



Netherlands
Bioinformatics
Centre

Enacting Taverna Workflows in Galaxy

Dr. Konstantinos Karasavvas



Leiden University Medical Center



Introduction

- Bioinformatics tools
 - large number and diversity
 - more than one tool for the job
 - different access mechanisms
 - non-interoperable
- Experiments need multiple tools
- Need to aggregate tools together
 - specify execution ordering
 - handle tools I/Os
 - pipelines, workflows, ...

Some Background: Galaxy (1)

- Web portal and framework for bioinformatics
 - data sources (UCSC, BioMart, ...) and analysis tools
 - new tools
 - unified intuitive interface
 - other: pipelining, provenance, sharing, open source

Some Background: Galaxy (2)

The screenshot displays the Galaxy web interface. The top navigation bar includes 'Galaxy' and tabs for 'Analyze Data', 'Workflow', 'Data Libraries', 'Admin', 'Help', and 'User'. The left sidebar contains a 'Tools' menu with categories like 'NGS: SAM Tools', 'NGS: Peak Calling', 'SNP/WGA: Data; Filters', 'SNP/WGA: QC; LD; Plots', 'SNP/WGA: Statistical Models', 'REST client', 'MyTools', 'Taverna Workflows', and 'Workflows'. The main panel 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 that the workflow extracts protein names from MedLine documents based on a user query. An 'Inputs' section lists 'Query' and 'maxHits_parameter' with their respective examples and descriptions. An 'Outputs' section lists 'ValidatedProtein' and 'UniProtID'. A warning note at the bottom states: '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.' The right sidebar shows a 'History' panel with a list of previous 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'.

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

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

Some Background: Taverna (1)

- Workflow management system
 - for bioinformatics resources (WS) and other domains
 - flexible and expressive workflow language
 - several extensibility points (e.g. shims and plugins)
 - other: provenance, open source

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 on the left, it contains a search filter, a "Clear" button, and an "Import new services" button. Below these, there is a list of "Available services" including "Service templates", "Local services", and several "WSDL" services from various sources like Biomoby, Soaplab, and EBI.
- Workflow explorer:** Located below the service panel, it shows a tree view of the workflow components. The tree includes "Download pathways for external references list", "Workflow input ports" (with ports like file_type, output_path, xreflist), "Workflow output ports" (with ports like written_files), and "Services" (with services like clone_list, copy_number, input, clones, count, list, and count).
- Workflow diagram:** The main area on the right, it shows a complex flowchart of the workflow. The diagram consists of various nodes (represented by colored boxes) connected by arrows, indicating the sequence and flow of data and control within the workflow. The nodes are organized into a hierarchical structure, with some nodes being sub-workflows themselves.

Some Background: myExperiment (1)

- Scientific social networking web site
 - sharing and re-use of scientific experiments
 - workflow repository for sharing and reuse

Some Background: myExperiment (2)

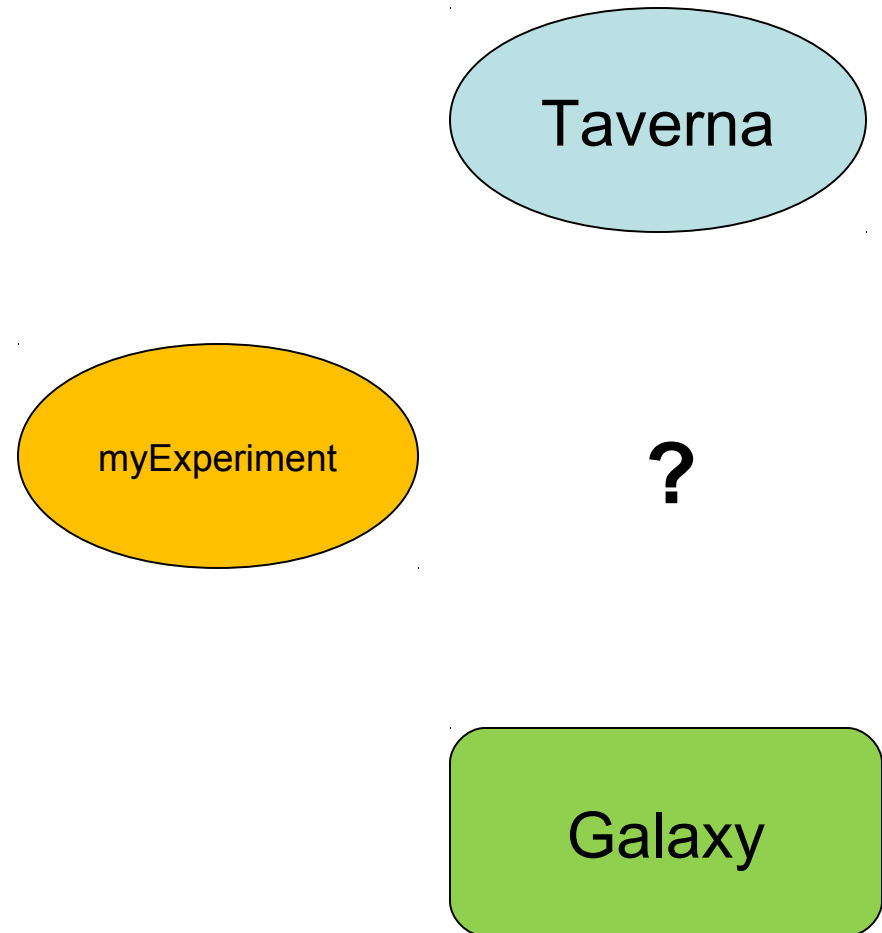
The screenshot shows a web browser window displaying the myExperiment website. The address bar shows the URL <http://www.myexperiment.org/workflows/820.html>. The page features a navigation bar with links for [About](#), [Mailing List](#), and [Publications](#), along with [Logout](#), [Give us Feedback](#), and [Invite](#) buttons. The main content area is titled "Workflow Entry: EBI_InterProScan for Taverna 2" and includes a search bar and tabs for [Home](#), [Users](#), [Groups](#), [Workflows](#), [Files](#), and [Packs](#). The workflow entry details show it was created on 26/01/10 at 14:45:46 by Stian Soiland-Reyes and last updated on 24/11/10 at 10:04:09. It includes 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 workflow type is "Taverna 2" and the original uploader is Stian Soiland-Reyes. A preview of the workflow diagram is shown at the bottom, with a note to click on the image to get the full size. The right sidebar contains a "New/Upload" section, a user profile for Kostas, and links for [My Profile](#), [My Messages](#), [My Memberships](#), [My History](#), and [My News](#). It also shows "My Stuff" with 0 friends and 0 groups, and "My Favourites" with 0 favourites.

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

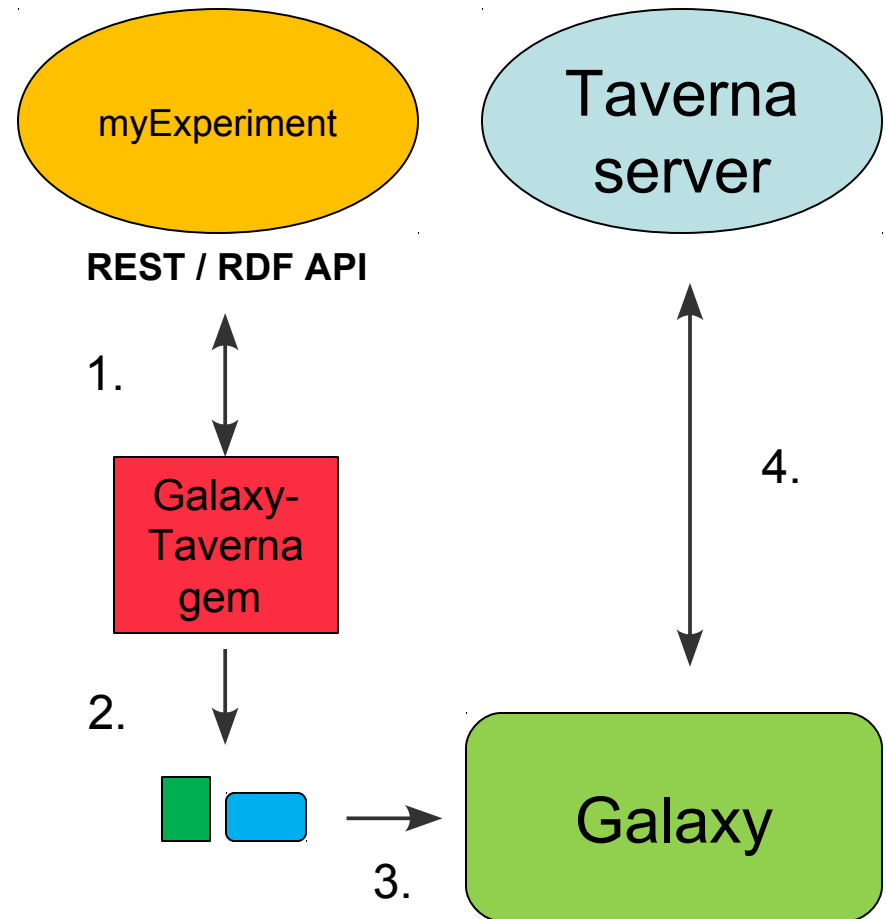
Taverna Workflows in Galaxy: requirements

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



Taverna Workflows in Galaxy: phase one

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



myExperiment

The screenshot displays the myExperiment web application interface within a browser window. The browser's address bar shows the URL <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 myExperiment interface features a top navigation bar with links for About, Mailing List, Publications, Logout, Give us Feedback, and Invite. Below this is a secondary navigation bar with tabs for Home, Users, Groups, Workflows (selected), Files, and Packs. A search bar is located in the center of this bar. 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'. A sidebar on the right contains a 'New/Upload' section with a 'Workflow' dropdown and a 'GO' button, 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'. The main content area also includes a 'Version 2 (latest) (of 2)' section with a 'View version: 2 (latest)' dropdown, a 'Workflow Type' section for 'Taverna 2', an 'Original Uploader' section for 'Stian Soland-Reyes', and a 'License' section. A 'Preview' section at the bottom shows a workflow diagram with the instruction '(Click on the image to get the full size)'. The browser's status bar at the bottom left shows 'Done'.

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

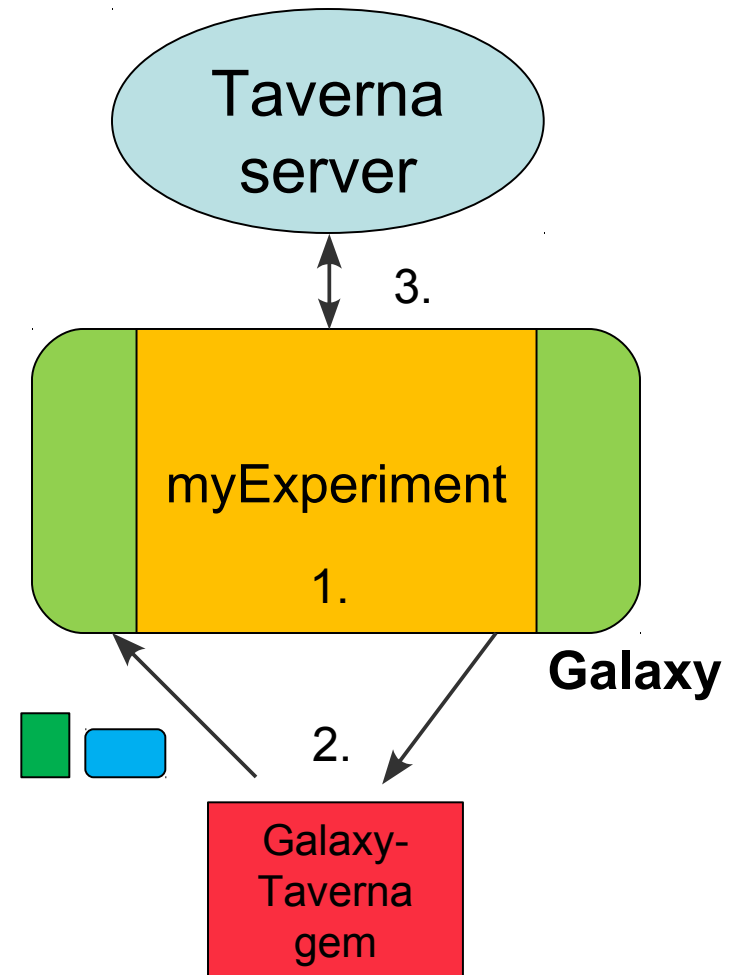
13 viewings
15 downloads
[see breakdown]

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
- Please test (www.myexperiment.org)
 - trivial if you already have Galaxy installed
 - useful taverna workflows for your pipelines
 - I would love to know

Taverna Workflows in Galaxy: phase two

- Future work
- Galaxy
 - integrate myExperiment as an external display app.
- Galaxy new functionality
 - dynamic loading of new tools
 - selected workflows will be generated and installed in one step
- Further investigation is needed



- 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