



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. At the top, the navigation bar includes 'Galaxy', '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', 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 a checkbox for 'Would you also like the raw results as a zip file?' set to 'Yes' and an 'Execute' button. Below the configuration, the 'What it does' section explains that the workflow extracts protein names from MedLine documents based on a user query, filtering by UniProt ID. The 'Inputs' section lists 'Query' and 'maxHits\_parameter' with examples. The 'Outputs' section lists 'ValidatedProtein' and 'UniProtID'. A warning note states that some workflows may not be up-to-date or have dependencies. The right sidebar shows a 'History' panel with 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'.

# 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. 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 search filter 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 explorer:** Located at the bottom left, it shows a tree view of the workflow components. The selected workflow is "Downloadpathwaysforexternalreferenceslist". Underneath, it lists "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 and largest area, titled "Workflow diagram", shows a complex flowchart of the workflow. It consists of numerous interconnected nodes, each representing a task or service. The nodes are color-coded (blue, purple, orange, green) and connected by arrows indicating the flow of data and control. The diagram is organized into several sub-workflows, each enclosed in a light blue box.

# Some Background: myExperiment (1)

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



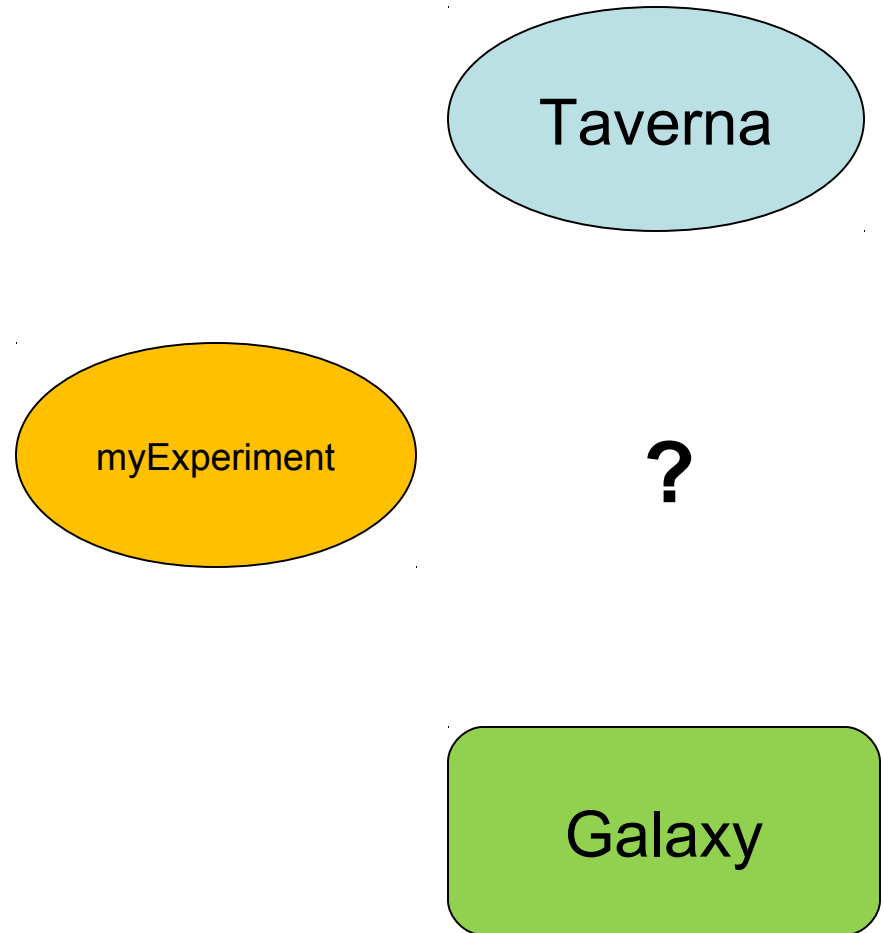


# Taverna Workflows in Galaxy: motivation

- 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

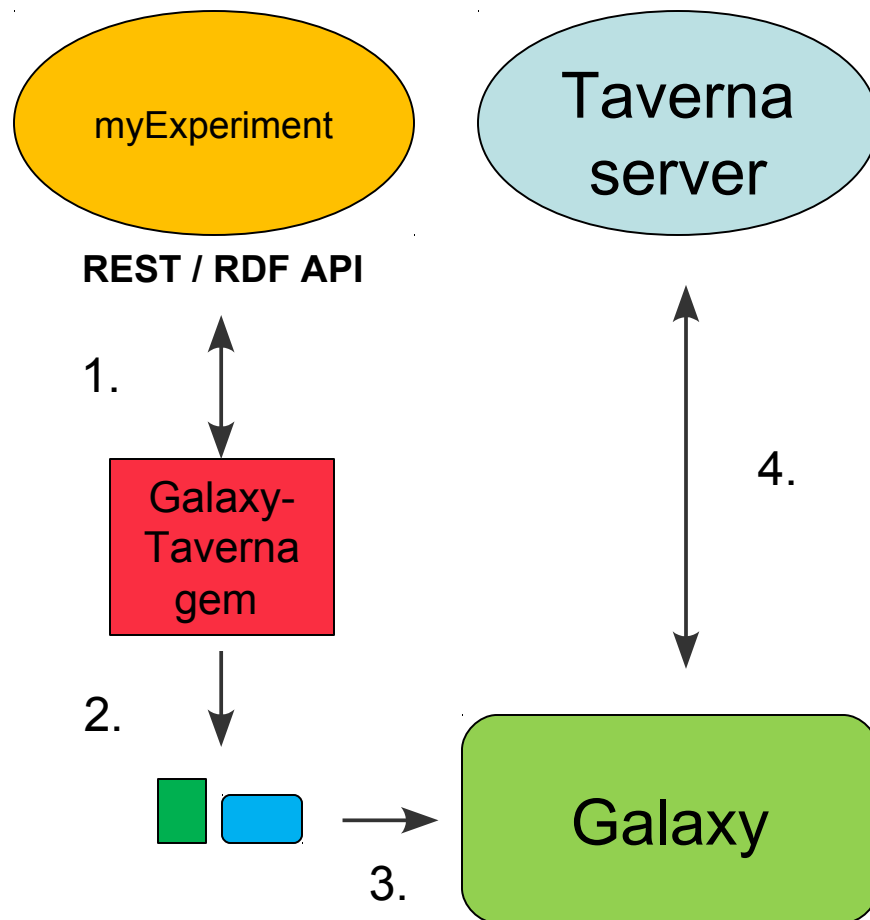
- ruby gem
  - behind the scenes
- generates a Galaxy tool
- requires a workflow description

- Workflow description

- myExperiment
- why?

- Galaxy

- tool needs to be manually installed



# myExperiment

The screenshot 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 header includes the myExperiment logo and navigation links for About, Mailing List, Publications, Logout, Give us Feedback, and Invite. Below the header is a navigation menu with Home, Users, Groups, Workflows, Files, and Packs. The main content area displays the 'Workflow Entry: EBI\_InterProScan for Taverna 2'. It includes a search bar, a breadcrumb trail (Home > Workflows > EBI\_InterProScan for Taverna 2), and a 'New/Upload' sidebar. The workflow entry details show it was created on 26/01/10 and last updated on 24/11/10. It lists the original uploader as Stian Solland-Reyes and provides a preview of the workflow diagram. The sidebar on the right contains a user profile for Kostas with links to My Profile, My Messages, My Memberships, My History, and My News, along with sections for My Stuff (0 Friends, 0 Groups) and My Favourites (0 favourites).

# Download Workflow as a Galaxy Tool

The screenshot shows a web browser window with the URL <http://www.myexperiment.org/workflows/1767.html>. The page content is as follows:

- 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:** Two buttons are present: "Download Workflow File/Package (T2FLOW)" and "Download Workflow as a Galaxy tool". A red arrow points to the second button.
- Run:** Section titled "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)

The right sidebar contains the following information:

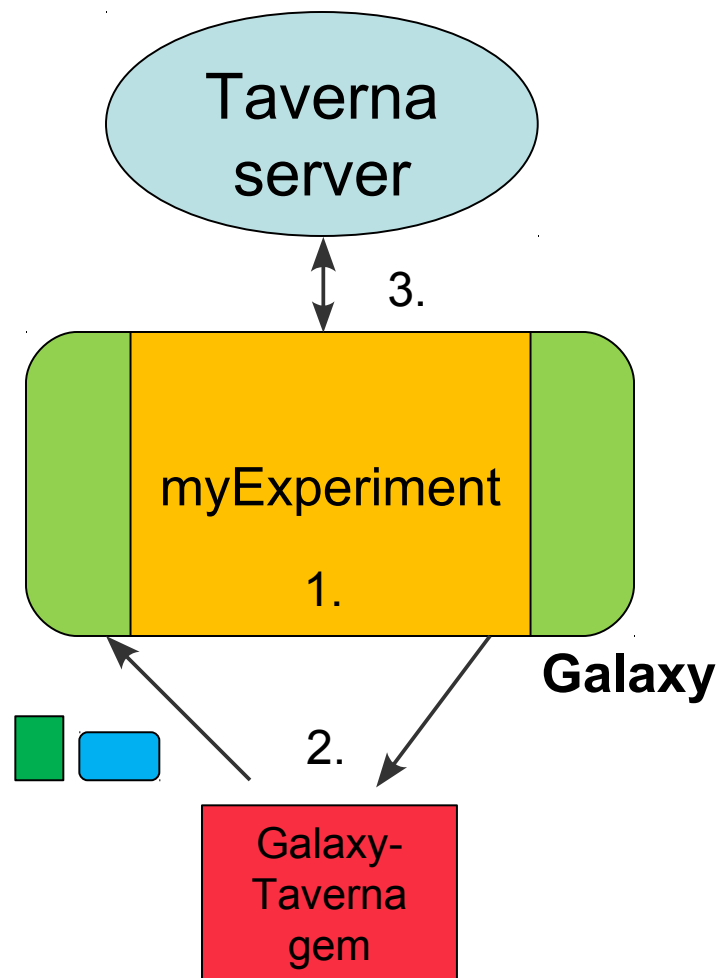
- Add Tags:** [Dropdown]
- Shared with Groups (0):** None
- Featured In Packs (0):** None
- Ratings (0):** Hover and click to rate. Current: 0.0 / 5 (0 ratings). You haven't rated yet.
- Breakdown:** [Dropdown]
- Attributed By (0):** (Workflows/Files) None
- Favourited By (0):** No one. Add to your Favourites [Star icon]
- 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](http://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](mailto:kostas.karasavvas@nbic.nl)