

# Enacting Taverna Workflows in Galaxy

Dr. Konstantinos Karasavvas  
Senior E-Lab Engineer



# Introduction

- Bioinformatics tools
  - large number and diversity
  - more than one tool for the job
  - different access mechanisms
  - non-interoperable
- E-Laboratory Components
  - reusability (of existing and new components)
  - interoperability (open source, standards, ...)
  - system integration (possible at several layers)

# Some Background

- Galaxy: web portal and framework for bioinformatics
  - data sources (UCSC, BioMart, ...) and analysis tools
  - new tools
  - unified intuitive interface
  - other: pipelining, provenance, sharing, open source
- Taverna: 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
- myExperiment: web site
  - workflow repository for sharing and reuse

# Galaxy

The screenshot shows a web browser window with the Galaxy interface. The address bar shows `http://localhost:8080/`. The browser tabs include "Galaxy" and "myExperiment...". The Galaxy navigation bar has "Analyze Data", "Workflow", "Data Libraries", "Help", and "User".

**Tools Panel (Left):**

- FASTA manipulation
- NGS: QC and manipulation
- NGS: Mapping
- NGS: Indel Analysis
- NGS: Expression Analysis
- 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)
  - EBI InterProScan for Taverna 2
  - Fetch PDB flatfile from RCSB server
  - Fetch today's xkcd comic

**Main Panel (Center):**

### BioAID\_ProteinDiscovery

**Enter Query:**

**Enter maxHits\_parameter:**

**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

For more information on that workflow please visit <http://www.myexperiment.org/workflows/74>.

**History Panel (Right):**

- 92: todaysXkcd
- 86: pdbFlatFile
- 81: RNorIDs
- 80: MMusIDs
- 79: HSapIDs
- 78: outputPlot
- 77: status
- 76: InterProScan\_GFF
- 75: Job\_ID
- 74: InterProScan\_XML\_result
- 73: InterProScan\_text\_result
- 2: Validated Protein
- 1: UniProt IDs

URL at the bottom: `http://localhost:8080/tool_runner?tool_id=BioAID_ProteinDiscovery_id`

# Taverna2 Workbench

The screenshot displays the Taverna2 Workbench interface, which is divided into several main sections:

- Service panel:** Located on the left, it features a search filter, a "Clear" button, and an "Import new services" button. Below these are "Available services" categorized into "Service templates", "Local services", and several "WSDL" entries from various sources like <http://moby.ucalgary.ca/moby/MOBY-Central.pl>, <http://www.ebi.ac.uk/soaplab/services/>, and <http://www.ebi.ac.uk/xembl/XEMBL.wsdl>.
- Workflow explorer:** Located below the service panel, it shows a tree view of the workflow components, including "Download pathways for external references list", "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, showing a complex flowchart of interconnected workflow components. The diagram includes nodes for "Workflow input ports", "Workflow output ports", and various service calls, all connected by arrows indicating the flow of data and control.

# 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 menu with options like Home, Users, Groups, Workflows, Files, and Packs. The main content area displays the details for a workflow entry titled "Workflow Entry: 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 \(1\)](#) | [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)

The workflow diagram shows a sequence of steps: job\_params\_generic\_definitive, job\_params\_spec\_definitive, job\_params\_cc\_definitive, job\_params\_spec\_definitive, Email\_address, Sequence\_of\_1, job\_params\_definitive, and Content\_mf. The steps are connected by arrows, indicating a flow from left to right.

**Workflow Type: Taverna 2**

**Original Uploader:** [Stian Soiland-Reyes](#)

**License:** All versions of this Workflow are licensed under:

**My Profile** [edit]

- My Messages
- My Memberships
- My History
- My News

**My Stuff**

0 Friends | 0 Groups

**My Favourites**

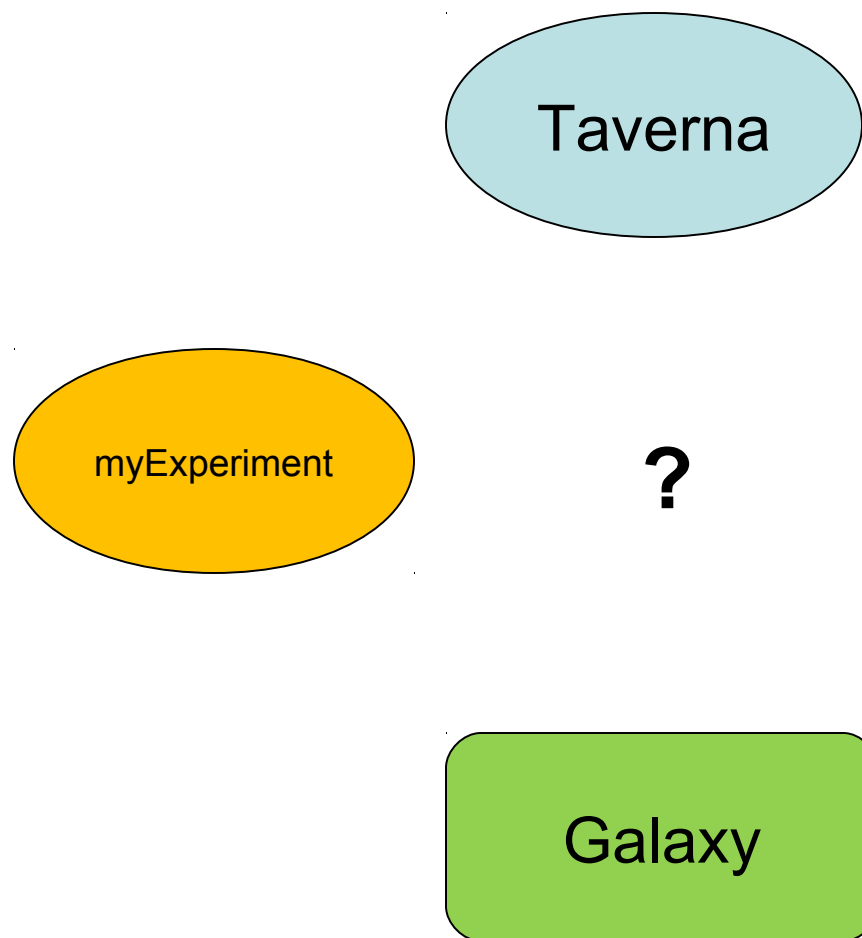
0 favourites

# 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

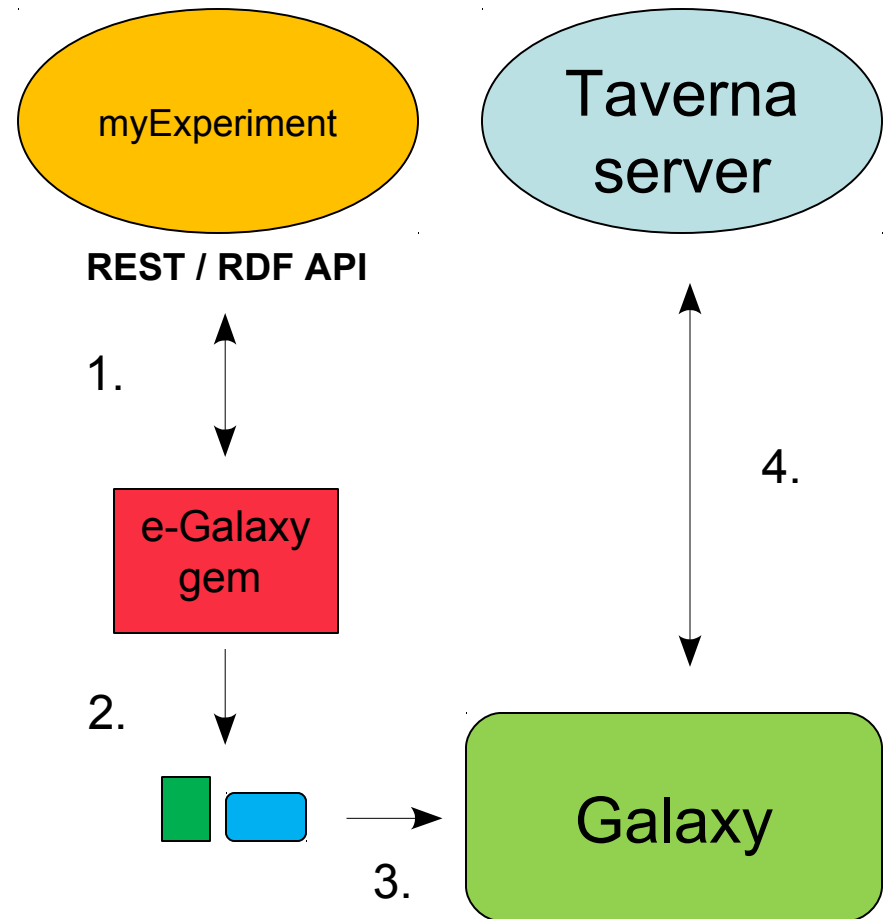
- Taverna
  - a server (access)
- myExperiment
  - browse workflows
- Galaxy
  - a server (admin)
  - the new tool
    - GUI plus config file
    - program (binary or script)





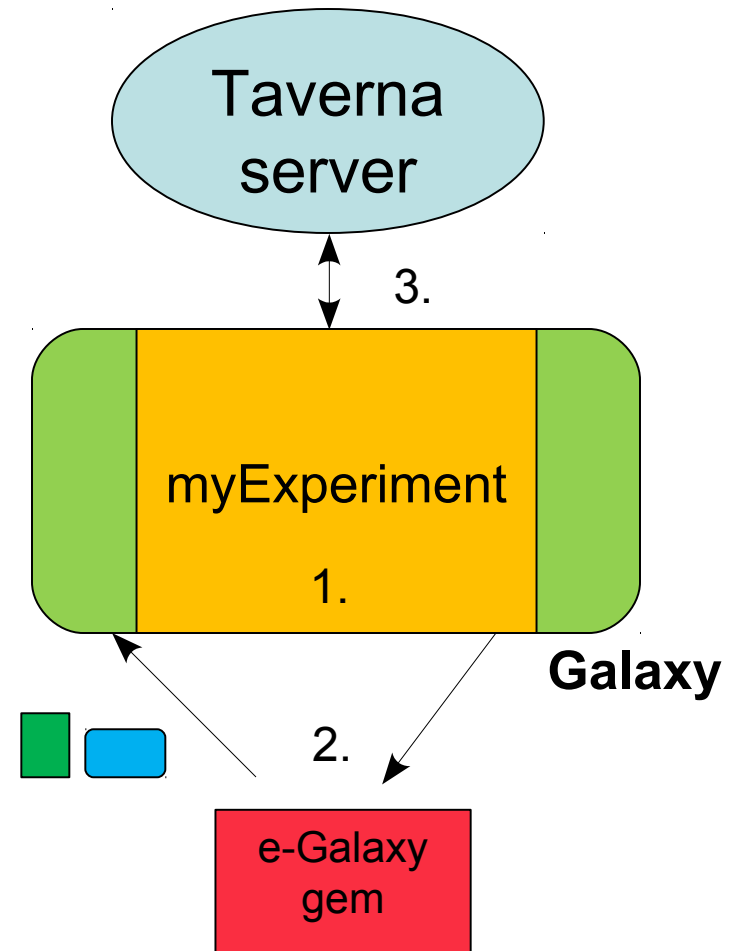
# Taverna Workflows in Galaxy: phase one

- Ruby gem
  - generates a Galaxy tool
  - requires a workflow description
- Workflow description
  - myExperiment
  - why?
- Galaxy
  - tool needs to be manually installed
- TODO
  - part of myExperiment



# Taverna Workflows in Galaxy: phase two

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



- More information

- <http://galaxy.psu.edu/>
- <http://www.taverna.org.uk/>
- <http://www.myexperiment.org/>

- Questions?