Enacting Taverna Workflows in Galaxy

Dr. Konstantinos Karasavvas
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 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.
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
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)
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

Diagram:

1. myExperiment
   - REST / RDF API
   - Galaxy-Taverna gem
2. Galaxy-Taverna gem
3. Galaxy
4. Taverna server
myExperiment
Download Workflow as a Galaxy Tool

This workflow performs an interproscan on provided sequences. The 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.

Option 1:
Copy and paste this link into File > 'Open workflow location...':
http://www.myexperiment.org/workflows/1767/download?version=2

Authors (1)
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