



Netherlands
Bioinformatics
Centre

Enacting Taverna Workflows through Galaxy

Konstantinos Karasavvas



Leiden University Medical Center



Introduction

- Software pipeline systems
 - Need to aggregate tools together
 - specify execution ordering
 - handle tools I/Os
 - pipelines, workflows, ...
- Galaxy: web portal and framework for bioinformatics
- Taverna: workflow management system
- myExperiment: scientific social networking web site

Some Background: Galaxy

The screenshot displays the Galaxy web interface. At the top is a navigation bar with links: Analyze Data, Workflow, Data Libraries, Admin, Help, and User. The left sidebar contains a 'Tools' menu with categories like NGS, SNP/WGA, REST client, and MyTools. The main workspace shows the 'BioAID_ProteinDiscovery' workflow configuration. It includes fields for 'Select source for Query' (set to 'Type manually'), 'Enter Query' (containing '"transmembrane proteins" AND amyloid'), 'Select source for maxHits_parameter' (set to 'Type manually'), and 'Enter maxHits_parameter' (set to '10'). There is a checkbox for 'Would you also like the raw results as a zip file?' set to 'Yes' and an 'Execute' button. Below the configuration, a 'What it does' section explains the workflow's purpose, followed by 'Inputs' and 'Outputs' sections. The right sidebar shows a 'History' panel with a list of previous workflow runs, each with a name, a thumbnail, and icons for viewing, deleting, and downloading. The bottom of the interface has a 'Workflows' section.

Galaxy

Analyze Data Workflow Data Libraries Admin Help User

Tools Options

NGS: SAM Tools
NGS: Peak Calling
SNP/WGA: Data; Filters
SNP/WGA: QC; LD: Plots
SNP/WGA: Statistical Models
REST client
MyTools
Taverna Workflows
Taverna BioAID Protein Workflow
Convert column to boolean query
BioAID_ProteinDiscovery
BioMart and Emboss Analysis (T2)
Biomart and EMBOSS analysis
EBI InterProScan for Taverna 2
EBI InterProScan for Taverna 2
Fetch PDB flatfile from RCSB server
Fetch today's xkcd comic
NCBI Gi to Kegg Pathways
What is known about HIV using Bio2RDF's SPARQL endpoints ?
DataBiNS with Kegg ID
G-language Genome Analysis Environment - Reading manuals
Demonstration of configurable iteration
Split text/string into its lines and filter on those to return subset or values

BioAID_ProteinDiscovery

Select source for Query:
Type manually

Enter Query:
"transmembrane proteins" AND amyloid

Select source for maxHits_parameter:
Type manually

Enter maxHits_parameter:
10

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

Execute

What it does

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.

Inputs

- Query Examples include:
 - "transmembrane proteins" AND amyloid
- maxHits_parameter Maximum number of documents to extract proteins from. Use <10 for testing, 100 as default, >100 if you want to live dangerously and can wait (may cause memory problems). Maximum number of documents to extract proteins from. Examples include:
 - 100

Outputs

- ValidatedProtein
- UniProtID

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.

History Options

Unnamed history

- 20: Compressed Results (zip)
- 19: UniProtID
- 18: ValidatedProtein
- 8: status
- 7: InterProScan_GFF
- 6: Job_ID
- 5: InterProScan_XML_result
- 4: InterProScan_text_result
- 3: Compressed Results (zip)
- 2: UniProtID
- 1: ValidatedProtein

Workflows

Some Background: Taverna Workbench

The screenshot displays the Taverna Workbench interface, which is used for designing and executing workflows. The interface is divided into several panels:

- Service panel:** Located at the top left, it contains a filter input, a 'Clear' button, and an 'Import new services' button. Below these, a list of 'Available services' is shown, including 'Service templates', 'Local services', and several 'WSDL' services from various sources like 'moby.ucalgary.ca' and 'ebi.ac.uk'.
- Workflow explorer:** Located at the bottom left, it shows a tree view of the workflow components. It includes 'Workflow input ports' (file_type, output_path, xreflist), 'Workflow output ports' (written_files), and 'Services' (clone_list, copy_number, input, clones, count, list, count, createFileName).
- Workflow diagram:** The central area, titled 'Workflow diagram', shows a complex flowchart of the workflow. It consists of various nodes (represented by colored boxes) connected by arrows, illustrating the sequence of operations and data flow within the workflow.

Some Background: myExperiment

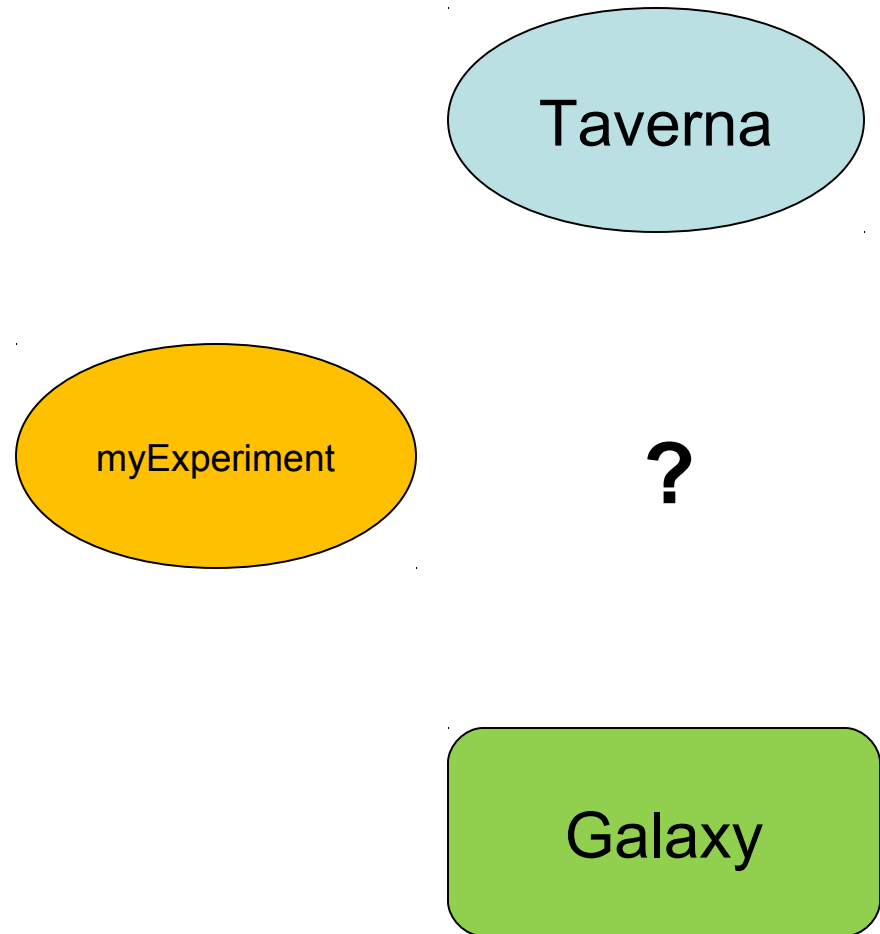
The screenshot shows a web browser window displaying the myExperiment website. The browser's address bar shows the URL <http://www.myexperiment.org/workflows/820.html>. The website has a navigation bar with links for [About](#), [Mailing List](#), [Publications](#), [Logout](#), [Give us Feedback](#), and [Invite](#). Below the navigation bar, there are tabs for [Home](#), [Users](#), [Groups](#), [Workflows](#) (selected), [Files](#), and [Packs](#). A search bar is located below the tabs. The main content area displays the details of a workflow entry titled "Workflow Entry: EBI_InterProScan for Taverna 2". It includes the creation and last update dates, a list of links for [License](#), [Credits](#), [Attributions](#), [Tags](#), [Featured in Packs](#), [Ratings](#), [Attributed By](#), [Favourited By](#), [Citations](#), [Version History](#), [Reviews](#), and [Comments](#). The workflow is currently at Version 2 (latest) of 2, created on 26/01/10 at 14:45:46 by Stian Soiland-Reyes and last edited on 24/11/10 at 10:04:09 by Alan Williams. The workflow type is Taverna 2, and the original uploader is Stian Soiland-Reyes. The license is stated as "All versions of this Workflow are licensed under:". A preview of the workflow diagram is shown at the bottom, with a note to click on the image to get the full size. On the right side of the page, there is a sidebar with a "New/Upload" section, a user profile for Kostas, and sections for "My Profile", "My Messages", "My Memberships", "My History", "My News", "My Stuff", "My Favourites", and "My Groups".

Taverna Workflows in Galaxy: motivation

- Galaxy
 - easy to use for biologists
 - e.g. NGS analysis, ...
- Taverna
 - powerful expressive workflows
 - e.g. text mining, ...
- active communities that add new functionality
 - some overlap but added value if combined
- approaches
 - incorporating Galaxy tools in Taverna
 - incorporating Taverna workflows in Galaxy

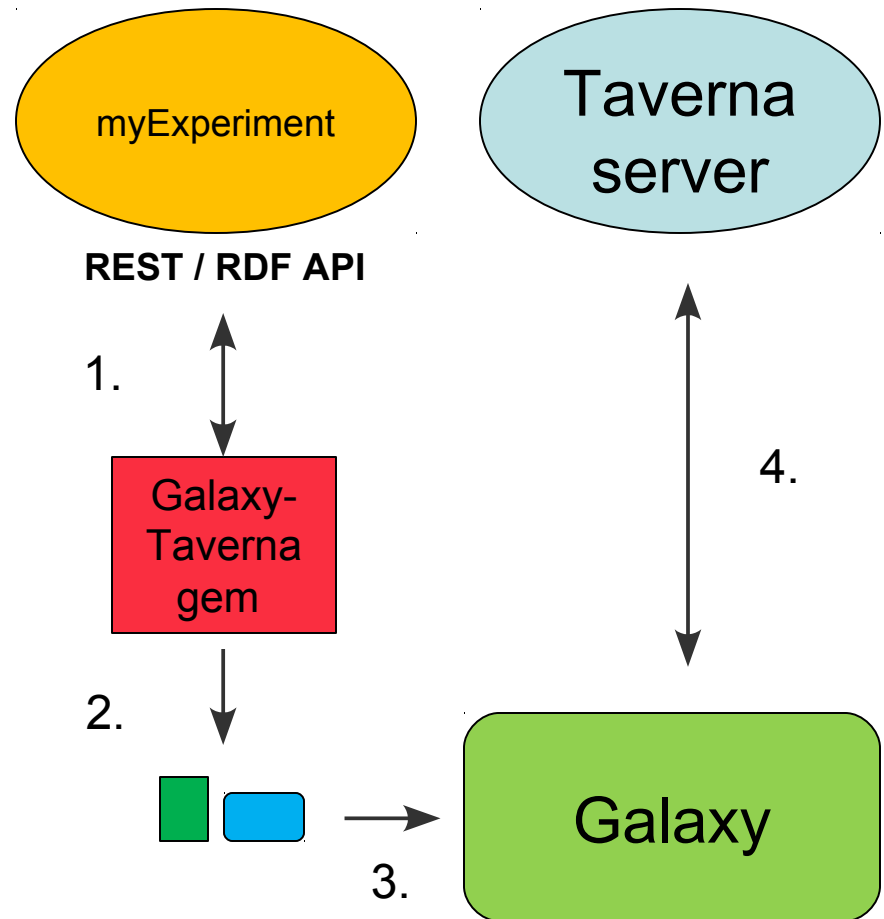
Taverna Workflows in Galaxy: components

- Taverna
 - a server (access)
- myExperiment
 - browse workflows
 - optional
- Galaxy
 - a server (admin)
 - tool to connect to Taverna
 - GUI plus config file
 - program (binary or script)



Taverna Workflows in Galaxy

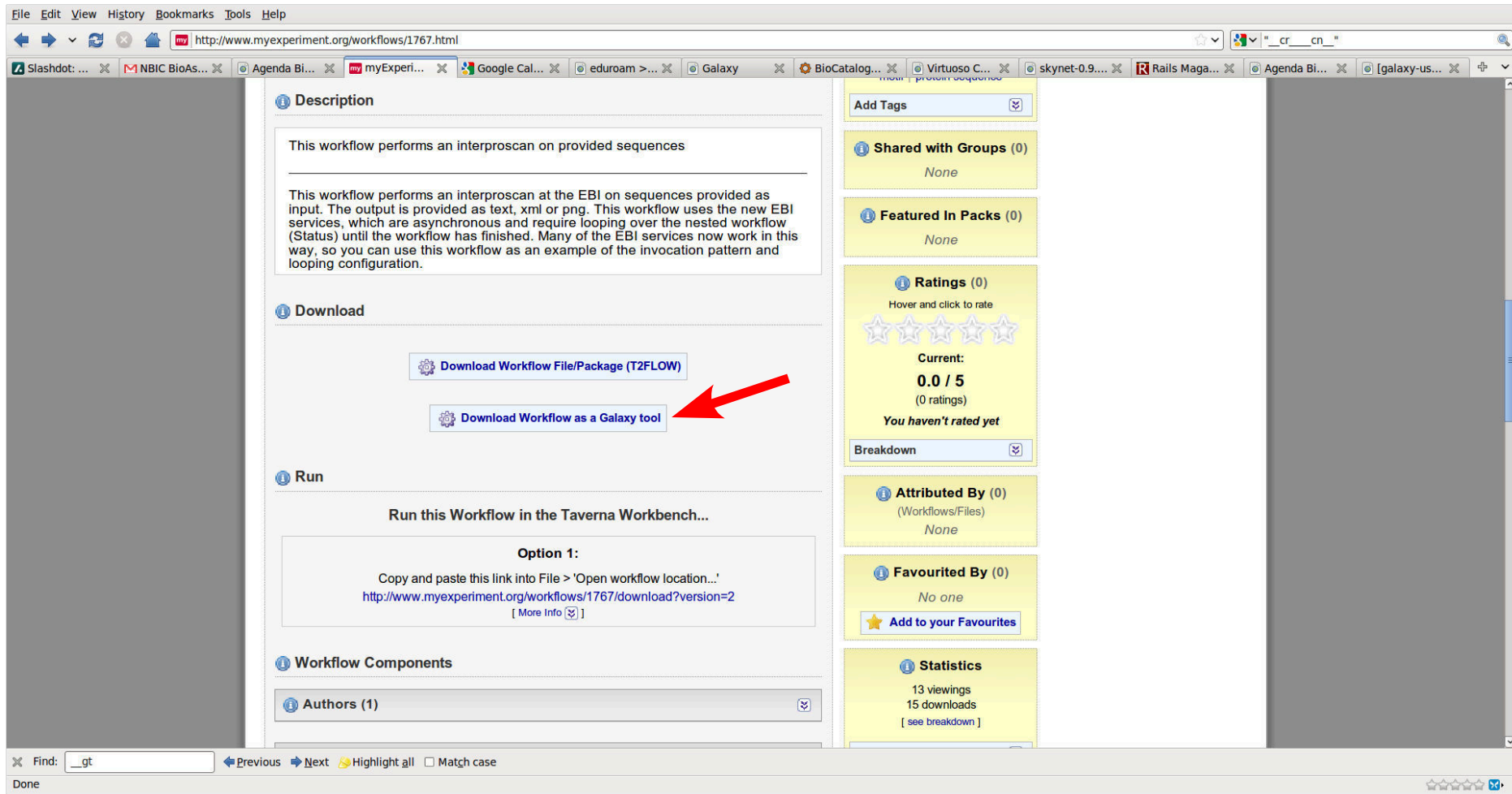
- Galaxy-Taverna component
 - ruby gem
 - behind the scenes
 - generates a Galaxy tool
 - requires a workflow description
- Workflow description
 - myExperiment
 - workflow file (t2flow)
- Galaxy
 - tool needs to be manually installed



myExperiment

The screenshot shows a web browser window with the myExperiment website. The address bar displays <http://www.myexperiment.org/workflows/820.html>. The browser's tab bar includes several open tabs: Slashdot: N..., Gmail - Inb..., NBIC BioAs..., nbic elabfactory, Error, Rake Tutorial, Galaxy, myExperim..., and Google Tra... The website's navigation bar features the myExperiment logo, links for About, Mailing List, and Publications, and user actions like Logout, Give us Feedback, and Invite. Below this is a secondary navigation bar with Home, Users, Groups, Workflows (selected), Files, and Packs. A search bar is positioned to the right of the Workflows tab. The main content area displays the 'Workflow Entry: EBI_InterProScan for Taverna 2'. It includes metadata such as 'Created at: 26/01/10 @ 14:45:46' and 'Last updated: 24/11/10 @ 10:04:09', along with links for License, Credits (4), Attributions (5), Tags (5), Featured in Packs (1), Ratings (0), Attributed By (1), Favourited By (0), Citations (0), Version History, Reviews (0), and Comments (0). The 'Version 2 (latest) (of 2)' section shows the version was created on 26/01/10 by Stian Soland-Reyes and last edited on 24/11/10 by Alan Williams. The title is 'EBI_InterProScan for Taverna 2' and the type is 'Taverna 2'. A 'Preview' section shows a workflow diagram with the instruction '(Click on the image to get the full size)'. The diagram consists of several rectangular nodes connected by arrows, representing a Taverna workflow. On the right side of the page, there is a sidebar with a 'New/Upload' section containing a 'Workflow' dropdown and a 'GO' button. Below this is a user profile for 'Kostas' with links to My Profile, My Messages, My Memberships, My History, and My News. Further down are sections for 'My Stuff' (0 Friends | 0 Groups) and 'My Favourites' (0 favourites). The bottom of the browser window shows a 'Done' status bar.

Download Workflow as a Galaxy Tool



Description

This workflow performs an interproscan on provided sequences

This workflow performs an interproscan at the EBI on sequences provided as input. The output is provided as text, xml or png. This workflow uses the new EBI services, which are asynchronous and require looping over the nested workflow (Status) until the workflow has finished. Many of the EBI services now work in this way, so you can use this workflow as an example of the invocation pattern and looping configuration.

Download

[Download Workflow File/Package \(T2FLOW\)](#)

[Download Workflow as a Galaxy tool](#)

Run

Run this Workflow in the Taverna Workbench...

Option 1:

Copy and paste this link into File > 'Open workflow location...'

<http://www.myexperiment.org/workflows/1767/download?version=2>

[More Info]

Workflow Components

Authors (1)

Add Tags

Shared with Groups (0)

None

Featured In Packs (0)

None

Ratings (0)

Hover and click to rate

☆☆☆☆☆

Current:

0.0 / 5

(0 ratings)

You haven't rated yet

Breakdown

Attributed By (0)

(Workflows/Files)

None

Favourited By (0)

No one

[★ Add to your Favourites](#)

Statistics

13 viewings
15 downloads
[see breakdown]

Taverna Workflows in Galaxy: requirements

- Galaxy server
 - admin access
 - Ruby environment and two gems
 - `$ sudo apt-get install ruby rubygems ruby1.8-dev libxml2-dev`
 - `$ sudo gem install t2-server --version 0.5.3`
 - `$ sudo gem install rubyzip`
- Taverna server
 - for testing, a taverna server is provided
- You can try it out (www.myexperiment.org)
 - trivial if you already have a Galaxy installation

Demo usage

- ISMB Technology track
 - Enacting Taverna Workflows through Galaxy
 - Lounge 1 on Monday, July 18: 12:15 p.m. - 12:40 p.m.

Future Work

- Taverna-Galaxy
 - integrate myExperiment as an external display application
 - requires dynamic loading of new tools
- Taverna workflows via a web interface
 - a web application to configure and run a workflow

- More information

- <https://trac.nbic.nl/elabfactory/wiki/eGalaxy>
- <http://galaxy.psu.edu/>
- <http://www.taverna.org.uk/>
- <http://www.myexperiment.org/>

- Questions?

- kostas.karasavvas@nbic.nl