

# An Introduction to Designing and Executing Workflows with Taverna

Stian Soiland-Reyes and Christian Brenninkmeijer  
University of Manchester  
materials by Dr Katy Wolstencroft and Dr Aleksandra Pawlik

<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 3.0 Unported License](http://creativecommons.org/licenses/by/3.0/)*

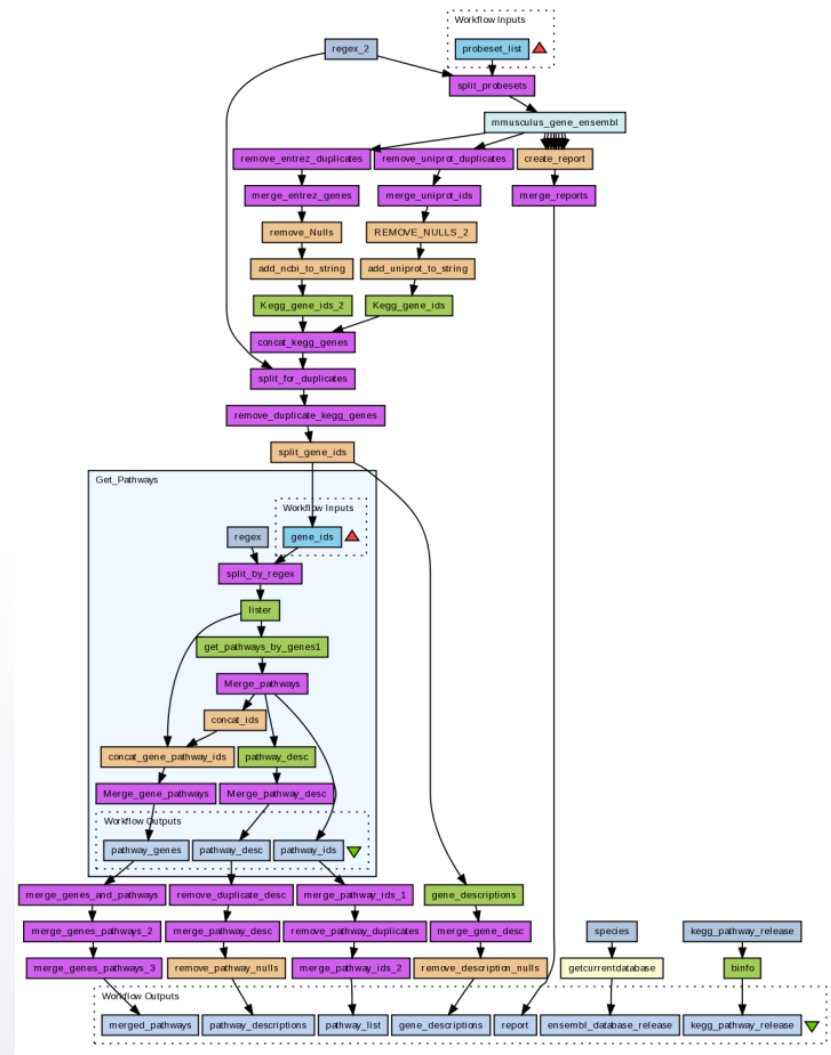
Bonn University, 2014-09-01

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



# Workflows

- Sophisticated analysis pipelines
- A set of services to analyse or manage data (either local or remote)
- Data flow through services
- Control of service invocation
- Iteration
- Automation
- Access to intermediate results





# Workflows in Taverna

- This tutorial will give you a basic introduction to designing, and reusing workflows in Taverna and some of its main features.
- Workflows in this practical use small data-sets and are designed to run in a few minutes. In the real world, you would be using larger data sets and workflows would typically run for longer



# Taverna: Download and installation

- Taverna can be downloaded from <http://www.taverna.org.uk/>
- Go to the page and find the latest (2.5)
- Follow the instructions on the website to install Taverna for your operating system (this is a simple one-click install for windows and Mac. For Linux, you may also need the GraphViz program. Follow the link on the Taverna download page if so)

The following page shows a screenshot of Taverna and the different panels that make up the workbench



# Taverna Workbench

The screenshot displays the Taverna Workbench 2.4.0 interface. The top menu bar includes File, Edit, Insert, View, Workflows, Components, and Advanced Help. Below the menu is a toolbar with icons for Design, Results, myExperiment, and Service Catalogue. The interface is divided into several panels:

- Services Panel:** Located on the left, it contains a search filter, an 'Import new services' button, and a list of available services categorized into Service templates, Local services, Biomart, file-based services, Interaction, OAuth services, and Soaplab. A box labeled 'Services Panel' is overlaid on this area.
- Workflow Explorer:** Located below the Services Panel, it shows a tree view of the current workflow, 'EBI\_InterproScan\_NewServices'. It lists workflow input ports (email, sequence), workflow output ports (getResult\_3\_output\_output, getResult\_output\_output, Graphical\_output, Workflow16\_getStatus\_output\_status), and services (getResult, getResult\_graphic, getResult\_graphic\_input). A box labeled 'Workflow Explorer' is overlaid on this area.
- Workflow Diagram:** The main workspace on the right shows a detailed workflow diagram for 'EBI\_InterproScan\_NewServices'. It features a central 'Status' component with its own input and output ports. The workflow starts with 'Workflow input ports' (sequence, email) leading to 'run\_input\_2', 'run\_input', and 'run'. The 'run' component produces a 'run\_output' which feeds into 'getResult\_input', 'getResult\_graphic\_input', and 'getResult\_xml\_input'. These lead to 'getResult', 'getResult\_graphic', and 'getResult\_xml' components, which then produce 'getResult\_output', 'getResult\_graphic\_output', and 'getResult\_xml\_output'. Finally, these outputs are mapped to the workflow's output ports. A box labeled 'Workflow Diagram' is overlaid on the top part of this area.



# Workflow Diagram

The workflow diagram is the visual representation of the workflow, it:




- Shows inputs, outputs, services and data flows
- Allows editing of the workflow by dragging and dropping and connecting services together
- Enables saving of workflow diagrams for publishing and sharing



# Workflow Diagram

Type	Description
Beanshell	A user editable scripting operation using interpreted Java
Nested_workflow	A sub-workflow exposed as a single operation
Rshell	A user editable scripting operation using the R programming language
SpreadsheetImport	An operation for loading spreadsheet data in CSV and Excel format
String_constant	A local operation to emit a single constant string
XPath_Service	A local operation to extract data from XML data
local_java	A local operation coded as a Java class
biomart	A configurable parameterised query over a Biomart data warehouse dataset. See link
soaplab	A service wrapper around a legacy commandline tool such as the EMBOSS tool set
wsdl_web_service	A processor accessing a standard SOAP service
REST_Service	A processor accessing a RESTful web service
Tool_service	Tool service

Other elements of the workflow are coloured as follows:

Element	Description
input_port 	Input port
output_port 	Output port
	Merge
XML_splitter	XML Splitter



# Workflow Explorer

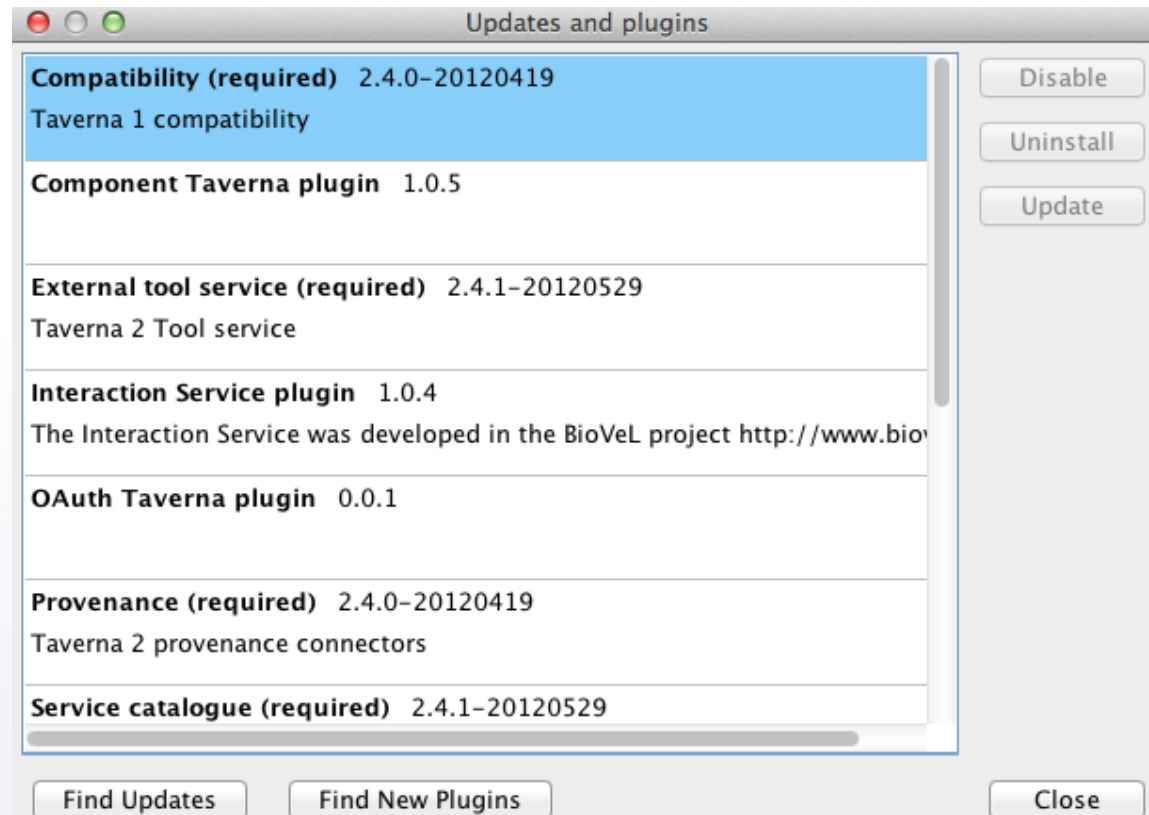
- The Workflow Explorer shows the detailed view of your workflow. It shows default values and descriptions for service inputs and outputs and it shows where remote services are located. It also shows configuration details, such as iteration and looping
- Workflow validation details can also be found here. Before a workflow is run, Taverna checks to see if it is connected correctly and if its services are available.





# Updates and Plugin Installation

- Taverna updates are issued on the regular basis
- There is also a number of plugins which are developed for Taverna
- To get the updates and plugins select “Advanced -> Updates and plugins”





# Available Services Panel

## Lists services available by default in Taverna

- Local java services
- WSDL Web Service – secure and public
- RESTful Services
- R Processor services (for statistical analyses)
- Beanshell scripts
- Xpath scripts
- Spreadsheet import service

The services panel also allows you to add new services or workflows from the web or from file systems – there are loads more available!



# Taverna Workflow Management

## Open extensibility

- Plug-in framework
  - Command line tool
  - Data Services: VOTables for AstroTaverna
  - Optimisations: E.g. Holl. model parameter sweeps
  - Infrastructures: Grid, HPC, Web Services
  - Domains: CDK, BioMart, VOTable
  - Commodities: Excel Spreadsheets, Open Refine, R
- Plug into other frameworks & platforms
  - Portals: Scratchpads
  - Interactive platforms: iPython Notebook
  - Wfms: KNIME Node, Galaxy tool, Kepler Actor
- Third party applications
  - Taverna Online
  - XworX
  - OGC chainer

Registry: <http://registry.euro-vo.org/services/RegistrySearch>

Keywords: table

Cone Search   SIA Search   SSA Search

100 results for ConeSearch: table

Short name	Title	Subjects	Identifier	Publisher
A2PIC	HEAO 1 A2...	[Extragalacti...	ivo://nasa.h...	NASA/GSFC...
A3	HEAO 1 A3...	[Observation]	ivo://nasa.h...	NASA/GSFC...
Abell	Abell Clusters	[Cluster of G...	ivo://nasa.h...	NASA/GSFC...
AGNSDSSXMM	Sloan Digital...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
ASCALSS	ASCA Large...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
ASCA	ASCA Maste...	[Observation]	ivo://nasa.h...	NASA/GSFC...
ASIAGOSN	Asiago Supe...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
ATNF	ATNF Pulsar...	[Pulsar]	ivo://nasa.h...	NASA/GSFC...
BATSE4B	CGRO/BATS...	[GRB]	ivo://nasa.h...	NASA/GSFC...
BATSEOCAT	CGRO/BATS...	[Observation]	ivo://nasa.h...	NASA/GSFC...
BATSEGRB	CGRO/BATS...	[GRB]	ivo://nasa.h...	NASA/GSFC...
BATSE/Spec	CGRO/BATS...	[GRB]	ivo://nasa.h...	NASA/GSFC...
BATSE/PLSR	CGRO/BATS...	[Pulsar]	ivo://nasa.h...	NASA/GSFC...
BAXGalClus	BAX X-Ray...	[Cluster of G...	ivo://nasa.h...	NASA/GSFC...
BBXRT	BBXRT Archive	[Observation]	ivo://nasa.h...	NASA/GSFC...
Be	Be Stars Cat...	[Star]	ivo://nasa.h...	NASA/GSFC...
BHROSATOpt	Byurakan/H...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
BMWCHANC...	Brera Multi...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
CARINAXMM	Carina OB1...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
Chan/DF25	ChandraDee...	[Optical Cou...	ivo://nasa.h...	NASA/GSFC...
CDFSAGNCXO	Chandra De...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
CENACXO	Centaurus A...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
CEPAXMM	Cepheus A S...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
CEPBOB3CXO	Cep B/OB3...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
CHAINHCXO	Chamaeleon...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
CHAMPHXA...	CHAMP (Cha...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
ChamPPS	Chandra Mu...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
ChanDF	Chandra De...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
ChanDFSS	Chandra Ext...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
Chandra	Chandra Ob...	[Observation]	ivo://nasa.h...	NASA/GSFC...
ChanNSpiral	Chandra Ne...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
Chan/Seren...	Chandra Ser...	[Optical Cou...	ivo://nasa.h...	NASA/GSFC...
CHANULXCAT	Chandra Arc...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
ChanNGalPSC	Chandra Ne...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...
CLANS	Chandra Loc...	[Survey Sour...	ivo://nasa.h...	NASA/GSFC...

**AGNSDSSXMM: Sloan Digital...**

*CATALOG*  
Bright XMM-Newton data an sample consists of 359 source completeness (100 and 85 per obscured AGN falls with incr effect can most probably be a XMM-Newton/Sloan Digital well as 4 sources with X-ray <http://www.sdss.org/>.

Source  
2006A&A...459..693/

Subjects  
Survey Source

**Service**  
ivo://ivoa.net/std/cones

Verbose  
Maximum records  
99999  
Maximum search radius  
180.0  
Test query  
SR  
0.016666666666666666  
DEC  
-10.523  
RA  
358.456

GET <http://heasarc.gsfc.nasa.gov/>  
Version: 1.0



## First tutorials:

- Downloading/installing Taverna
- Building a simple workflow
- Finding services in BioCatalogue