

Building workflows with Apache Taverna

Stian Soiland-Reyes
University of Manchester

Including materials by Katy Wolstencroft,
Aleksandra Pawlik, Christian Brenninkmeijer

<http://orcid.org/0000-0001-9842-9718>
<http://orcid.org/0000-0002-2937-7819>
<http://orcid.org/0000-0002-1279-5133>
<http://orcid.org/0000-0001-8418-6735>



This work is licensed under a
Creative Commons Attribution 4.0 International License



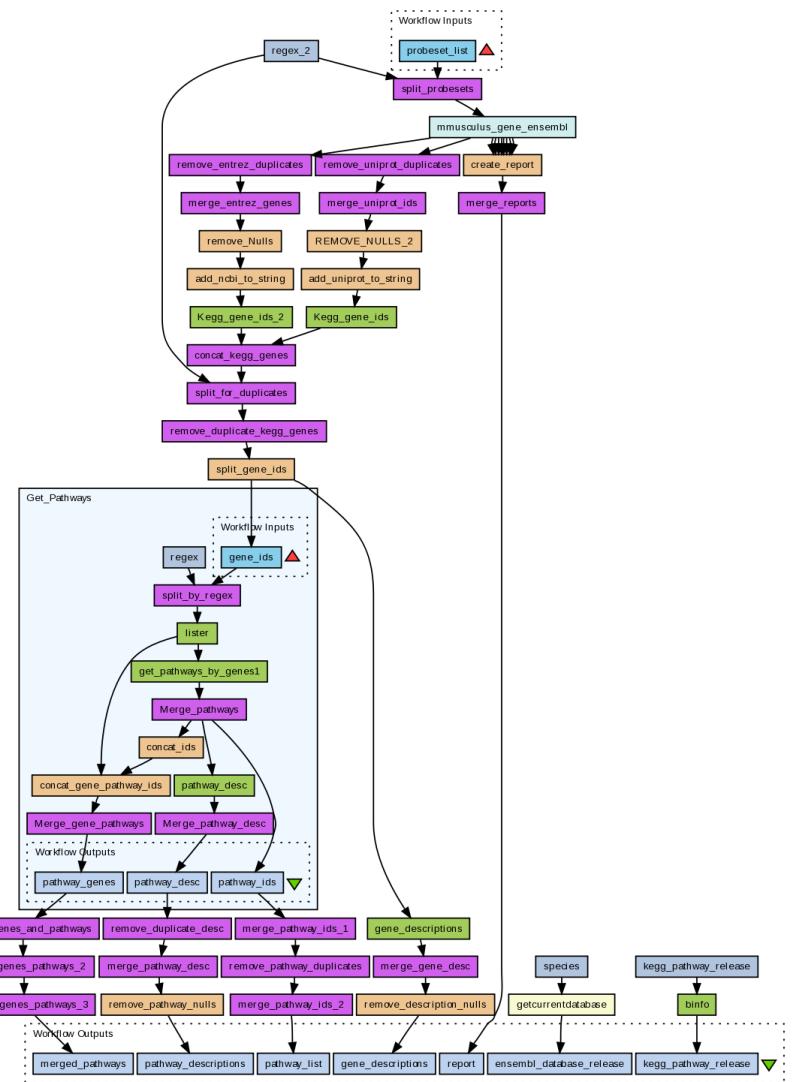
Apache Taverna

Barcelona, 2016-10-20
<http://taverna.incubator.apache.org/>

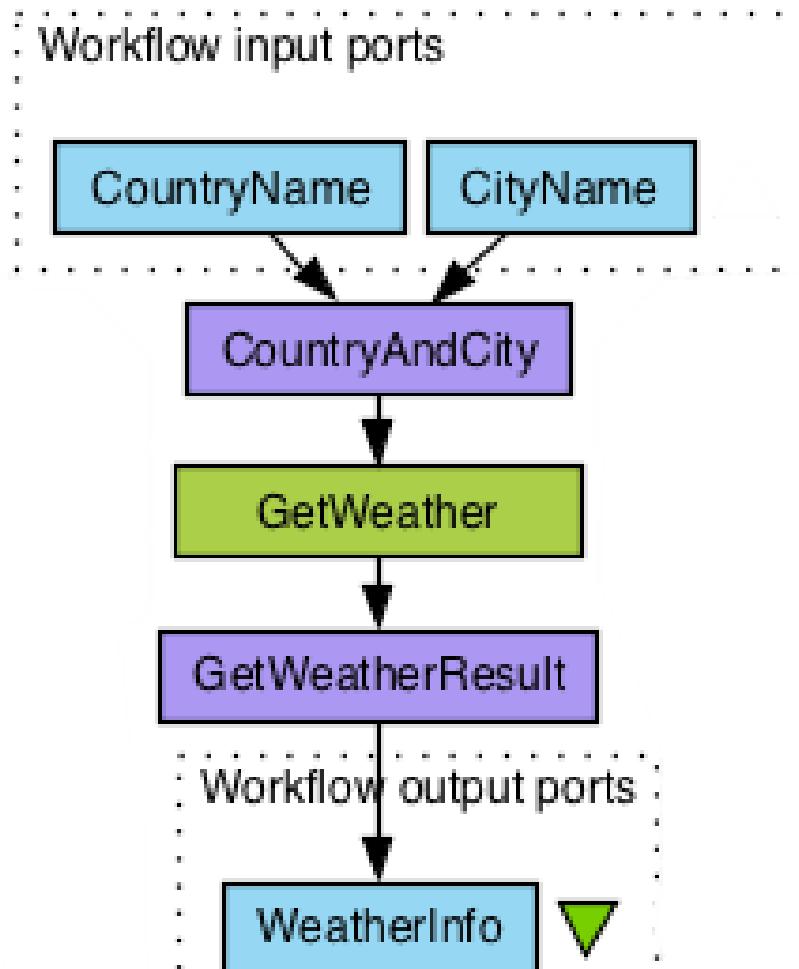


Taverna workflows

- Sophisticated analysis pipeline
- Graphical representation of executable analysis
- Combine a set of **services** to analyse or manage data (local or remote)
- Data **flow** from one service (boxes) to the next (connected with arrows)
- **Iteration** - process multiple data items
- Automation – **rerun** workflow



Example Taverna Workflow



Workflow: Get the weather forecast of the day given the city and the country

Green box is a **Web Service**

Purple boxes are local **XML** services to assemble/ extract XML

Blue boxes are workflow **input** and **output** ports



- **Flow of data** from one tool to the next is automatic - just connect inputs and outputs
- Incompatibilities overcome in the workflow with helper services (*shims*)
 - Allowing new tool combinations
- Workflow engine records parameter values and algorithms – **provenance**
- Workflows can include data **integration** and **visualization**
- **Iteration** over large data sets automatic - ideal for high throughput analysis (e.g. omics)



Taverna Workbench

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

Freely available,
open source

57,000+ downloads
(workbench 2.5.0)

Installers for Windows,
Mac OS X, Linux

Versions:
2.5.0 (workbench)
3.1.0 (command line)
2.5.4 (server)

 Apache Taverna Introduction ▾ Documentation ▾ Download ▾ Community ▾ ASF ▾ Search

Apache Taverna (incubating)

ⓘ Taverna has transitioned to the Apache Incubator. This website is for the Apache Taverna project.
You might still find links to the earlier taverna.org.uk website which is not affiliated with Apache and provided for archive purposes only.

ⓘ Apache Taverna Command Line 3.1.0 has now been released.

Taverna Workflow System

Powerful, scalable, open source & domain independent tools for designing and executing workflows
3500+ resources.

[Download](#)

[Follow](#)

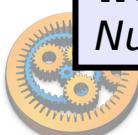
Build scientific workflows

Taverna is an open source and domain-independent Workflow Management System –
a suite of tools used to design and execute *scientific workflows* and aid *in silico*

Taverna has moved to the Apache

Taverna has been accepted as an Apache Incubator project

Wolstencroft et al. (2013): **The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud**,
Nucleic Acids Research, **41**(W1): W557-W561. doi:[10.1093/nar/gkt328](https://doi.org/10.1093/nar/gkt328)



Apache Taverna

Taverna Workflow System

History:

2003: Taverna 0.1

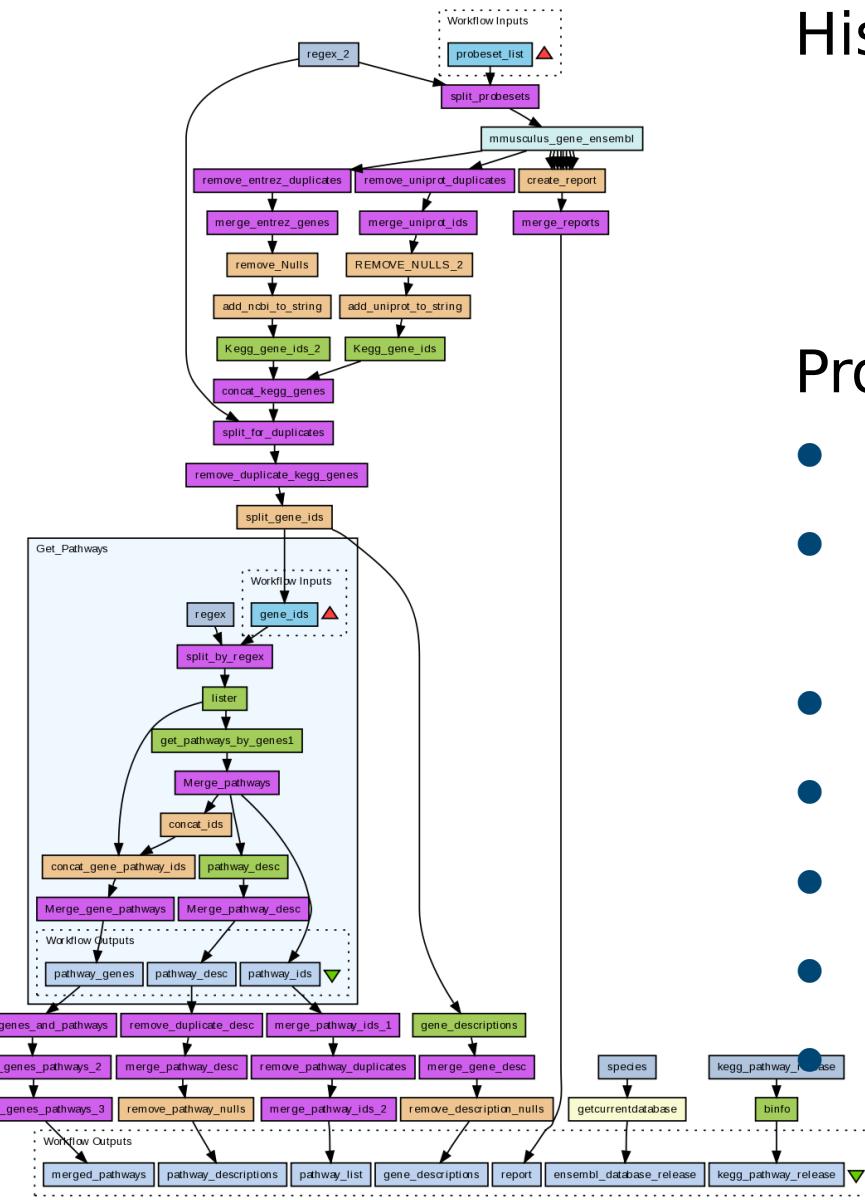
2014: Taverna 2.5.0

2016: Apache Taverna 3.1.0

Products:

- Apache Taverna Command-line
- Apache Taverna Engine / Language (API)
- Taverna Workbench
- Taverna Server
- Taverna Player
- Taverna Mobile

Plugins and integrations



Taverna editions and extensibility

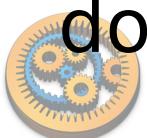
Taverna is a generic workflow system that can be extended by **plugins** and customized for use in different domains.

- Core
- Astronomy
- **Bioinformatics**
- Biodiversity
- Digital Preservation
- Enterprise

The Taverna **editions** are pre-built downloads of Taverna with plugins for the most popular domains.

<http://www.taverna.org.uk/download/workbench/>

2-5/



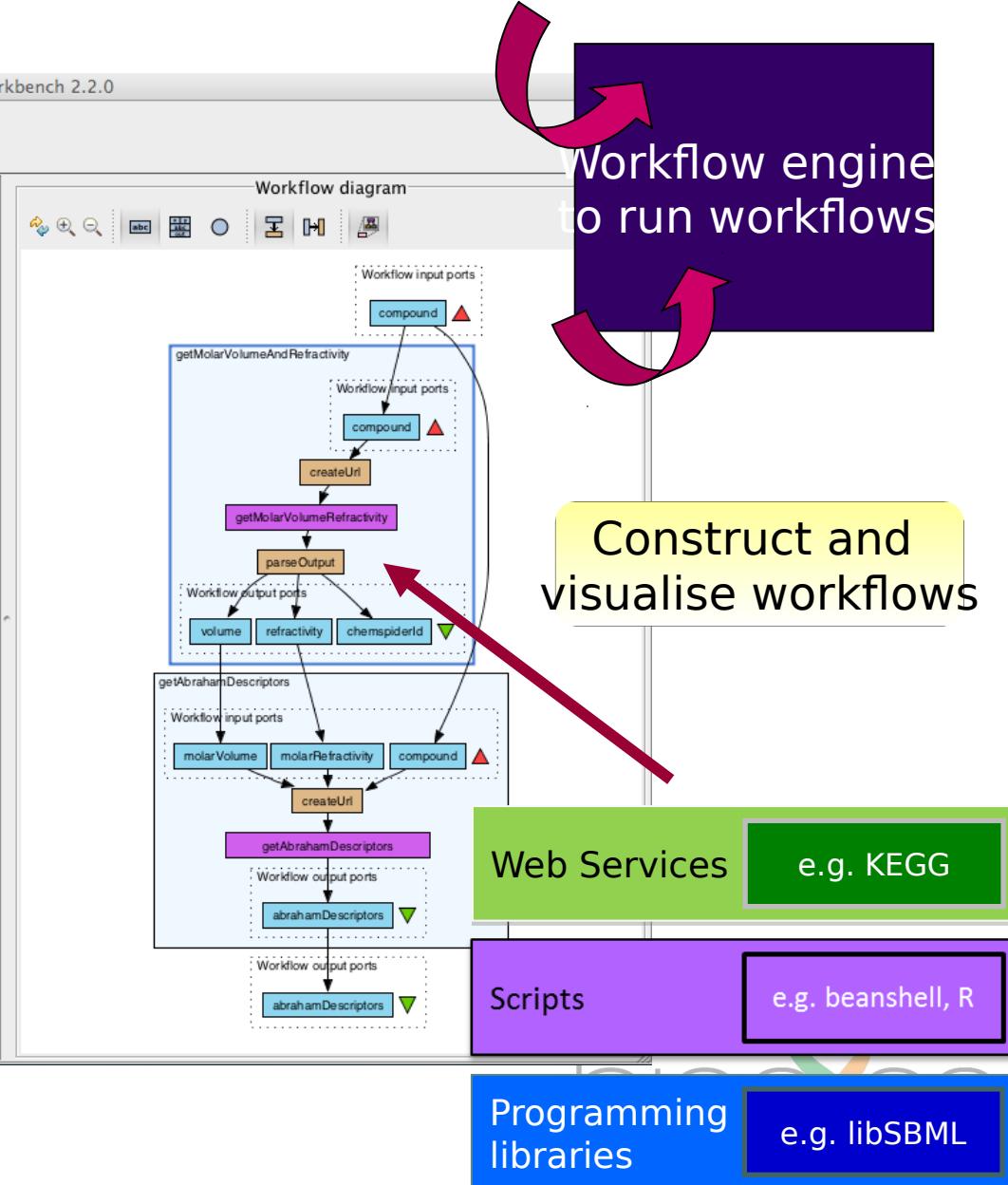
Apache Taverna



Taverna Workbench

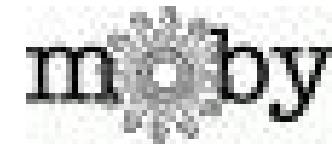
List of services

The screenshot shows the Taverna Workbench 2.2.0 interface. The top menu bar includes 'Design', 'Results', and 'myExperiment'. The main window has a 'Service panel' on the left containing a 'List of services' tree view. Under 'Available services', there are several items like Beanshell, Nested workflow, Rshell, SpreadsheetImport, and String constant. Below this is a 'Workflow explorer' tab showing a workflow named 'Workflow9' with nodes for 'Workflow input ports' (compound), 'Services' (getAbrahamDescriptors, getMolarVolumeAndRefractivity), and 'Workflow output ports' (abrahamDescriptors). A 'Validation report' tab is also visible.



Using Tools and Services from Taverna workflows

- Web Services
 - WSDL
 - REST
- Data services
 - BioMart
- Local scripts:
 - R
 - Beanshell
 - Command line (e.g. Python,
- Other workflows
- Appendix: And more... Add your own!



Web Services

Web Services: HTTP-based programmatic access (API).

Similar to “*GET me the web page <http://example.com/cats/pics>*”,

but Web Services allow “*GET me a genome sequence http://www.uniprot.org/uniprot/WAP_RAT.xml*”

Use **remote services** (typically free) from your computer in an automated way

Not the same as services on the web (i.e. **forms** that shows results as a web page)

Two flavours: **REST** (light-weight) and **SOAP** (“rich”)



Apache Taverna



Who Provides the Services?

Open domain services and resources

- Taverna accesses thousands of services
- Third party - we don't own them - we didn't build them
- All the major providers
 - NCBI, DDBJ, EBI ...
- ..but **no** common data model.



National Center for
Biotechnology Information (USA)



Tokyo, Japan



Cambridge, UK



Cambridge, UK



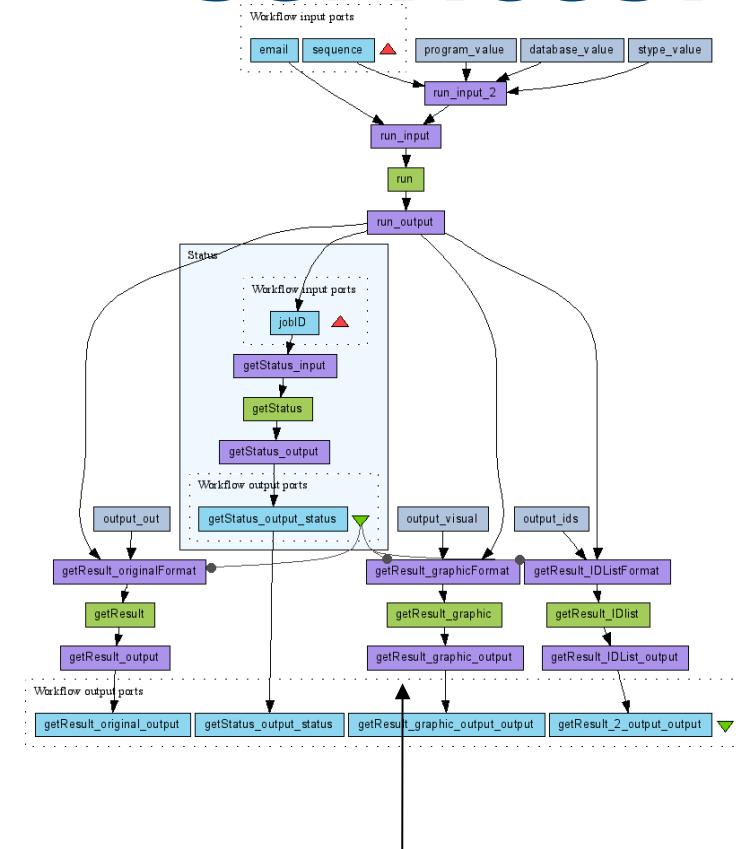
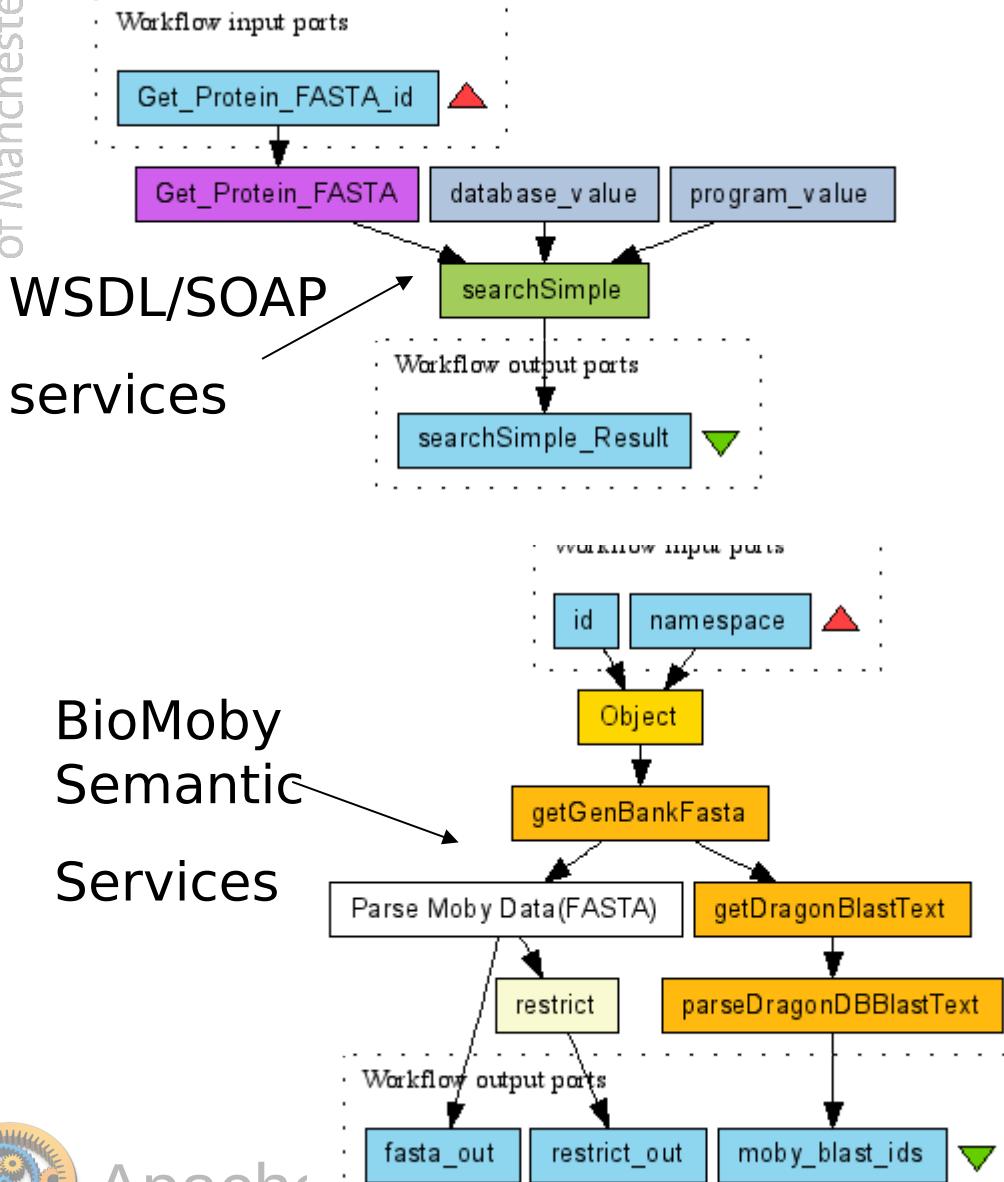
Apache Tavel



SeqHound



How do you use the services?



**Asynchronous services
(Submit, Wait, Fetch)**

Search: **Go!**[Home](#) [Services](#) [Register a Service](#) [Service Providers](#) [Search by Data](#) [Latest](#)[Home »](#)[!\[\]\(9fb35ce00785e0d1c8f42da5044e6593_img.jpg\) SHARE](#)     [EMAIL](#)

The BioCatalogue: providing a curated catalogue of life science Web services

Helpful Links

[Getting started with the Catalogue](#)[General info on Web services](#)[Best practice guide for developing Web services](#)[SSI's top tips on creating Web services](#)[Executing Web services](#)[Turn your command line application\(s\) into Web services](#)

Citing BioCatalogue

Bhagat, J., Tanoh, F., Nzuobontane, E., Laurent, T., Orlowski, J., Roos, M., Wolsencroft, K., Aleksejevs, S., Stevens, R., Pettifer, S., Lopez, R., Goble, C.A.

BioCatalogue currently has **1184 services**, **259 service providers** and **861 members** 

"Web Services are hard to find"

DISCOVER

-  Find the right Web Service
-  Powerful search and filtering
-  Information from providers and community

[More info](#)

"My Web Services are not visible"

REGISTER

-  Easily register Web Services
-  Instantly available to everyone
-  Providers can advertise, describe and monitor their Services

[More info](#)

"Web Services are poorly described"

ANNOTATE

-  Anyone can describe and annotate
-  Ongoing expert curation
-  Social curation by the community

[More info](#)

"Web Services are volatile"

MONITOR

-  Services change and get outdated
-  BioCatalogue monitors Services
-  Monitors availability and reliability

[More info](#)

Site Announcements

[ELIXIR Tools registry 'community paper' published in NAR](#)

By Niall Beard (11 months ago)

[BiodiversityCatalogue presents development of Web Map Service support at TDWG](#)

By Niall Beard (about 1 year ago)

[Scheduled Maintenance](#)

By Niall Beard (over 2 years ago)

[BioCatalogue Rails 3 Upgrade](#)

By Aleksandra Nenadic (almost 3 years ago)

[Soaplab EMBOSS services end of life: 31st January 2013](#)

By Robert Haines (almost 4 years ago)

[More](#)

 InterProScan 5

REST

No alternative names Categories: Protein Sequence Analysis Annotations: 5  0  5  0[Overview](#) [Rest Services \(3\)](#) [Examples](#) [Monitoring](#) [History](#)

What is an endpoint?

[Quick Browse](#) | [GET /parameterdetails/{parameter}](#) | [GET /parameters](#) | [GET /run](#) |<https://www.biocatalogue.org/services/3766/>[GET /parameterdetails/{parameter}](#)Part of Service: [InterProScan 5](#)Part of Endpoint Group: *none*Template:  <http://www.ebi.ac.uk/Tools/services/rest/iprscan5/parameterdetails/{parameter}>

No description(s) yet

Tags on this endpoint: *none*[GET /parameters](#)Part of Service: [InterProScan 5](#)Part of Endpoint Group: *none*Template:  <http://www.ebi.ac.uk/Tools/services/rest/iprscan5/parameters>

No description(s) yet



Apache Taverna



Search



Enter search query, e.g. "Proteomics", "Sequence alignment", "BAM".

Filter by attribute ▾



Filtering on Interface



- Name
- + Function
- Topic
- + Contact
- Resource type
- Interface
- Affiliation

SOAP WS

<https://bio.tools/>

170 entries found

Sort by

Latest



Display as

Cards



First

Previous

1

2

3

Next

Last

INB-BSC Genomics Cloud<http://transplant.bsc.es>

Cloud infrastructure deploying a series of genome oriented tools

[Genome annotation](#), [Sequence analysis](#)

Addition date 6 months ago
Affiliation ub.edu
Topic Sequence analysis
Resource Type Platform
Interface Web UI, SOAP WS
Maturity Stable
Cost Free with restrictions
Collection ELIXIR-ES, IRB-BSC

Documentation[Read the docs](#)**convert-matrix**<http://rsat.eu/>

Addition date 7 months ago
Affiliation RSAT
Topic Nucleic acid sites, features and motifs
Resource Type Tool, Service

**Apache Taverna**

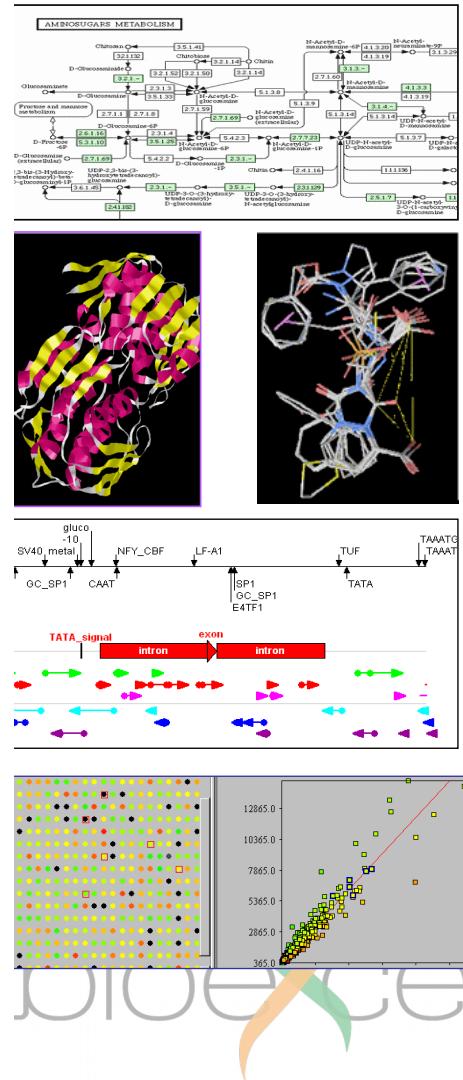
What do Scientists use Taverna for?

<http://taverna.apache.org/introduction/taverna-in-use/>

Taverna by domain

Various domains where scientists have used Taverna so far include:

- Annotation
- Arts
- Astronomy
- Biodiversity
- Bioinformatics
- Biology
- Chemistry
- Data and text mining
- Databases
- Document and image analysis (digital preservation)
- Education
- Engineering
- Geoinformatics
- Information quality
- Medicine
- Multimedia
- Natural language processing
- Protein and proteomics(/introduction/taverna-in-use/protein-and-proteomics)
- Service provision
- Service testing
- Social sciences



Workflows are ...

- ... records and protocols (i.e. your *in silico* experimental method)
- ... know-how and intellectual property
- ... hard work to develop and get right
-re-usable methods (i.e. you can build on the work of others)

So why not share and re-use them



Apache Taverna



Workflow Repository

[About](#) | [Mailing List](#) | [Publications](#) [Logout](#) | [Give us Feedback](#) | [Invite](#)[Home](#) [Users](#) [Groups](#) [Workflows](#) [Files](#) [Packs](#) [Services](#) [Topics](#) [Workflows](#) [Home](#) » [Workflows](#) [BOOKMARK](#) [f](#) [t](#) [e](#) ...[New/Upload](#)[Workflow](#)

Workflows

Search filter terms [« previous](#) [1](#) [2](#) [3](#) ... [191](#) [next »](#)

Sort by: Rank

Showing 1905 results. Use the filters on the left and the search box below to refine the results.

Filter by type

<input type="checkbox"/> Taverna 2	863
<input type="checkbox"/> Taverna 1	645
<input type="checkbox"/> RapidMiner	171
<input type="checkbox"/> Kepler	43
<input type="checkbox"/> Bioclipse Script...	34
<input type="checkbox"/> GWorkflowDL	24
<input type="checkbox"/> LONI Pipeline	22
<input type="checkbox"/> BioExtract Server	16
<input type="checkbox"/> Trident (Packa...	10
<input type="checkbox"/> LabTrove Tem...	9

[Taverna 2](#)[Pathways and Gene annotations for QTL region](#)

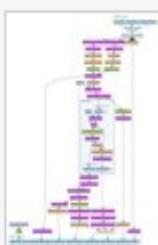
(v7)

Original Uploader[Paul Fisher](#)

Created: 19/11/09 @ 18:18:52 | Last updated: 02/09/11 @ 11:44:57

Credits: Paul Fisher

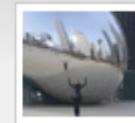
License: Creative Commons Attribution-Share Alike 3.0 Unported License



This workflow searches for genes which reside in a QTL (Quantitative Trait Loci) region in the mouse, *Mus musculus*. The workflow requires an input of: a chromosome name or number; a QTL start base pair position; QTL end base pair position. Data is then extracted from BioMart to annotate each of the genes found in this region. The Entrez and UniProt

View

Download (v7)

[Katy Wolstencroft](#)[My Profile](#) [\[edit\]](#)[My Messages](#)[My Memberships \(5\)](#)[My History](#)[My News](#)**4 new friendship requests**[Onlyhakanboz](#)[mihaiionita_me](#)[Pankaj chauhan](#)[Hanny](#)

Just Enough Sharing....

- myExperiment can provide a central location for workflows from one community/group
- myExperiment allows you to say
 - Who can look at your workflow
 - Who can download your workflow
 - Who can modify your workflow
 - Who can run your workflow
- Ownership and attribution



Spectrum of Users

Advanced users design and build workflows (informaticians)

myexperiment beta

Workflow Entry: Microarray CEL file

Version: 2 (latest)

Title: Microarray CEL file to candidate pathways

Version created on: Wednesday 03 October 2007 @ 18:35:55 (GMT)

Description:

This workflow takes in a http://www.bioinf.manchester.ac.uk/MADAT/ returned by this workflow are a list of the top differentially expressed genes on the number specified as input - geneNumber), which are then used to identify pathways which may be influencing the observed changes in the microarray. By identifying the candidate pathways, more detailed insights into the genes can be obtained.

CEL

Download

Download Scalable Diagram (SVG)

Taverna Workbench 2.2.0

Service panel

Available services

- Beanshell - A service that allows Beanshell scripts, with dependencies
- Nested workflow - A service that allows you to have one workflow inside another
- Rshell - A service that allows the calling of R scripts on an R server
- Spreadsheetimport - A service that imports data from spreadsheets
- String constant - A string value that you can set

Workflow diagram

```

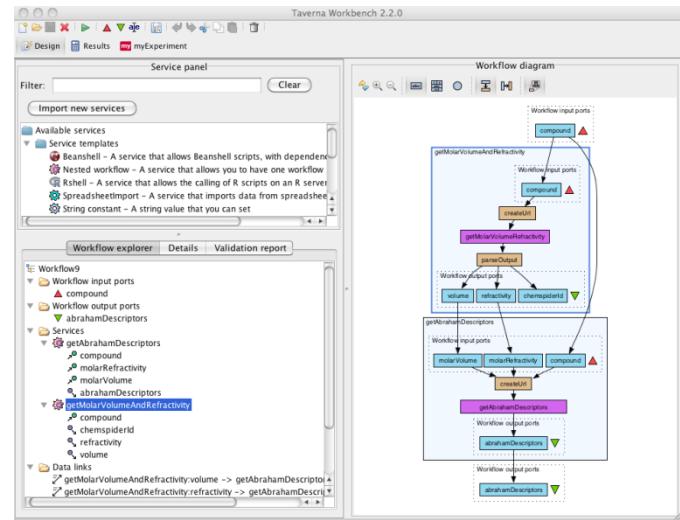
graph TD
    subgraph Workflow9 [Workflow9]
        direction TB
        A[getMolarVolumeAndRefractivity] --> B[parseOutput]
        B --> C[abrahamDescriptors]
        C --> D[getAbrahamDescriptors]
        D --> E[molarVolume]
        E --> F[molarRefractivity]
        F --> G[compound]
        G --> H[getAbrahamDescriptors]
        H --> I[abrahamDescriptors]
        I --> J[getMolarVolumeAndRefractivity]
        J --> K[parseOutput]
        K --> L[abrahamDescriptors]
        L --> M[getAbrahamDescriptors]
        M --> N[molarVolume]
        N --> O[molarRefractivity]
        O --> P[compound]
        P --> Q[getAbrahamDescriptors]
        Q --> R[abrahamDescriptors]
        R --> S[getMolarVolumeAndRefractivity]
        S --> T[parseOutput]
        T --> U[abrahamDescriptors]
        U --> V[getAbrahamDescriptors]
        V --> W[molarVolume]
        W --> X[molarRefractivity]
        X --> Y[compound]
        Y --> Z[getAbrahamDescriptors]
        Z --> AA[abrahamDescriptors]
    end

```

Others “replay” workflows through web page



Apache Taverna



Intermediate users reuse and modify existing workflows or components



Workflow
Repository**my experiment**

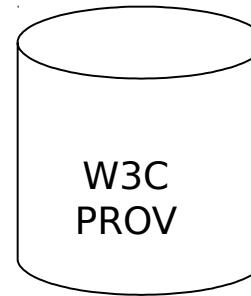
Service Catalogue

BioCatalogue 

"The Life Science Web Service Registry"

Activity and Service
Plug-in ManagerCOMMON
WORKFLOW
LANGUAGEmobythe Taverna

A Collection of Tools

Workflow GUI
Workbench
and 3rd party plug-ins**Taverna**Provenance
StoreWorkflow
ServerSecure Service Access, and
Programming APIs

Client User Interfaces

Taverna Player

Custom Web Portals

Mobile App

Programming and
APIsbioexcel

What's next?



COMMON
WORKFLOW
LANGUAGE

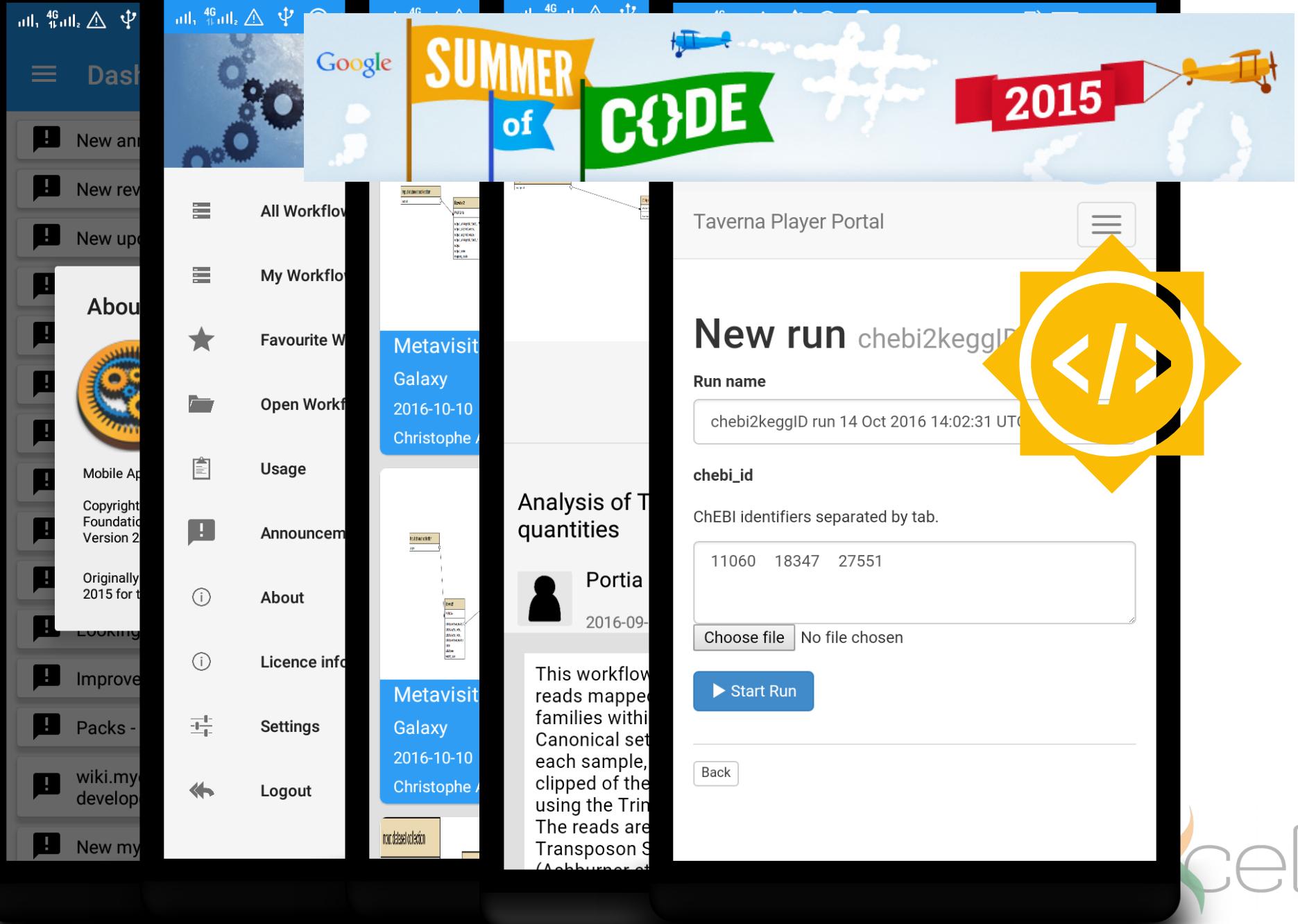


Apache Taverna



Apache Taverna





Common Workflow Language Viewer

Workflow: lobSTR-workflow.cwl

Missing workflow description

Inputs

Name	Description	Format Depth
	lobSTR's bwa reference files	0
	Use this in the read group SM tag	0
	list of files containing the first end of paired end reads in fasta or fastq format	0
	list of files containing the second end of paired end reads in fasta or fastq format	0
	prefix for output files. will output prefix.aligned.bam and prefix.aligned.stats	0
	Use this in the read group LB tag	0
	File containing statistics for each STR.	0
	File to read noise model parameters from (.stepmodel)	0

Outputs

Name	Description	Format	Depth
		0	
		0	
		0	
		0	



COMMON
WORKFLOW
LANGUAGE



Apache Taverna





About the Community

Apache Taverna is a [incubator project](#) within the [The Apache Software Foundation \(ASF\)](#).

Taverna is an open source community that's organized by its [members](#), including you! Whether you're running or [writing workflows](#), developing plugins, hosting a service, hacking the core, [citing us](#), or just want to [edit](#) these pages there are opportunities for you to get in touch and ask questions, contribute,

How to Communicate

Mailing Lists

You are highly encouraged to participate on the Apache Taverna [mailing lists](#):

Users

Build/run workflows; analytical science; services

[Subscribe](#)[Unsubscribe](#)[Archives](#)**Developers**

Coding on/with Taverna; workflow/provenance science

[Subscribe](#)[Unsubscribe](#)[Archives](#)**Commits**

Code changes; automated build reports

[Subscribe](#)[Unsubscribe](#)[Archives](#)

More Information

- Apache Taverna (incubating)
• <http://taverna.apache.org/>
- myExperiment
• <http://www.myexperiment.org>
- BioCatalogue
• <http://www.biocatalogue.org>



Apache Taverna



Download Taverna Workbench 2.5

for *Bioinformatics* (~230 MiB):

<https://s.apache.org/taverna-bio>

or: *Core edition* (~190 MB):

<https://s.apache.org/download-taverna>

then skip “Service Catalogue” in tutorial

Then follow “*day1*” in **tutorial**:

<https://s.apache.org/taverna-tutorial>



Apache Taverna

