Enacting Taverna Workflows through Galaxy

Konstantinos Karasavvas
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 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 Workbench
Some Background: myExperiment
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

myExperiment
REST / RDF API
Galaxy-Taverna gem
Galaxy
Taverna server
myExperiment
Download Workflow as a Galaxy Tool
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