

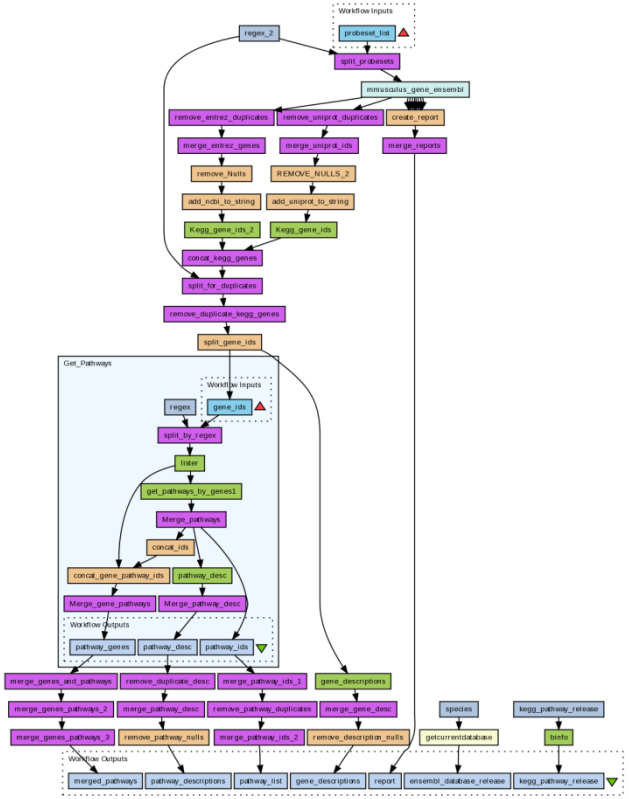

An Introduction to Designing and Executing Workflows with Taverna

Alan R Williams

University of Manchester

Workflows

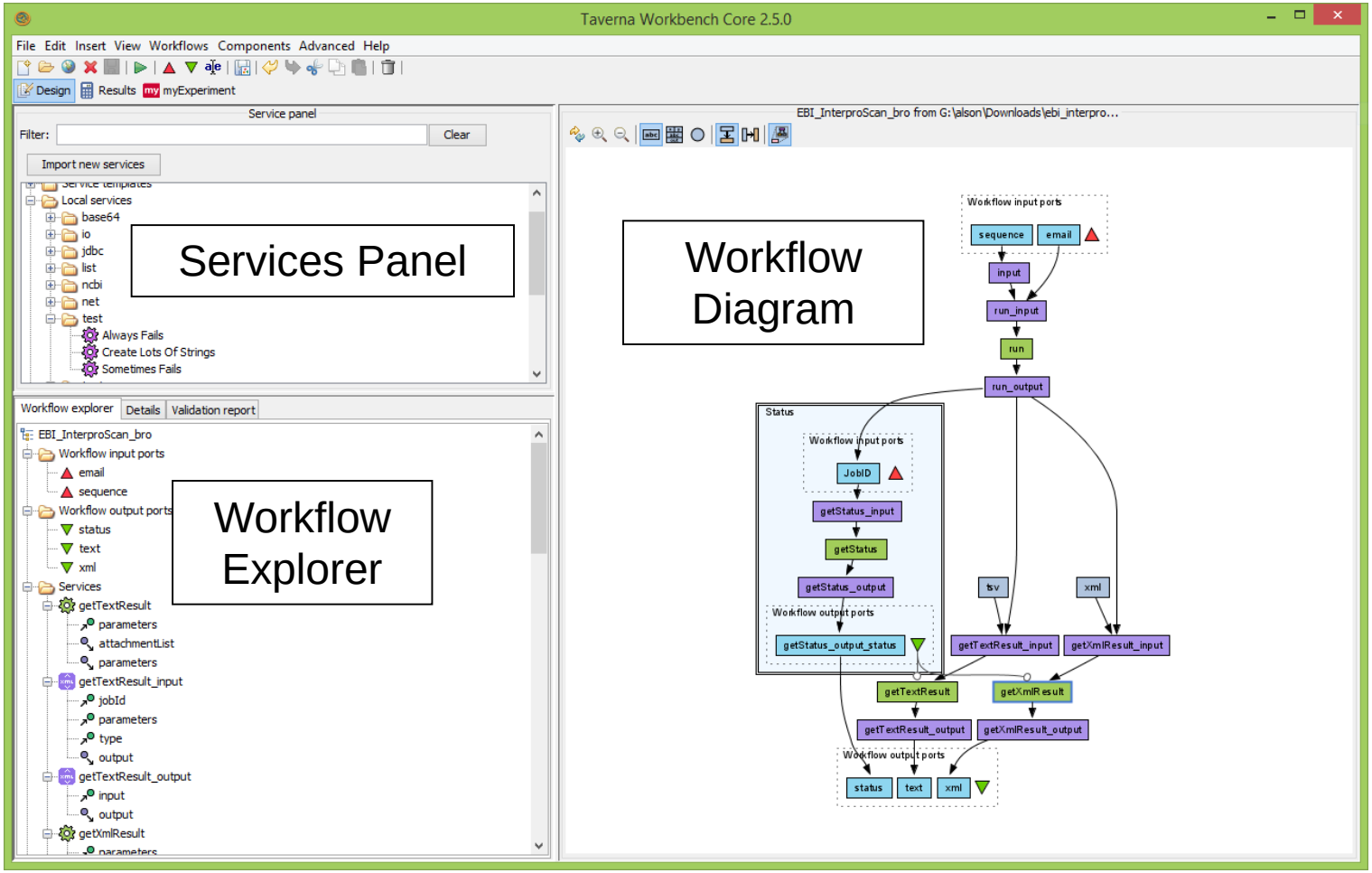
- Sophisticated analysis pipelines
- A set of services to analyse or manage data (either local or remote)
- Data flow through services
- Control of service invocation
- Iteration
- Automation
- Access to intermediate results



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 allows you to use different forms of data input and save output data in different formats too – we will look at that in this tutorial as well.

Taverna Workbench



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

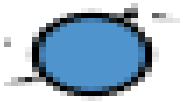
Taverna understands difference types of services

Type	Description
Beanshell	A user editable scripting operation using Beanshell script
Nested_workflow	A sub-workflow exposed as a single operation
Rshell	A user editable script in R
SpreadsheetImport	An operation for loading spreadsheet data in CSV and Excel format
String_constant	A single constant string
XPath_Service	Extraction of data from XML data

Type	Description
Local_service	A standard Beanshell
wsdl_web_service	A processor accessing a standard SOAP service
REST_Service	A processor accessing a RESTful web service
Tool_service	Calling a local or remote tool/script
Interaction	Interaction in a web browser
Component	Re-usable component workflow

Workflow Diagram

Taverna workflows also have

Element	Description
	Input port
	Output port
	“Merge” - make a list out of elements (often sub-lists)
	XML Splitter
	Data flow
	Control flow (run after)

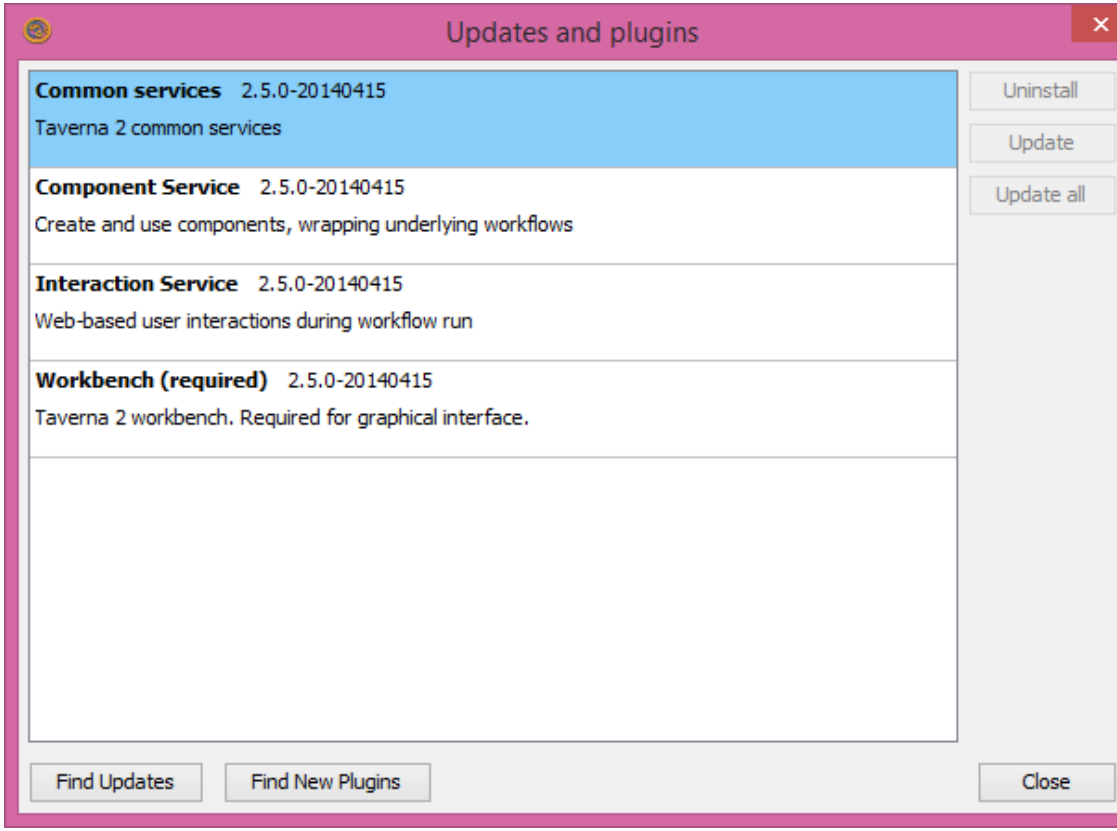
Workflow Explorer

- 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

- It's a good practice to update Taverna regularly
- Taverna updates are issued on a 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

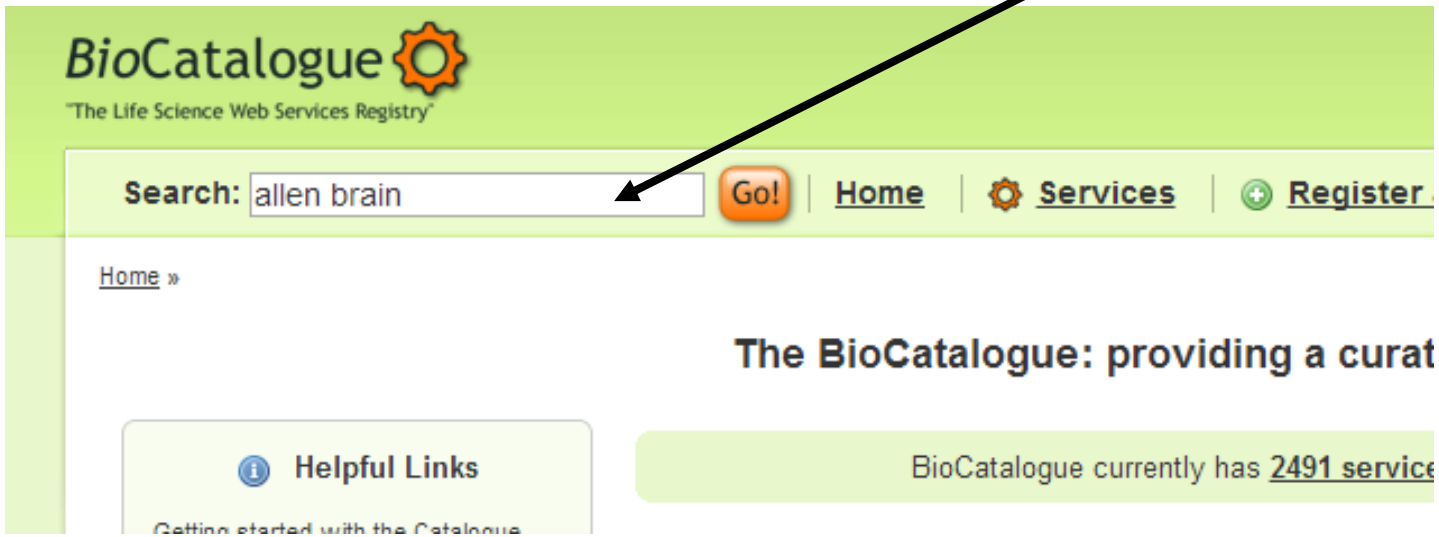
Lists services available by default in Taverna

- Local services – a standard set of utilities
- WSDL Web Service – secure and public
- RESTful Services
- R Processor services (for statistical analyses)
- Beanshell scripts
- XPath scripts
- Spreadsheet import service
- Interaction service – allows the workflow to “ask” for information in a web browser
- Component – re-usable workflows

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

Exercise 1: Building a Simple Workflow

- We will start with something easy – we will use an Allen Brain service to retrieve information about the experiments for a gene whose name we will provide
- Go to the www.biocatalogue.org and search for "Allen brain"

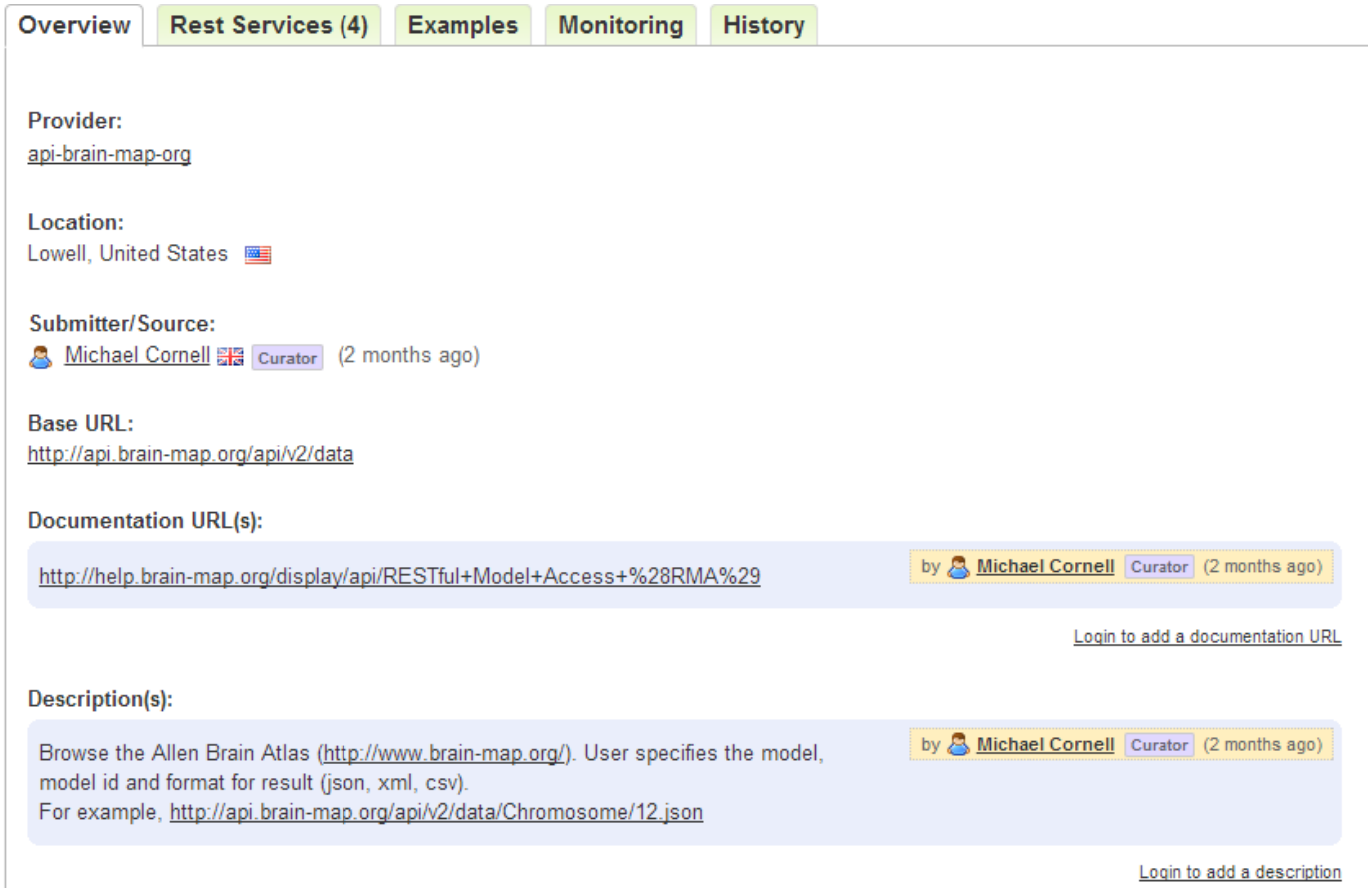


Exercise 1: Building a Simple Workflow

- From the results select Allen Brain Atlas Browse

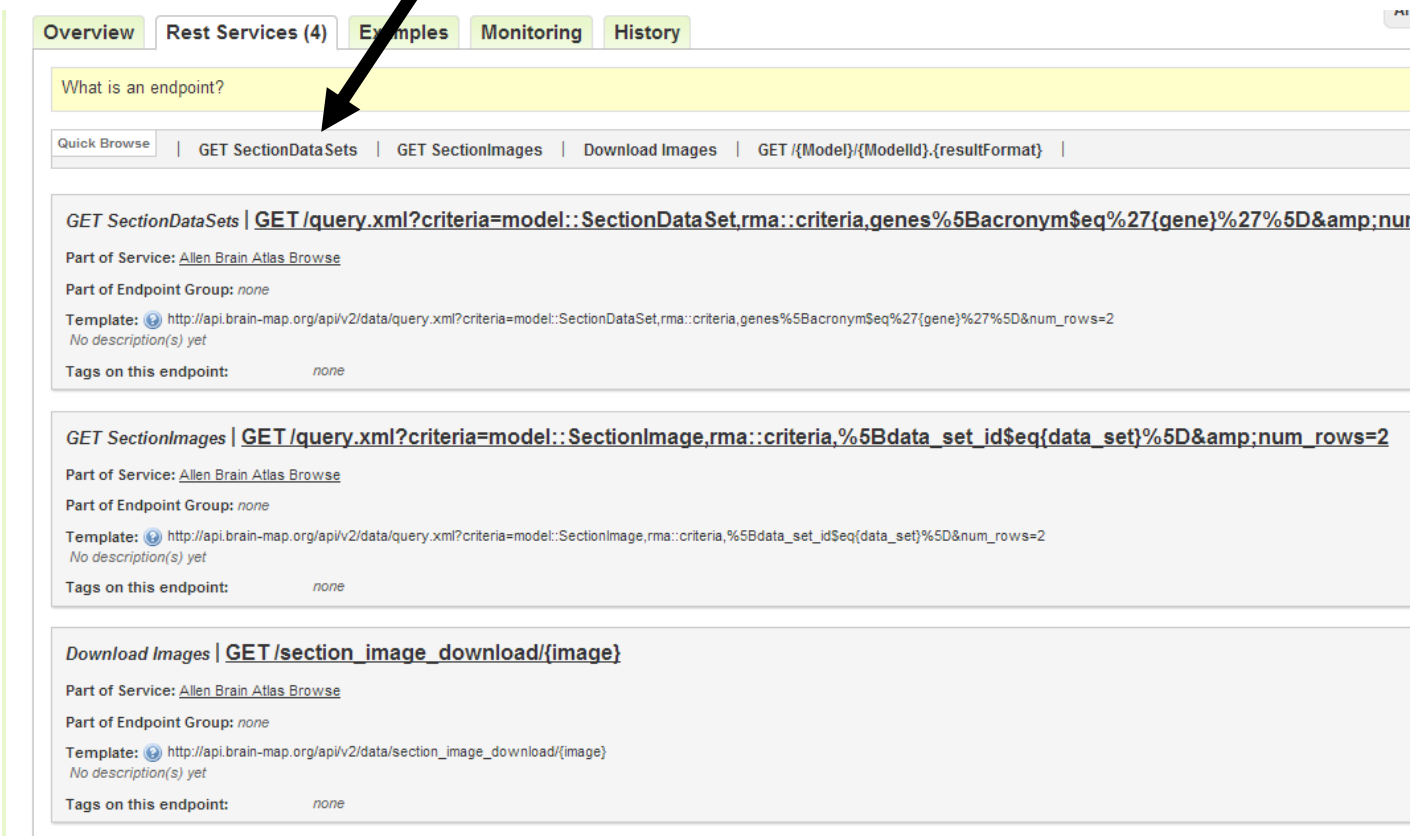
Exercise 1: Building a Simple Workflow

- Have a look at the service description



Exercise 1: Building a Simple Workflow

- Select the Rest Services tab and see how the service can be used
- Click on GET SectionDataSets

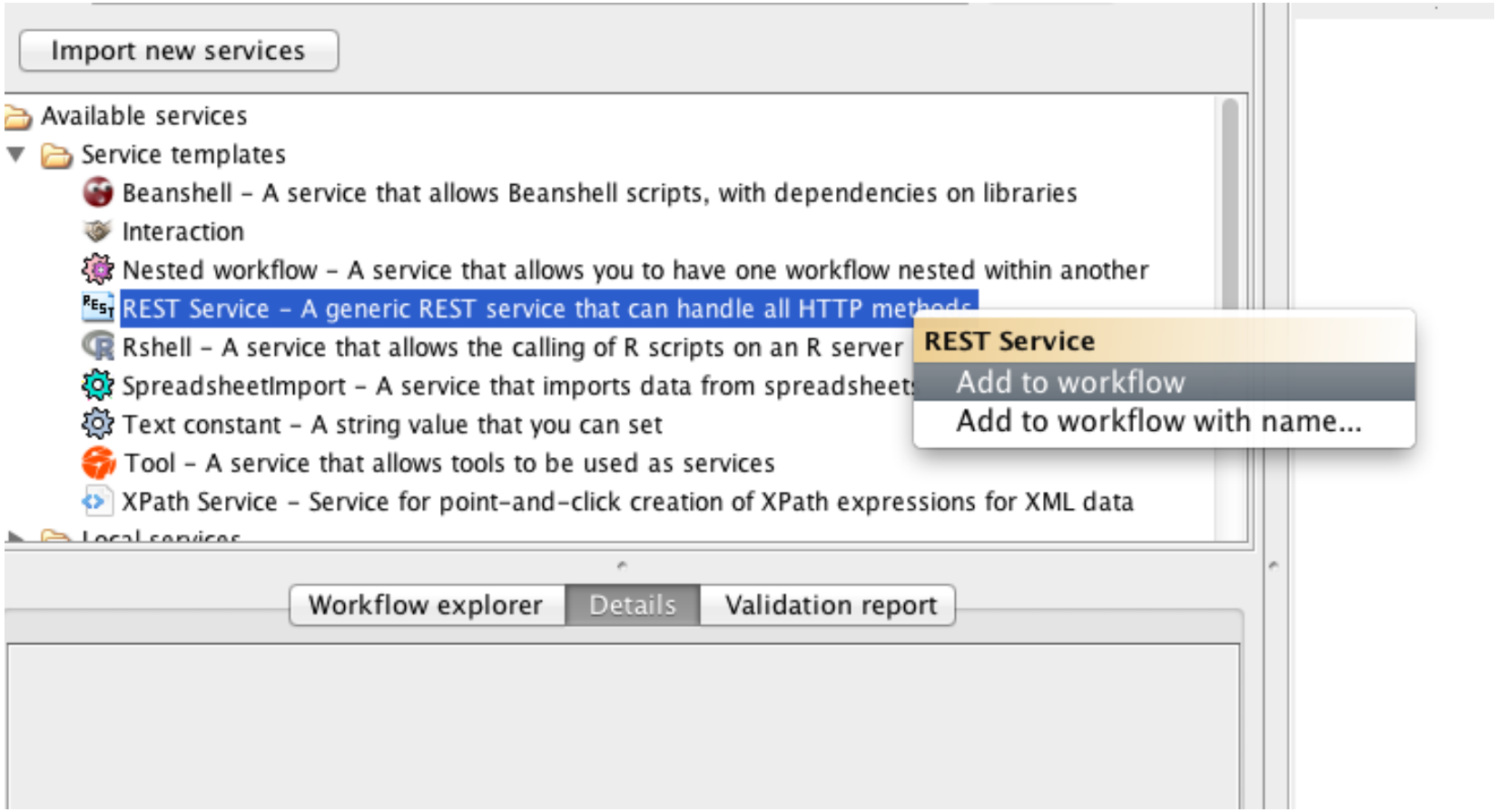


Exercise 1: Building a Simple Workflow

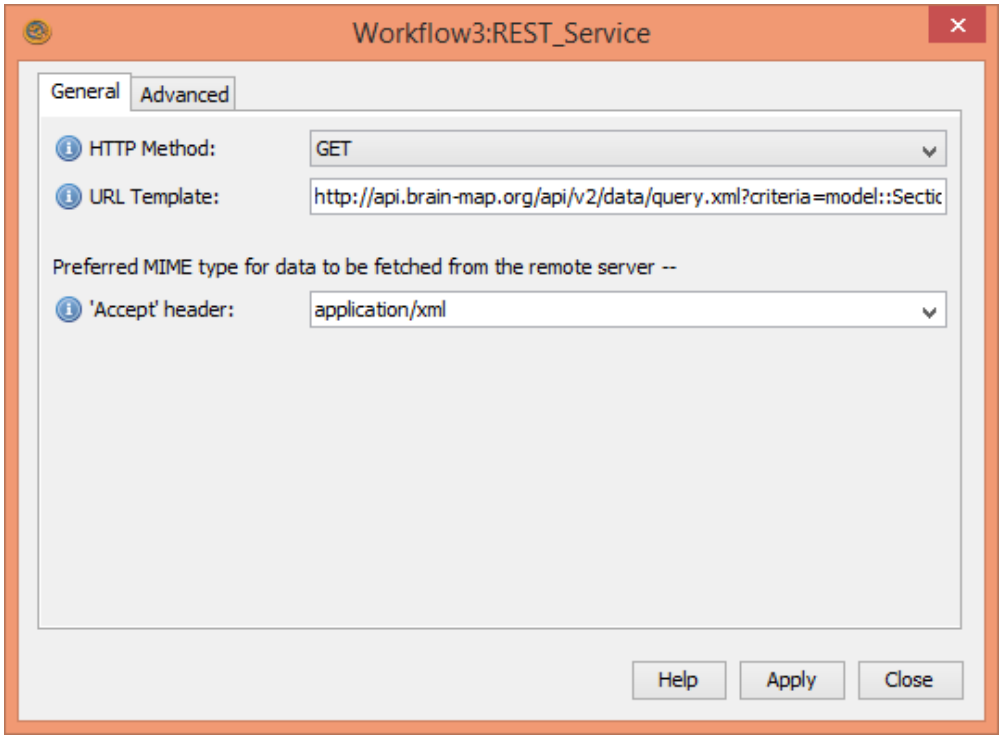
- This service returns the results for a gene name that we will provide, and limits the number of results to 2
- We can copy the Template

  http://api.brain-map.org/api/v2/data/query.xml?criteria=model::SectionDataSet,rma::criteria,genes%5Bacronym$eq%27{gene}%27%5D&num_rows=2
- In Taverna Workbench go to the **Services Panel** in the **Design** view
- From the **Available Services** select **Service Templates** and **REST**
- Right-click on it and select **Add to workflow** (see the next slide)

Exercise 1: Building a Simple Workflow



Exercise 1: Building a Simple Workflow

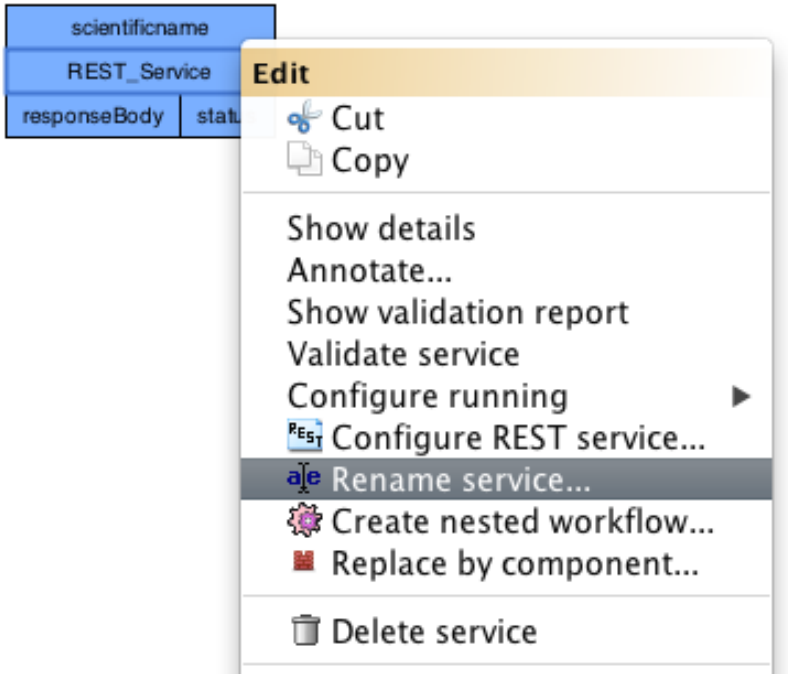


- Enter the template you copied into the URL template field:

http://api.brain-map.org/api/v2/data/query.xml?criteria=model::SectionDataSet,rma::criteria,genes%5Bacronym$eq%27{gene}%27%5D&num_rows=2

Exercise 1: Building a Simple Workflow

- Let's change the name of the service to: *AllenExperimentsForGene*



Exercise 1: Building a Simple Workflow

- At the top of the workflow diagram panel, change the view to show all ports by clicking on the icon shown below

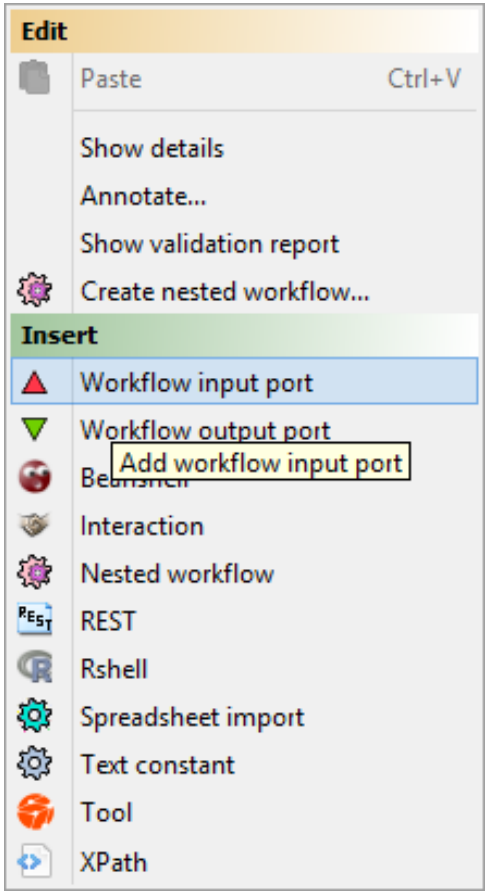
Show all ports icon

- This view allows you to see any data input/output or parameter value options for your chosen service

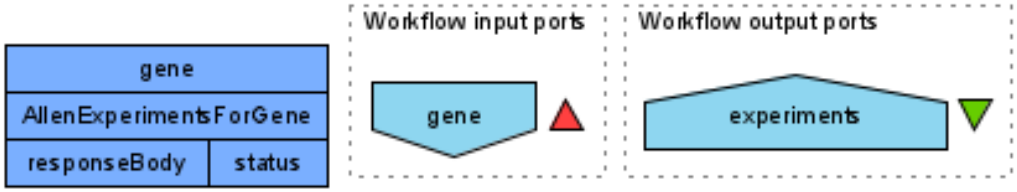
Exercise 1: Building a Simple Workflow

- In a blank space in the **workflow diagram**,right-click and select **Workflow input port** from the **Insert** section
- Type in a name for this input (e.g. *gene*) and click **OK**

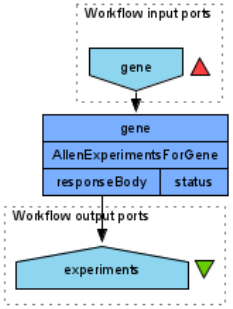
- Do the same to create a new workflow output. Call this output *experiments*



Exercise 1: Building a Simple Workflow

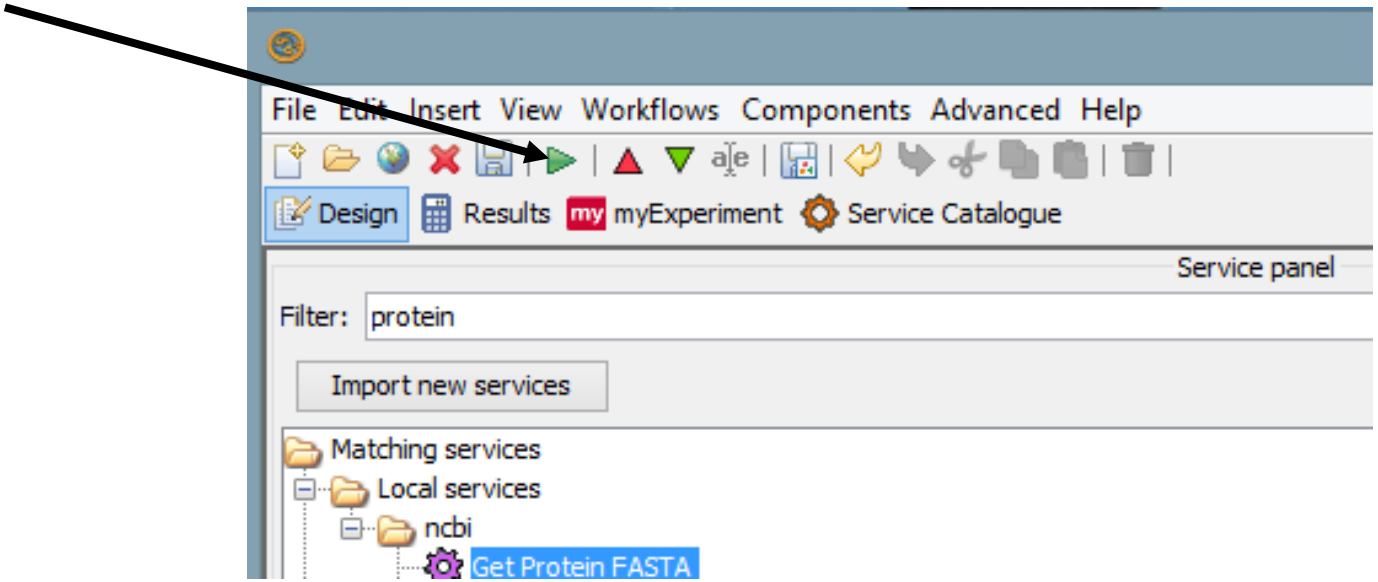


- Connect the input and output ports
- Your workflow should look like this



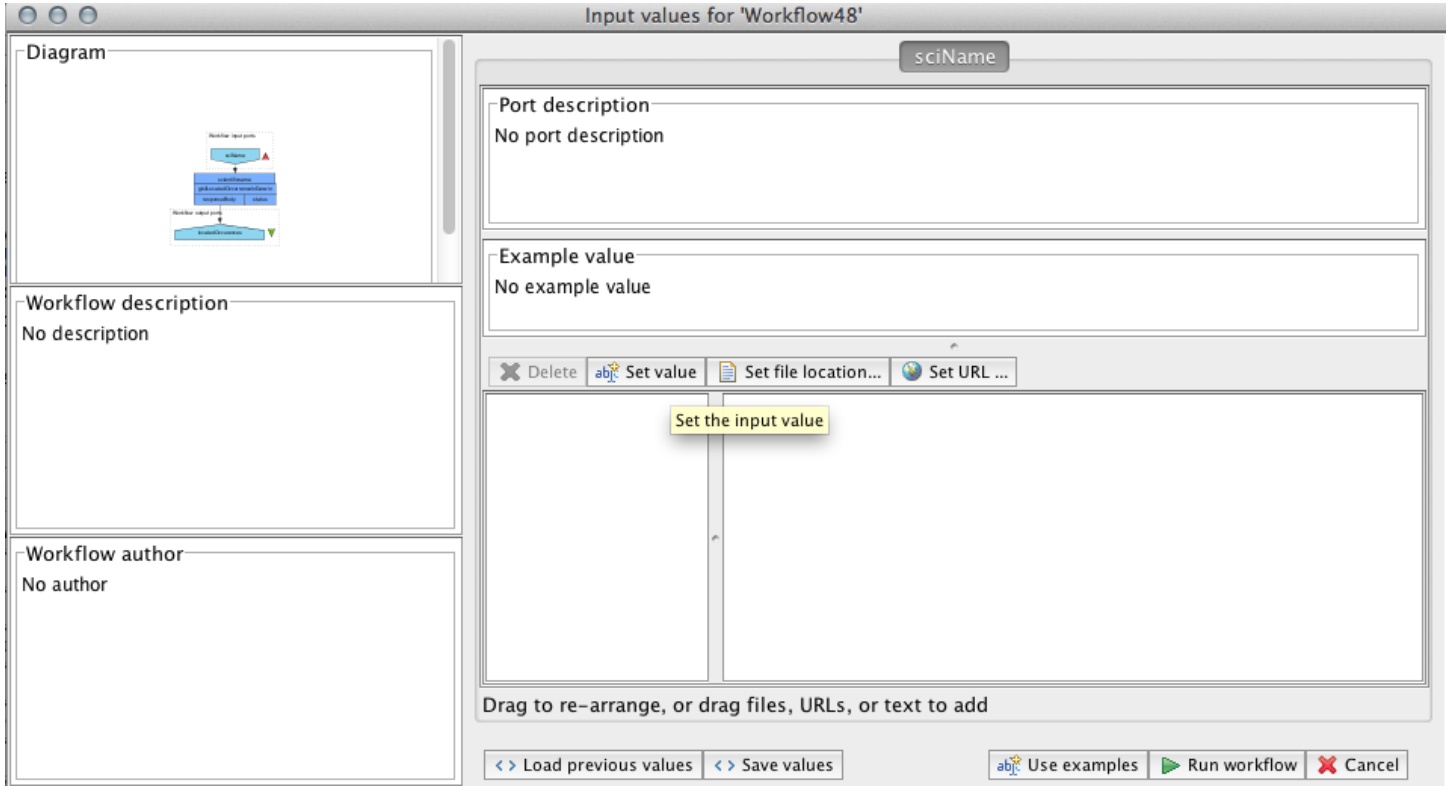
Exercise 1: Building a Simple Workflow

- Run the workflow by selecting **File** -> **Run workflow**, or by clicking on the play button at the top of the workbench



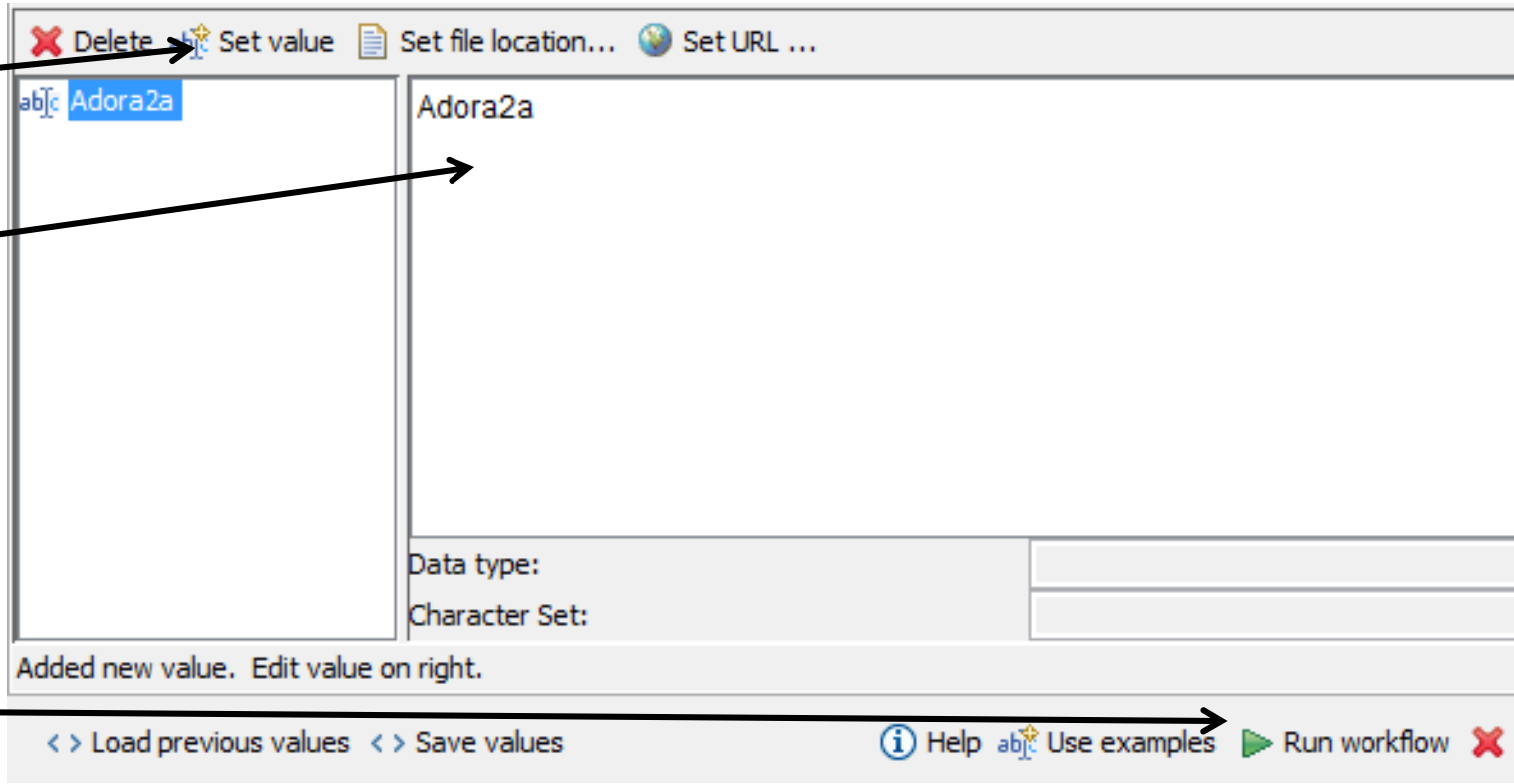
Exercise 1: Building a Simple Workflow

- You'll get a pop up window where you can enter the data for the workflow.

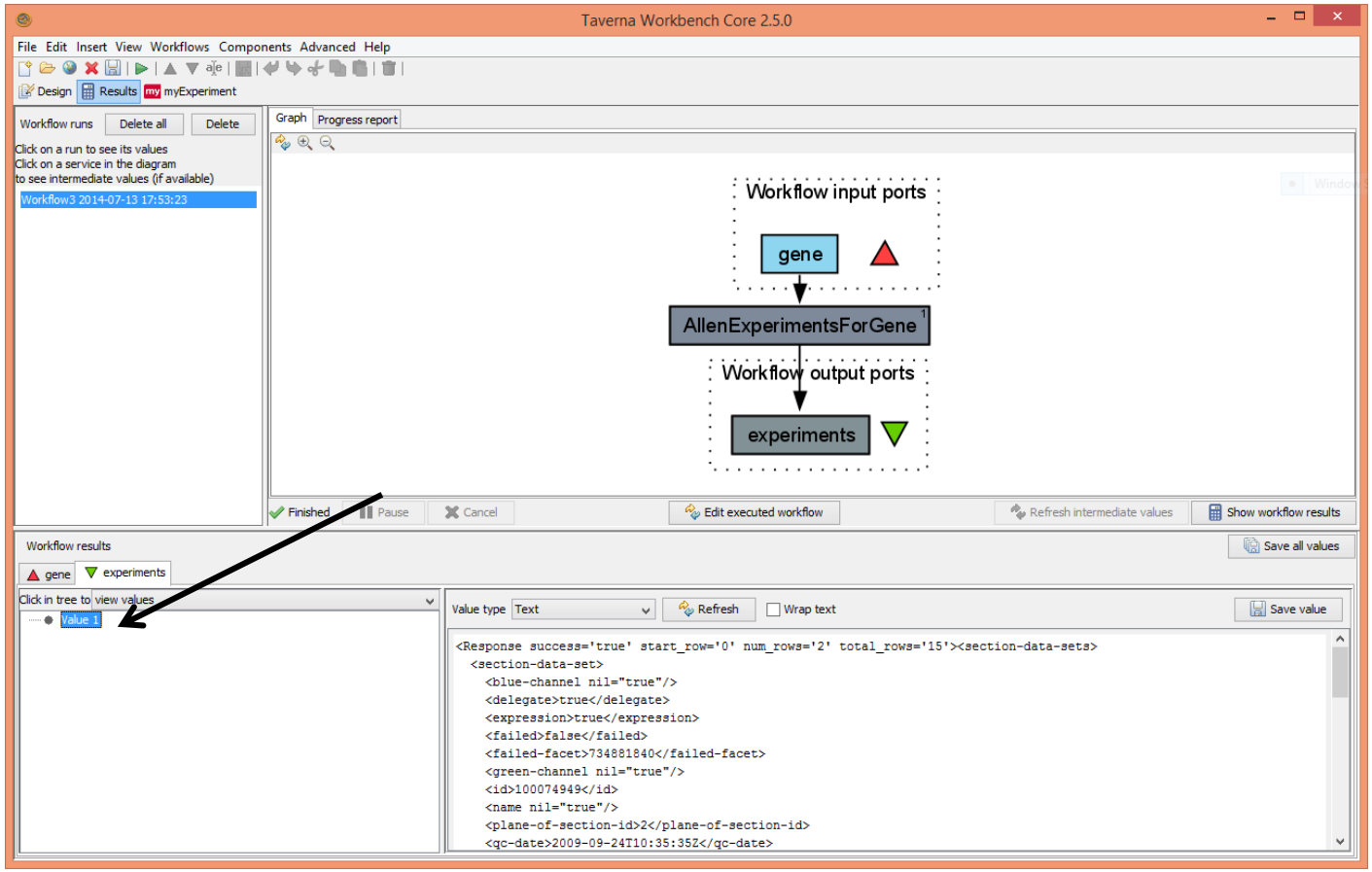


Exercise 1: Building a Simple Workflow

- Click **Set value**
- Enter *Adora2a* (note no newline)
- At the bottom of the window **Run workflow**



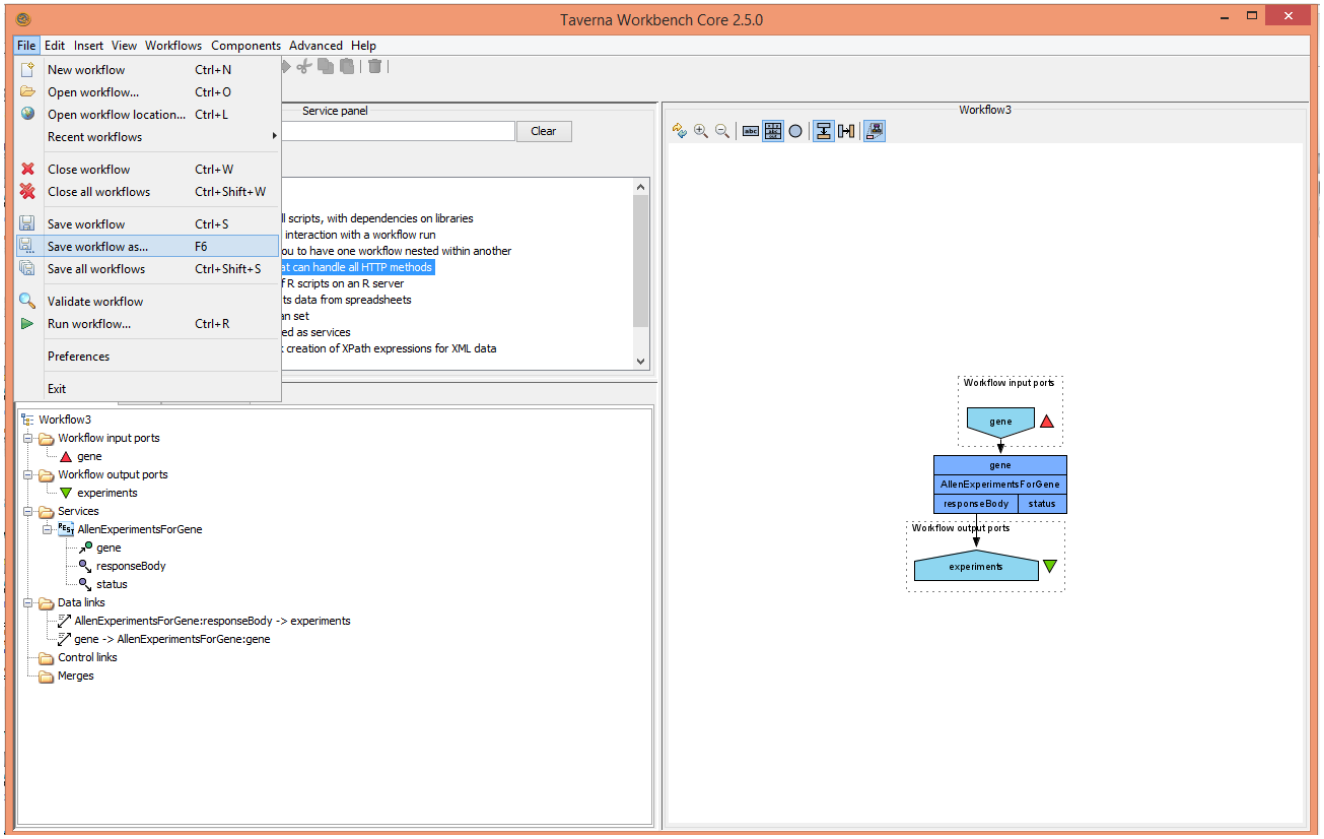
Exercise 1: Building a Simple Workflow



- You should see the workflow running
- Once the workflow finished running
- Click on *experiments* and **Value** to see the results

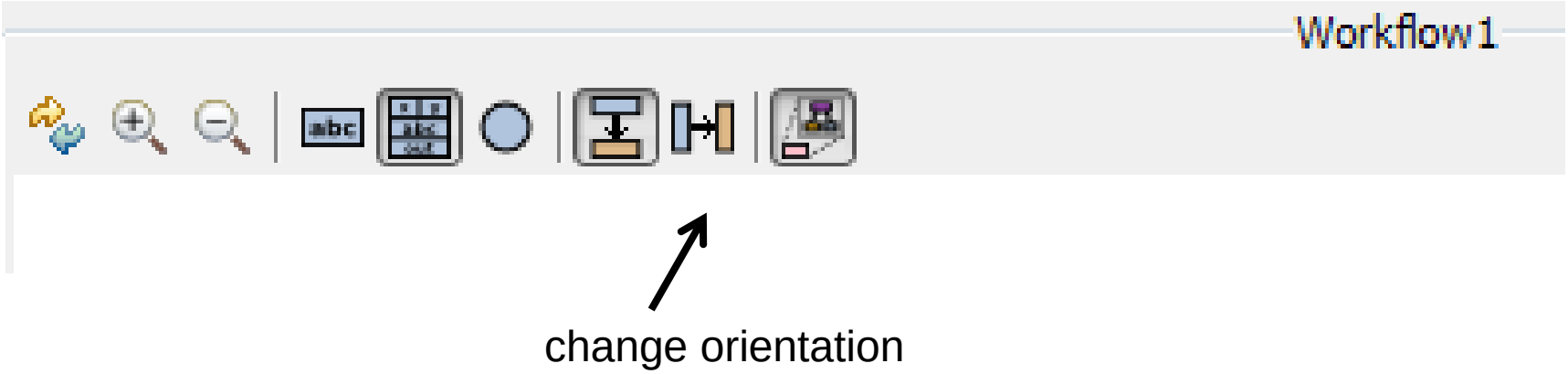
Exercise 1: Building a Simple Workflow

- Let's save the workflow now as "experiments_for_gene"



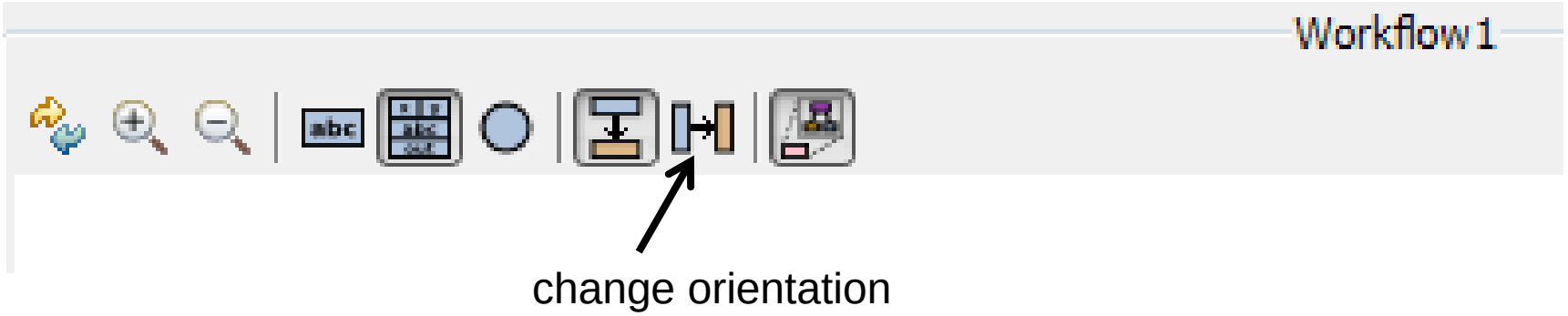
Service ports

- Most of the time, you don't need to connect all ports. Some are optional and some already have default values set.
- Service documentation should tell you this. You can use the BioCatalogue to find documentation and user descriptions
- Change the orientation of the port names to fit them on the screen more easily by clicking on the icon shown below



Service ports

- Most of the time, you don't need to connect all ports. Some are optional and some already have default values set.
- Service documentation should tell you this. You can use the BioCatalogue to find documentation and user descriptions
- You can change the orientation of the port names to fit them on the screen more easily by clicking on the icon shown below



Adding a Workflow Description

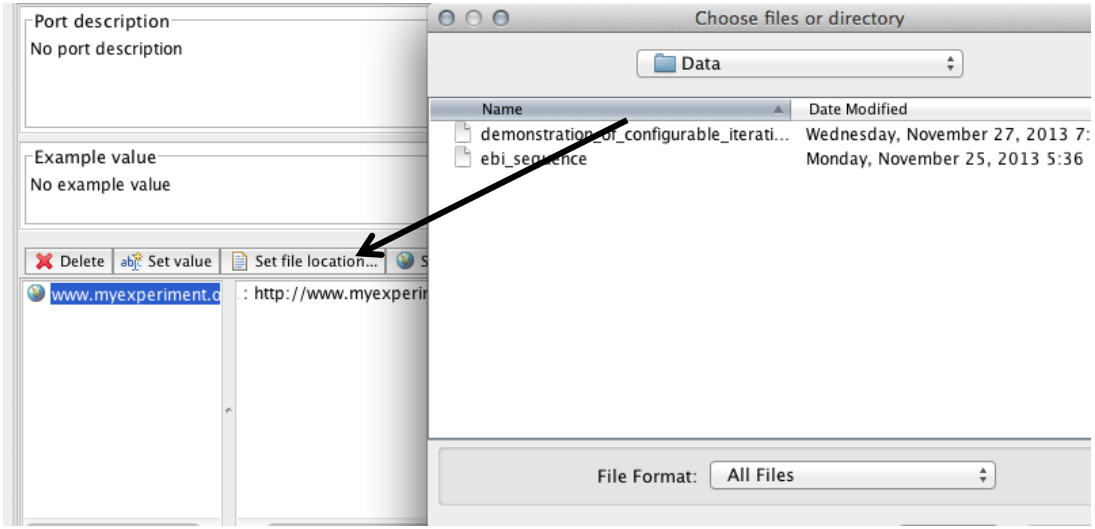
- In the **Design** view, right-click on a blank part of the workflow diagram and select **Annotate**
- Add some details about the workflow e.g. who is the author, what the workflow does
- You can also add examples and descriptions for the workflow inputs by selecting them and selecting **Annotate**
- Add an example for the gene *Adora2a*
- Save the workflow by going to “File -> save workflow”
- Run the workflow again and look at the results

Additional Exercise 1: Xpath Service

- This exercise is optional.
- Our workflow returns the result in the XML format.
- Taverna provides a service which helps to process XML data – **XPath service**
- Go to myExperiment and find *XPath service Tutorial*
- Using this tutorial try adding the **XPath Service** to the workflow to process the XML results

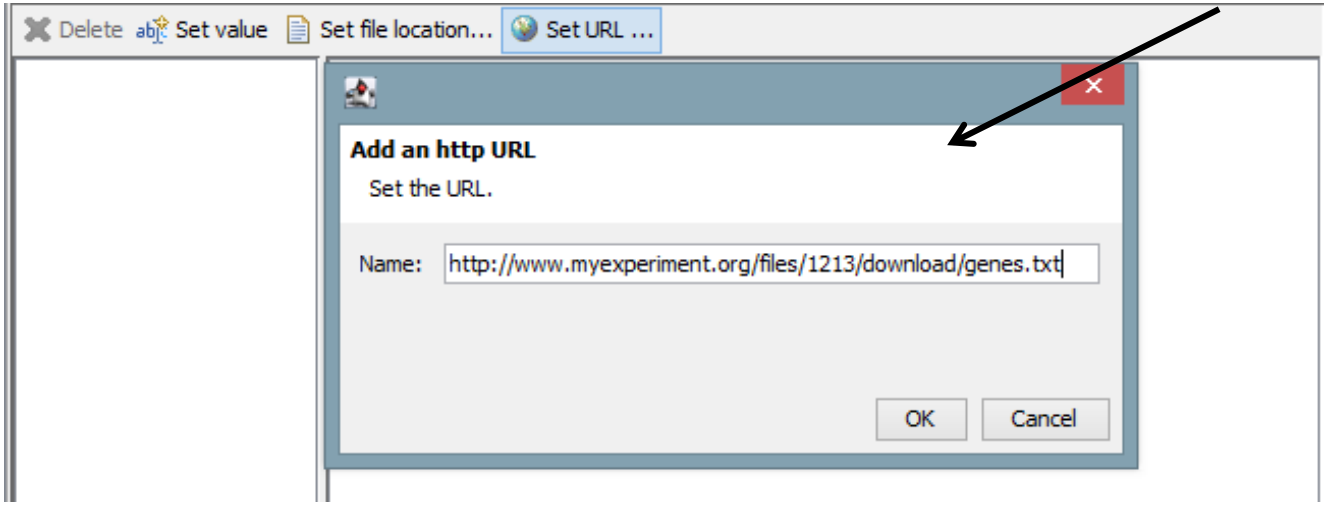
Exercise 2: Entering data from different sources

- We can add input data into the workflow not only manually but also from a file. Go to myExperiment pack and download the file *Genes for VPH-Share workshop*
- Click **Run workflow** again but instead of selecting Set value select **Set file location** and navigate to where you saved the file



Exercise 2: Entering data from different sources

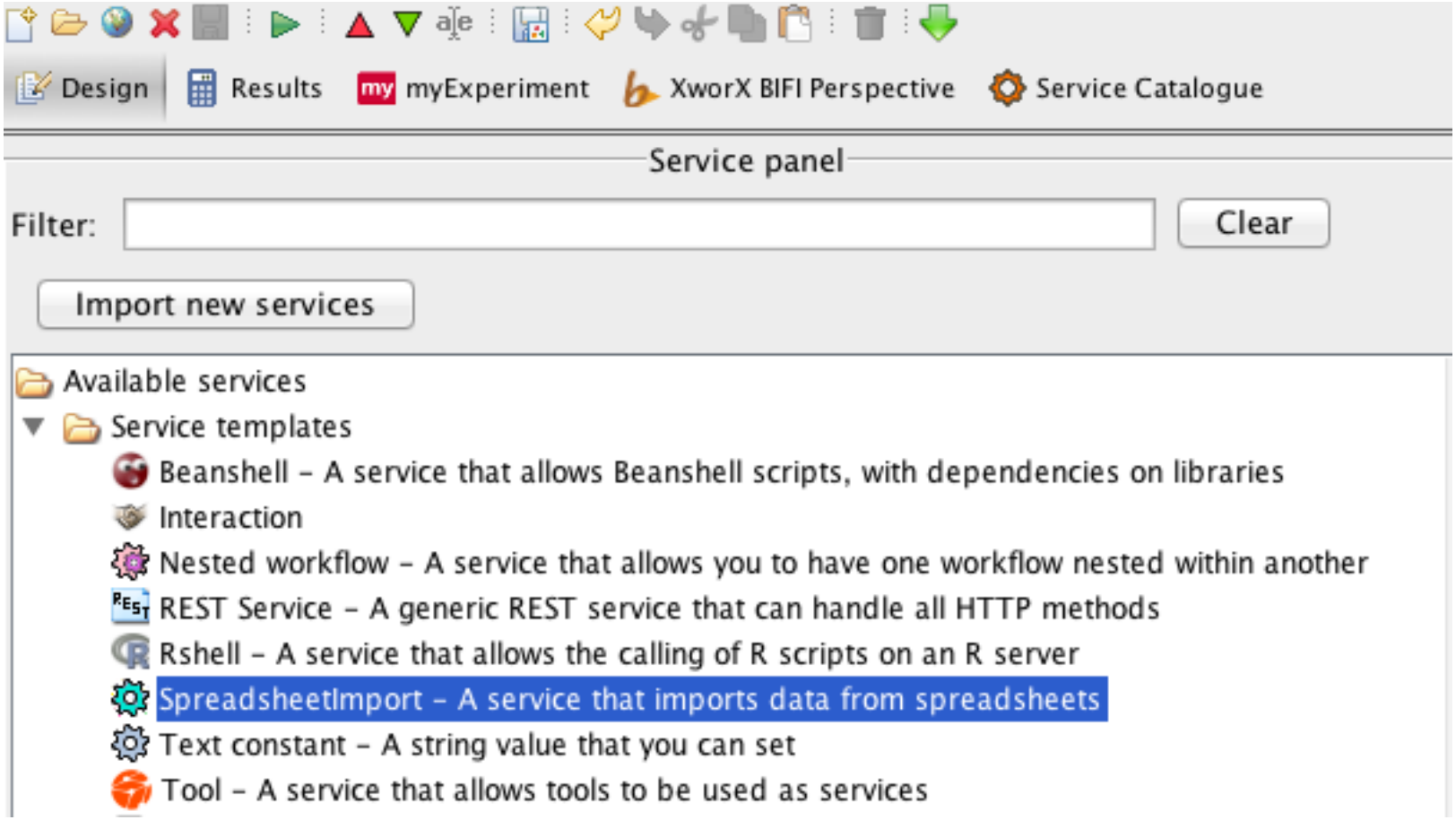
- Instead of downloading the file we can point the workflow to the file's URL (if we know it). Let's run the workflow again but this time select **Set URL** and paste in http://www.myexperiment.org/files/1213/download/genes.txt



Exercise 2: Entering data from different sources

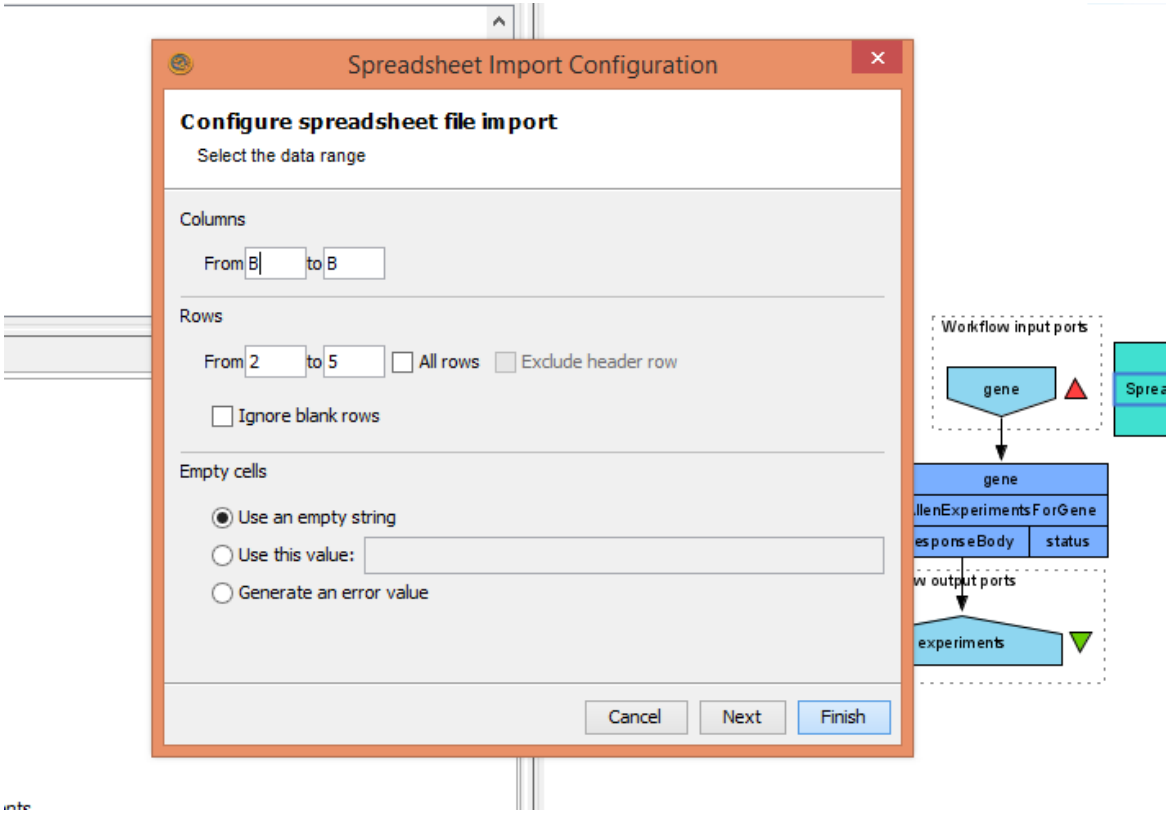
- So far we used simple text files but it is also possible to use spreadsheets as sources of input data. In order to do that we will need to add a Spreadsheet tool to our workflow.
- From the myExperiment group download the file *genes.xls* open it on your machine and see what it contains (the list of the genes is in cells B2 to B5)
- From the **Service Templates** select **SpreadsheetImport** right-click on it and add it to the workflow

Exercise 2: Entering data from different sources



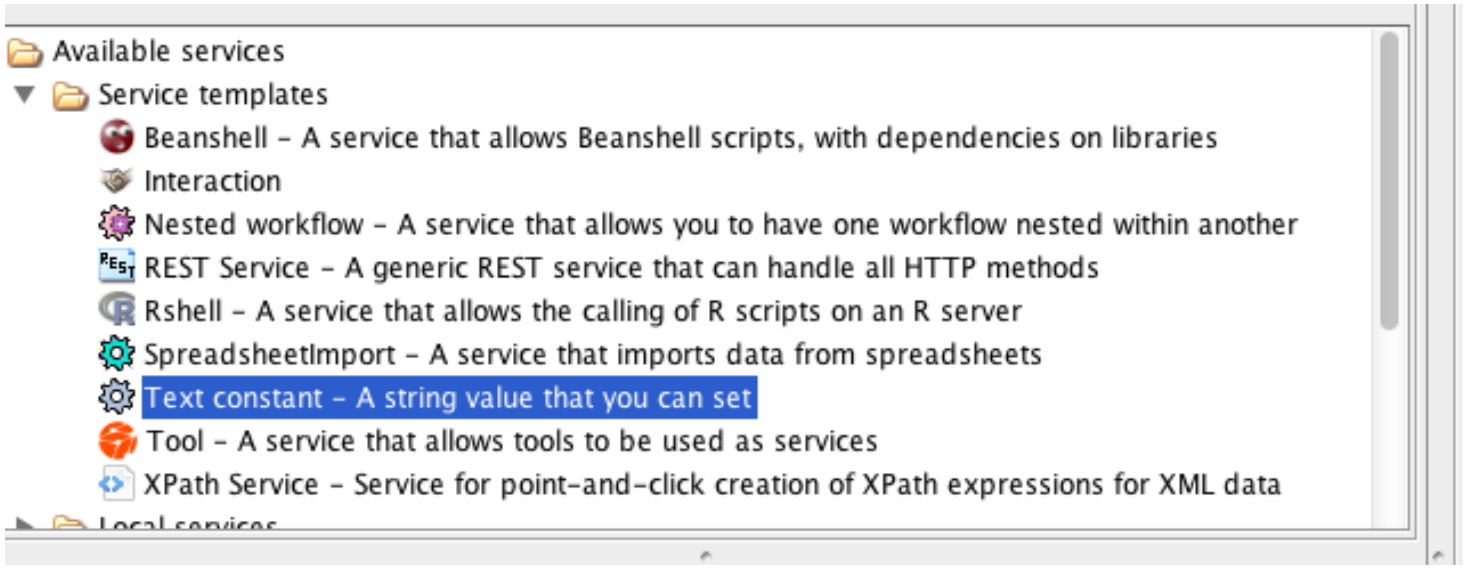
Exercise 2: Entering data from different sources

- In the pop up window set the correct range for columns and rows (untick the box “all rows”)



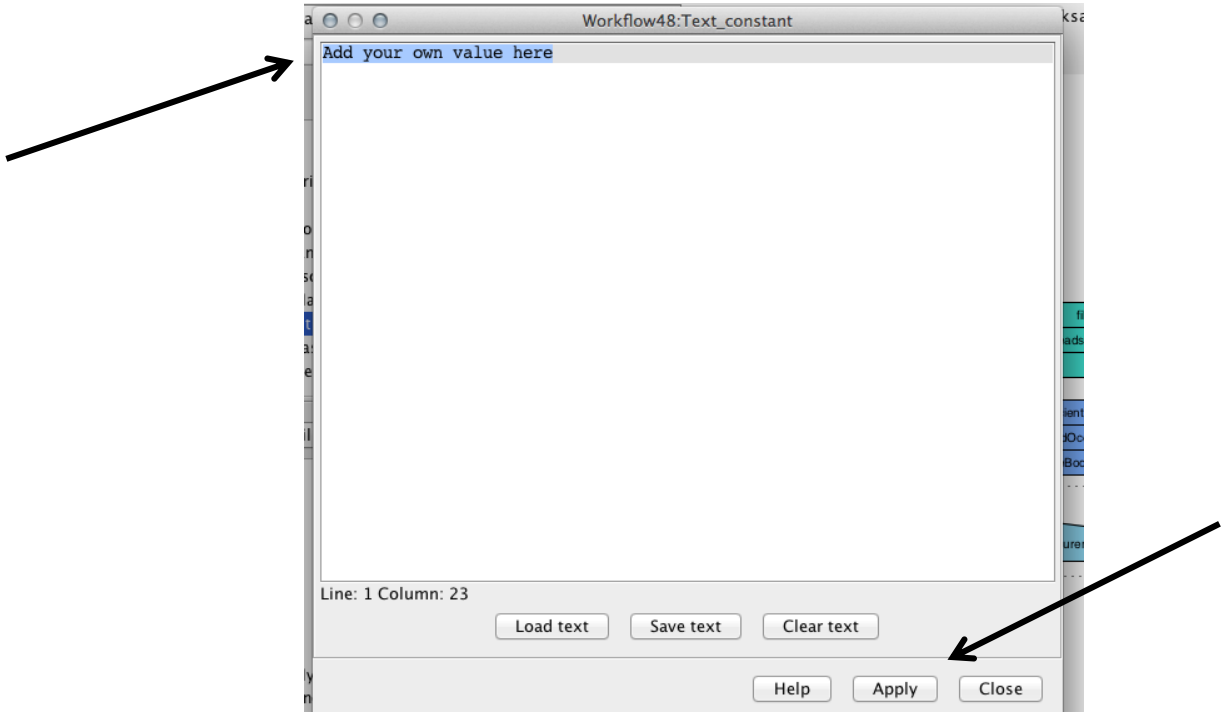
Exercise 2: Entering data from different sources

- We need to delete the input port for the workflow (right click on it and select **Delete**)
- The Spreadsheet tool expects as an input the URL (or path) to the file. The best way to feed in that URL/path is to add a service called **Text constant**



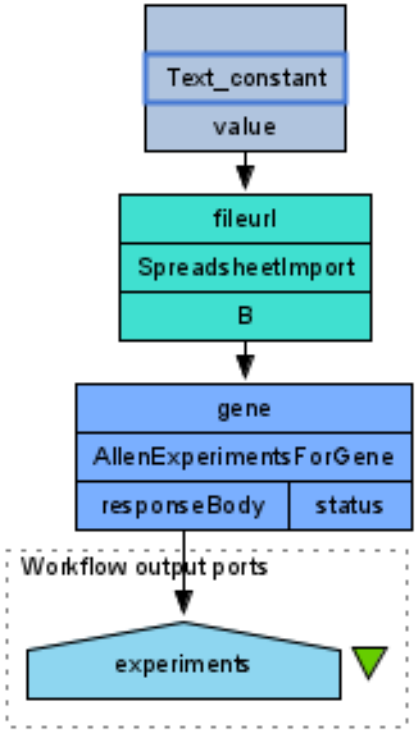
Exercise 2: Entering data from different sources

- Where is says **Add your own value here** enter: *http://www.myexperiment.org/files/1214/download/genes.xls* (or if you prefer the full path to your local file), then **Apply** and **Close**



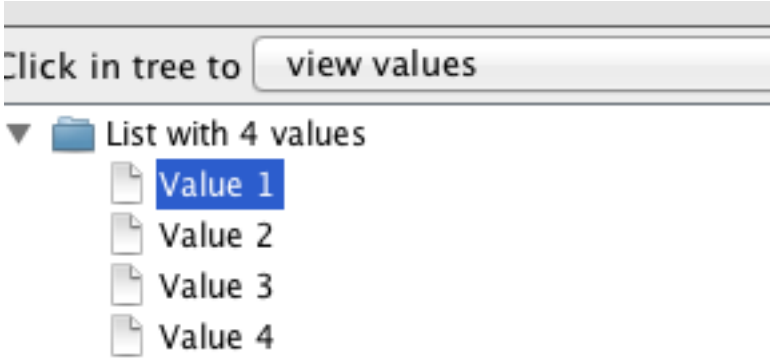
Exercise 2: Entering data from different sources

- Connect the *Text constant* with the *SpreadsheetImport* and the *SpreadsheetImport* with the input to the *AllenExperimentsForGene* service



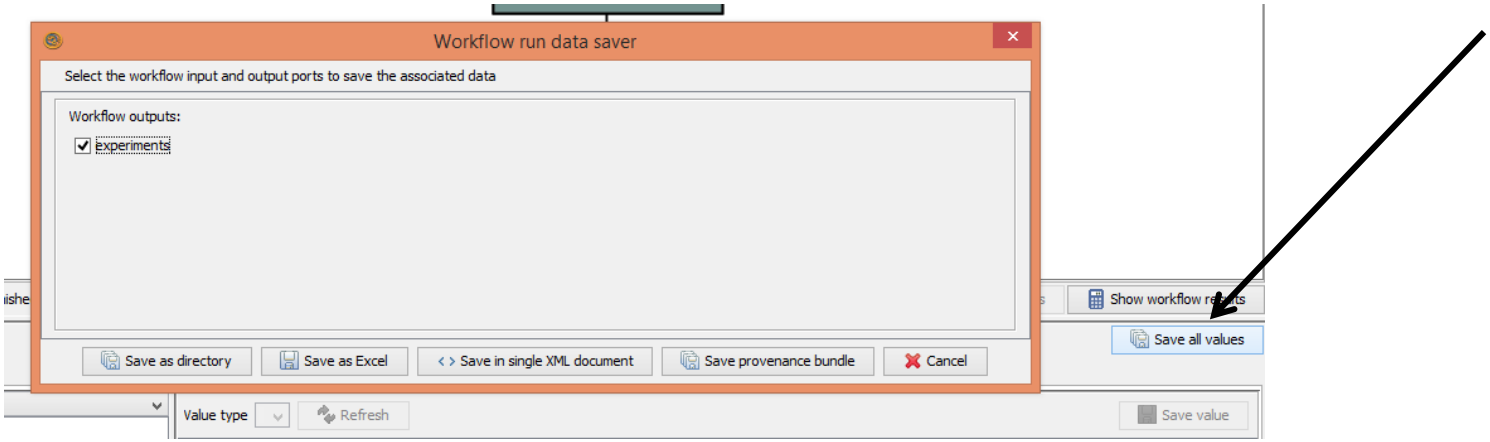
Exercise 2: Entering data from different sources

- When we run the workflow, we can see that there are four values for the results (as there were 4 gene names that we read from the spreadsheet).
- Taverna implicitly iterated over these 4 input values and processed them.



Exercise 3: Saving workflow results

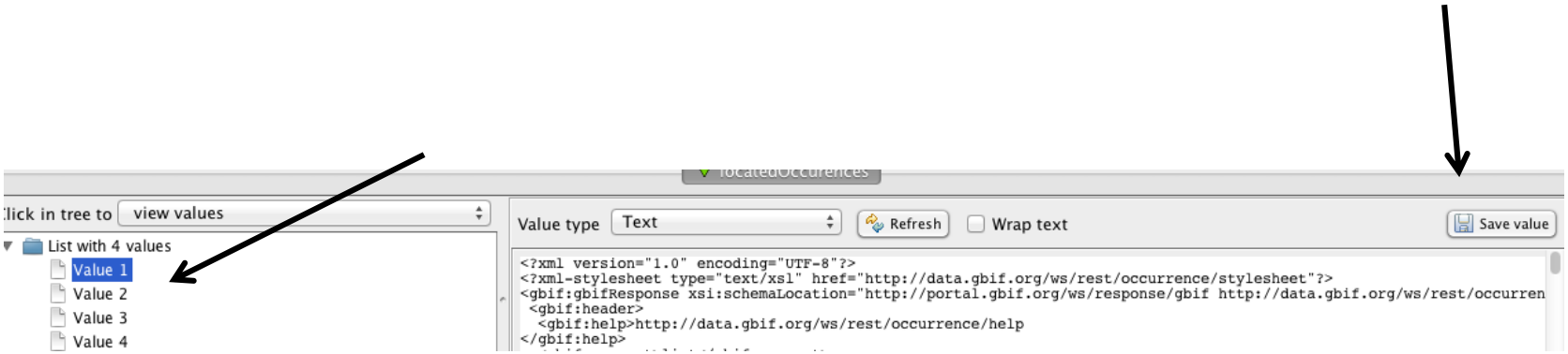
- Taverna allows you to save results in different formats and also allows you to save intermediate workflow results (which is very useful when you run a large workflow)
- You can save all result values:



- Taverna allows you to save values in a variety of formats

Exercise 3: Saving workflow results

- You can also save each single value separately:



- In order to save intermediate values, in the results tab select the part of the workflow which you want to save the values for, then in the results window you should see these values and you will be able to save them

Summary

- You can now
  - Create workflows from a variety of services
  - Connect services together
  - Run workflows
  - Use input values from different sources
  - Save workflows
  - Save results