T2FLOW)" and "Download Workflow as a Galaxy tool". A red dashed box and a red arrow highlight the "Download Workflow as a Galaxy tool" button.
- Run:** Section titled "Run this Workflow in the Taverna Workbench...". It includes "Option 1:" which instructs users to copy and paste a link into File > 'Open workflow location...'. The link is <http://www.myexperiment.org/workflows/820/download?version=2>. There is also a "[More Info]" link.

The right sidebar contains several sections:

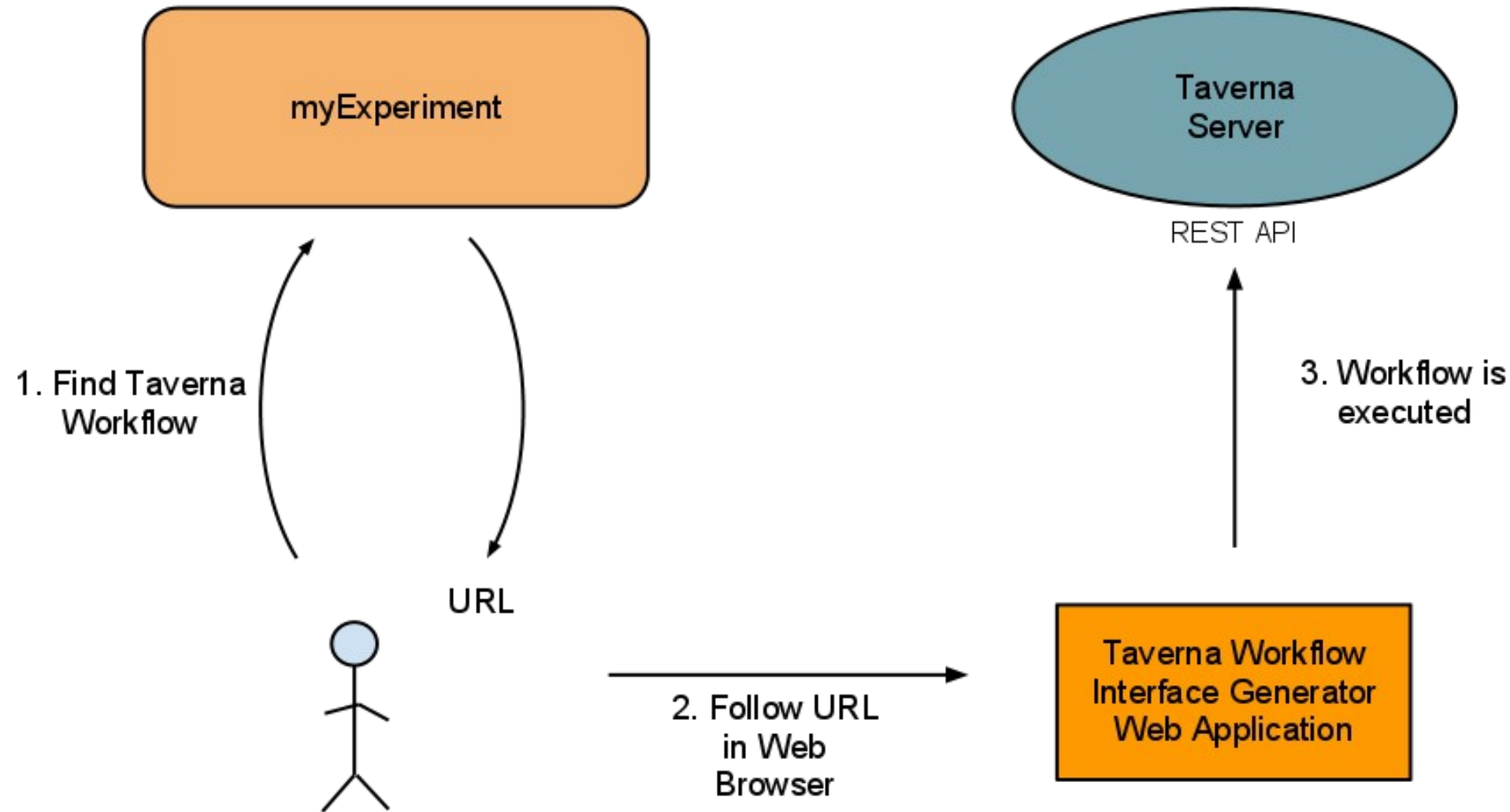
- (People/Groups):** Lists Stian Soiland-Reyes, Katy Wolstencroft, Paolo, and Hamish McWilliam.
- Attributions (5):** Lists EBI_InterProScan for Taverna 2, EBI InterproScan T2, EBI_InterProScan, and EBI InterProScan.
- Tags (5):** Includes a checkbox for "Original Uploader tags" and a list of tags: interpro, interproscan, looping, sequence, term. There is an "Add Tags" button.
- Shared with Groups (1):** Lists myGrid.
- Featured In Packs (1):** Lists Taverna 2.1 beta 2 example workflows.
- Ratings (0):** Shows a star rating system with the text "Current:".

On the far right, there are three panels: "My Favourites" (0 favourites), "My Tags" (0 tags), and "Popular Tags" (25 tags). The "Popular Tags" panel lists various tags such as benchmarks, bio2rdf, bioinformatics, BLAST, cheminformatics, data integration, ebi, example, gene, graph, impact, ist600, kegg, localworker, mygrid, ondex, pathway, pathways, phenotype, protein, pubmed, sequence, taverna, text mining, and workflow.

Taverna Workflows on a web browser (1)



Taverna Workflows on a web browser (2)



Taverna Workflows on a web browser (3)

The screenshot shows a Mozilla Firefox browser window with the URL <http://workflow.mybiobank.org/t2web/workflow/74>. The page title is "Workflow: BioAID_ProteinDiscovery" by Marco Roos. The interface includes a "Configure Workflow Inputs" section with two input fields: "Enter Query:" containing the text `"transmembrane proteins" AND amyloid`, and "Enter maxHits parameter:" containing the value `3`. Both fields have "Upload file?" checkboxes. A red arrow points to the "Execute" button below the second input field. Below this is a "Workflow Description" section with text explaining that the workflow extracts protein names from MedLine documents based on a user query, filtered by UniProt ID. At the bottom, there is a table with headers "Output", "Description", and "Examples". The table lists "ValidatedProtein" and "UniProtID" as outputs. A "Please Note" section is partially visible at the bottom.

Workflow: BioAID_ProteinDiscovery
workflow by Marco Roos

Configure Workflow Inputs

Enter Query:
`"transmembrane proteins" AND amyloid`

Upload file? ☐

Enter maxHits parameter:
`3`

Upload file? ☐

Execute

Workflow Description

The workflow extracts protein names from documents retrieved from MedLine based on a user Query (cf Apache Lucene syntax). The protein names are filtered by checking if there exists a valid UniProt ID for the given protein name.

Output	Description	Examples
ValidatedProtein		
UniProtID		

Please Note

Taverna Workflows on a web browser (4)

Results: - Mozilla Firefox

File Edit View History Bookmarks Tools Help

NBIC BioA... Main Page... Spark Results: BullionVa... Διαδικτυα... Problem L... Rich UI Wi... Outlook ... Screencas... Galaxy-Ta... eGalaxy - ... myExperi... myGrid » ... Taverna-G... The Galax...

http://workflow.mybiobank.org/t2web/enact

how to make arrows in gimp

nbic netherlands bioinformatics centre

Workflow: BioAID_ProteinDiscovery

workflow by Marco Roos

LU MC Leiden University Medical Center

[ValidatedProtein](#)

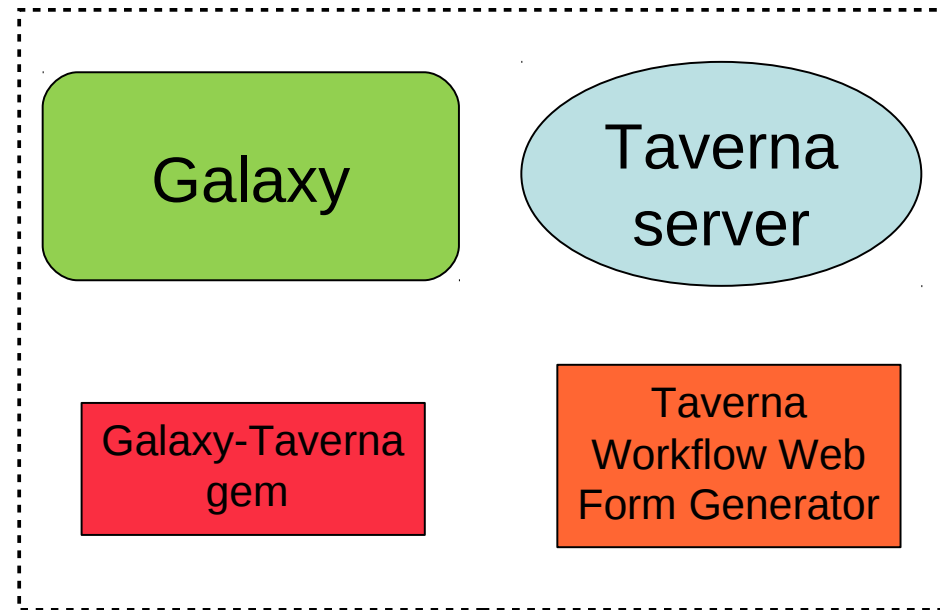
[UniProtID](#)

P70386
Q02527
Q09327
Q10470
Q14CK5
Q6IC49
Q9UH32
P70386
Q02527
Q09327
Q10470
Q14CK5
Q6IC49
Q9UH32
P70386
Q02527
Q09327
Q10470
Q14CK5
Q6IC49
Q9UH32
P70386
Q02527
Q09327
Q10470
Q14CK5
Q6IC49
Q9UH32
A8K7C2
O73815
P02571
P02579
P12714
P14104
P53478
P60010
P63259
P63260
P63261

Transferring data from workflow.mybiobank.org...

Demonstration Server (1)

- Galaxy+Taverna Server 'in one box' and more
 - demonstration
 - preconfigured
 - 'playground'
 - Taverna → Galaxy



Demonstration Server (2)

- Galaxy+Taverna Server 'in one box' and more
- Galaxy Server
 - <http://galaxy.nbic.nl/galaxy>
 - with some example taverna workflows
- Taverna Server
 - <http://galaxy.nbic.nl/demo/taverna-server>
- Taverna workflows to Galaxy tools generator
- Taverna workflows web interface generator web application
 - <http://galaxy.nbic.nl/t2web/workflow/74>

Conclusions

- More interoperable Galaxy – Taverna workflows
- Taverna workflows can be accessed in Galaxy
 - ... and thus take part in a Galaxy workflow
- Taverna workflows can be accessed via the web
 - Bioinformatician creates the workflow
 - ... sends the URL to biologist
- Demonstration Server
 - Virtual machine

Acknowledgements

- NBIC BioAssist developers
- myGrid team (developers of Taverna)
- Galaxy developers
- Users 'in the loop'
 - Eleni Mina
 - Harish Dharuri
 - Pieter Neerincx
 - Jelle Scholtalbers
- Our colleagues at the Human Genetics Department, LUMC, NL
 - BioSemantics group LUMC-Leiden/EMC-Rotterdam
- NBIC-BioAssist/Wf4ever (EU-FP7)

Questions?

- More information
 - **Taverna-Galaxy** → <https://trac.nbic.nl/elabfactory/wiki/eGalaxy>
 - **Taverna-Web** → <https://trac.nbic.nl/elabfactory/wiki/t2web>
 - **Virtual Machine** → https://wiki.nbic.nl/index.php/Galaxy_VM
 - **Demonstration Server** → <http://galaxy.nbic.nl/galaxy>
 - **Galaxy** → <http://galaxy.psu.edu/>
 - **Taverna** → <http://www.taverna.org.uk/>
 - **myExperiment** → <http://www.myexperiment.org/>
- Contact
 - elaboratory-users@trac.nbic.nl (preferred)
 - kostas.karasavvas@nbic.nl
 - m.roos@lumc.nl