

Apache Taverna – Scientific workflows for reproducible research

Authors: Denis Karyakin, Apache Taverna team (<http://taverna.incubator.apache.org/about/>)

Funding by the Google Summer of Code

E-mail: samhane.me@gmail.com

<http://taverna.incubator.apache.org/>



INTRODUCTION

A **Workflow Management System (WMS)** is a piece of software that provides an infrastructure to setup, execute, and monitor scientific workflows. In other words, the WMS provides an environment where **in silico experiments** can be defined and executed.

Apache Taverna is an open source domain independent Workflow Management System - a suite of tools used to design and execute scientific workflows. The Taverna suite includes **the Taverna Engine, Taverna Workbench and Taverna Server**. Taverna Workbench provides a range of features enabling scientists to design, execute and share workflows.

The Taverna Suite of Tools



DESIGN AND CONSTRUCTION

Taverna Workbench provides a graphical environment where researchers can drag and drop local and distributed services and data resources to design and build scientific data-driven workflows, producing a single explicit experimental protocol

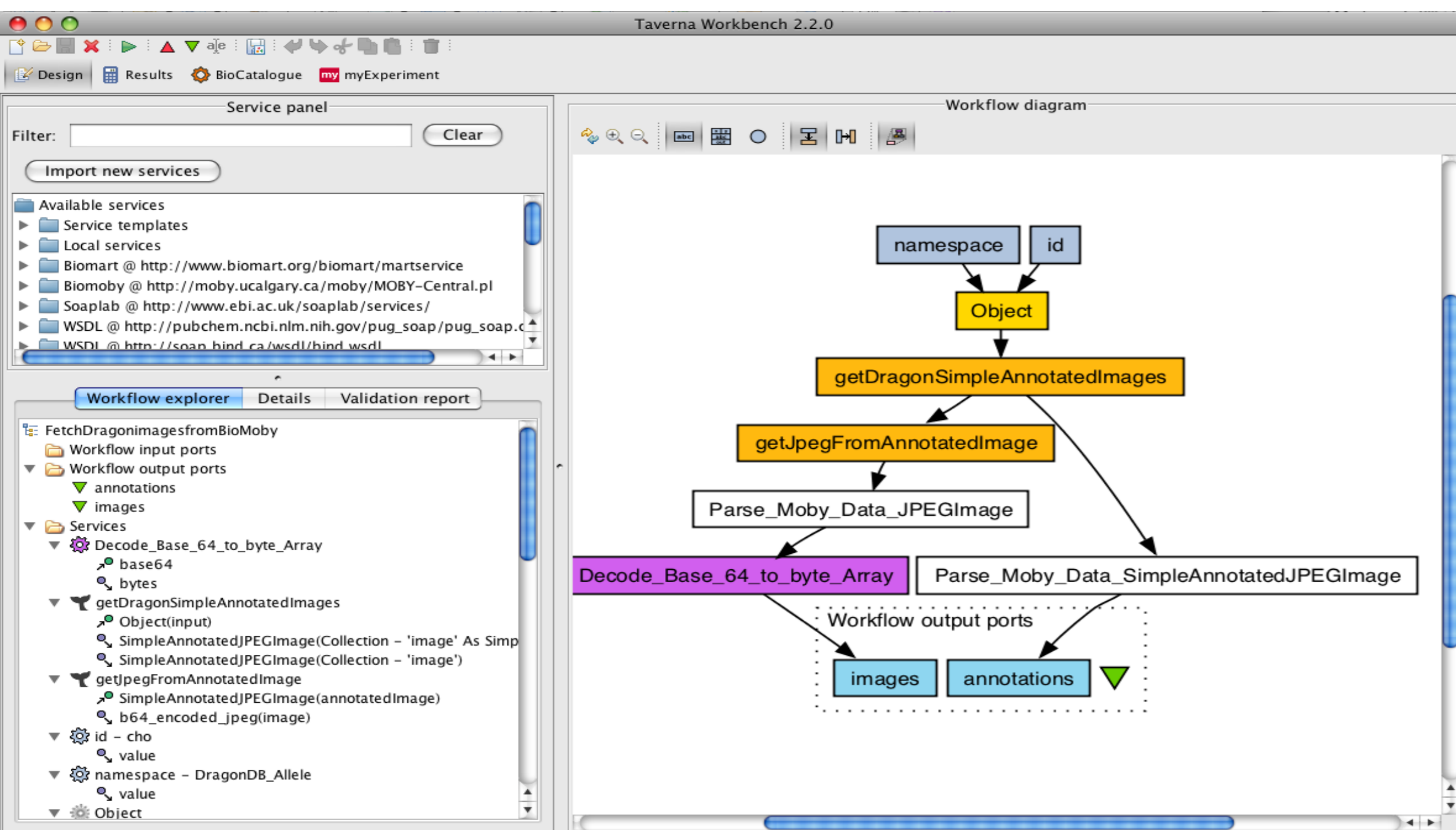
Standard web services (SOAP or REST)

Invoke any command line tool or script (local/SSH)

Secure invocations with credentials (HTTPS/SSL, SSH)

Scripting (Beanshell, R)

Pluggable components of reusable workflows



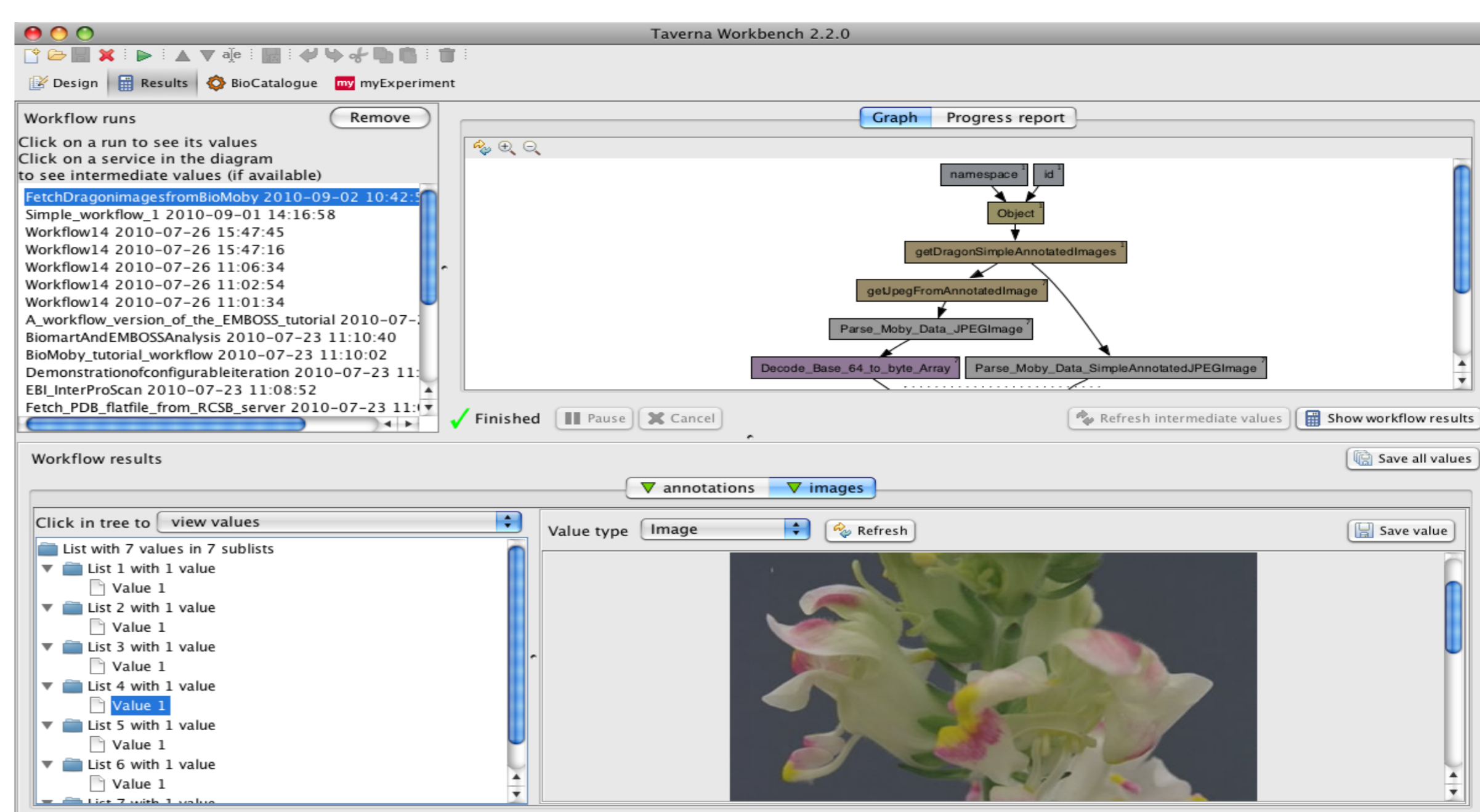
EXECUTION

Graphical workflow execution in the Taverna Workbench shows the progress of a workflow run.

Command line execution allows Workflows to also be executed on the command line, integrating with other software (e.g. Galaxy).

Remote workflow execution using Taverna Server allows REST integration of workflows with third-party platforms, **web applications** and **mobile apps**.

Full execution provenance and intermediate values captured for workflow runs, exported as research object bundle.



SHARING AND DISCOVERY

Workflows created with Apache Taverna can be **published, shared and reused**. Workflows can be re-used and re-purposed to suit the needs of an individual researcher, reducing workflow re-invention whilst facilitating scientific collaboration and sharing of research expertise.

myExperiment is a social web site for researchers sharing **Research Objects** such as Scientific Workflows. myExperiment has a REST API.

The myExperiment website was launched in November 2007 and contains a significant collection of scientific workflows for a variety of workflow systems, most notably Taverna, but also other tools such as Bioclipse. myExperiment facilitates the re-use and re-purposing of workflows through its large and active user communities.

BioCatalogue is a registry of web services, which can be annotated, discovered, tagged, curated and monitored. Open source and generic, can be customized to set up your own web service repository focused on your domain.