



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, there is a navigation bar with the 'Galaxy' logo and menu items: 'Analyze Data', 'Workflow', 'Data Libraries', 'Admin', 'Help', and 'User'. On the left side, there is a 'Tools' sidebar with a search box and a list of tool categories including 'NGS: SAM Tools', 'NGS: Peak Calling', 'SNP/WGA: Data; Filters', 'SNP/WGA: QC; LD; Plots', 'SNP/WGA: Statistical Models', 'REST client', 'MyTools', and 'Taverna Workflows'. The main content area shows the configuration for the 'BioAID_ProteinDiscovery' workflow. 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 also a checkbox for 'Would you also like the raw results as a zip file?' set to 'Yes' and an 'Execute' button. Below the configuration, there is a 'What it does' section, an 'Inputs' section with a list of parameters (Query and maxHits_parameter), and an 'Outputs' section with a list of results (ValidatedProtein and UniProtID). A warning icon and text note that some workflows may not be up-to-date or have dependencies. On the right side, there is a 'History' sidebar showing a list of workflow jobs, including '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', and '1: ValidatedProtein'. Each job entry has icons for viewing, deleting, and refreshing.

Some Background: Taverna Workbench

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

- Service panel:** Located at the top left, it contains a menu bar (File, Edit, Insert, View, Workflows, Advanced, Help) and a toolbar. Below the toolbar is a filter input field and a "Clear" button. A prominent "Import new services" button is visible. A list of "Available services" is shown, including "Service templates", "Local services", and several "WSDL" services from various sources like Moby, Soaplab, and KEGG.
- Workflow diagram:** The central area shows a complex flowchart of a workflow. It consists of numerous interconnected nodes, each representing a task or service. The nodes are color-coded (e.g., blue, purple, orange) and connected by arrows indicating the flow of data and control. Some nodes are grouped into larger boxes, possibly representing sub-workflows or specific stages.
- Workflow explorer:** Located at the bottom left, it provides a hierarchical view of the workflow. It shows the "Workflow input ports" (file_type, output_path, xreflist), "Workflow output ports" (written_files), and a list of "Services" used in the workflow, such as "clone_list", "copy_number", "input", "clones", "count", "list", and "count".

Some Background: myExperiment

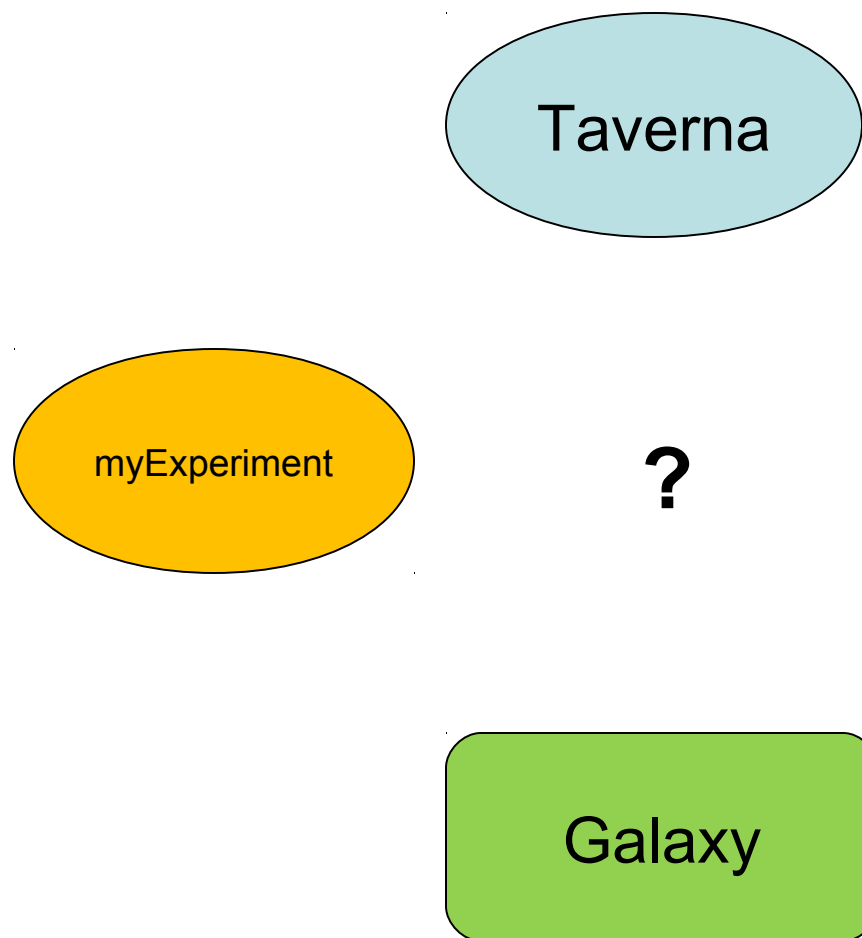
The screenshot shows a web browser window displaying the myExperiment website. The browser's address bar shows the URL <http://www.myexperiment.org/workflows/820.html>. The website's navigation bar includes links for 'About', 'Mailing List', 'Publications', 'Logout', 'Give us Feedback', and 'Invite'. The main content area features a search bar and a navigation menu with options like 'Home', 'Users', 'Groups', 'Workflows', 'Files', and 'Packs'. The 'Workflows' section is active, displaying a workflow entry titled 'Workflow Entry: EBI_InterProScan for Taverna 2'. This entry was created on 26/01/10 at 14:45:46 and last updated on 24/11/10 at 10:04:09. It lists various statistics such as license, credits, attributions, tags, featured in packs, ratings, attributed by, favoured by, citations, version history, reviews, and comments. The workflow is currently at version 2 (latest), created by Stian Soiland-Reyes and last edited by Alan Williams. The workflow type is 'Taverna 2', and the original uploader is Stian Soiland-Reyes. A license section indicates that all versions of this workflow are licensed under a specific license. On the right side, 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', and 'My Favourites'. At the bottom of the workflow entry, there is a preview section with a diagram showing the workflow steps and their connections.

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

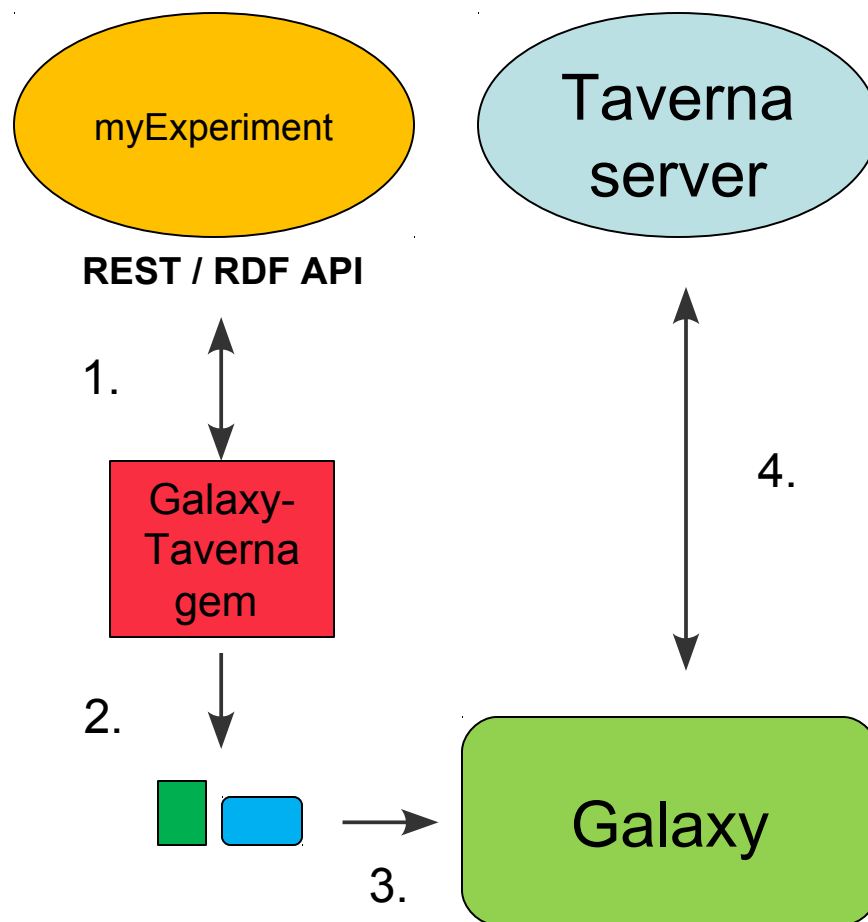
Taverna Workflows in Galaxy: components

- 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



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