Inputting data

Aleksandra Pawlik
myGrid Team
University of Manchester

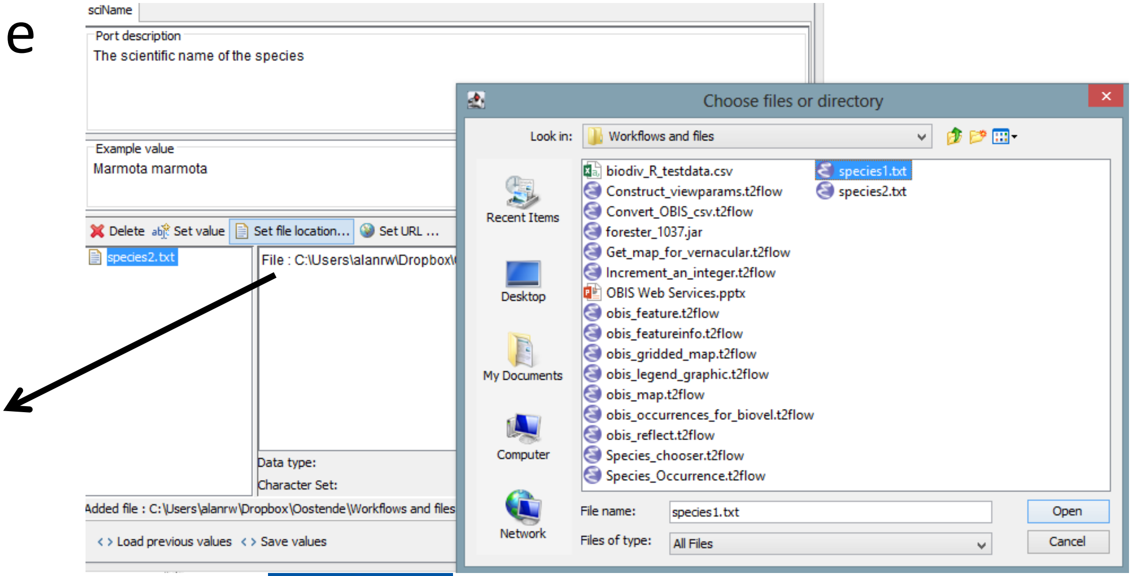
VLIZ, 2014-10-06 / 2014-10-08
http://www.taverna.org.uk/

Re-opening your workflow

- This tutorial uses the Species_Occurrence workflow that you created before
  - Available from http://www.myexperiment.org/workflows/4484.html
- You need to open it in Taverna
- We have already seen how to enter values explicitly with **Set value** and also how to specify them by interaction in a browser
- This tutorial considers some other ways.

Entering data from different sources

- Go to myExperiment group and download a file *Data file for workflow retrieving species occurrence* – call it *species1.txt*
    - *See: http://www.myexperiment.org/files/1051.html*
- Click run workflow again but instead of selecting Set value select Set file location and navigate to where you saved the *species1.txt* file



The scientific name of the species

Marmota marmota

species2.txt

File : C:\Users\alanrw\Dropbox\

Data type:

Character Set:

Added file : C:\Users\alanrw\Dropbox\Oostende\Workflows and files

Look in: Workflows and files

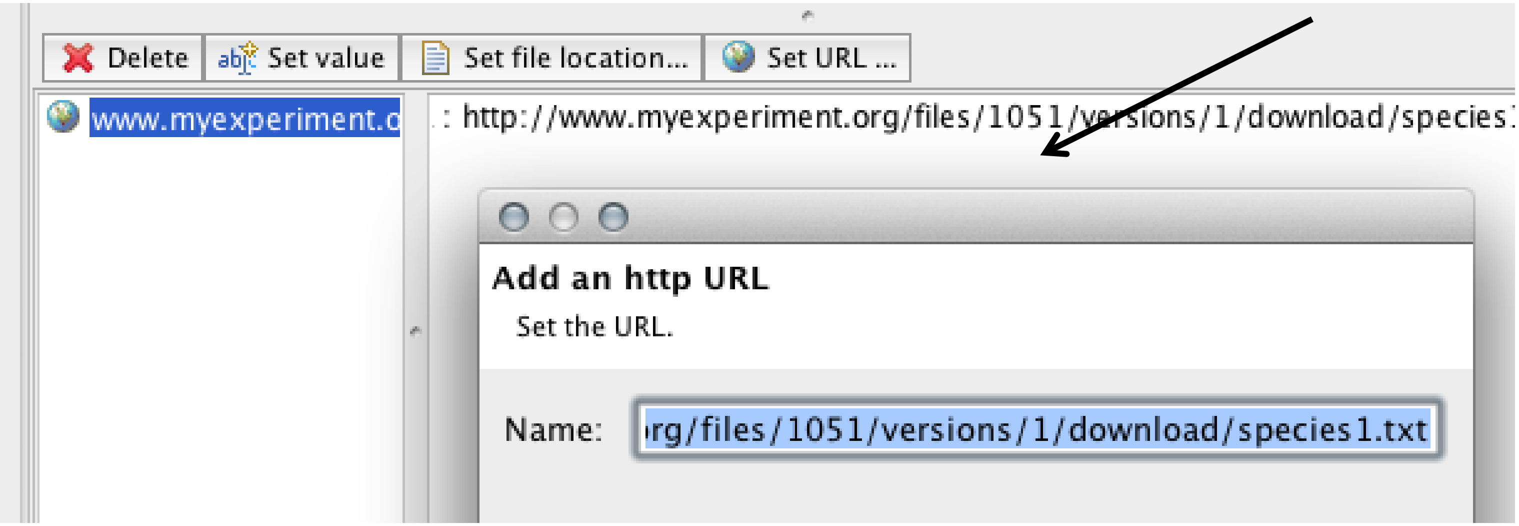
biodiv_R_testdata.csv
Construct_viewparams.t2flow
Convert_OBIS_csv.t2flow
forester_1037.jar
Get_map_for_vernacular.t2flow
Increment_an_integer.t2flow
OBIS Web Services.pptx
obis_feature.t2flow
obis_featureinfo.t2flow
obis_gridded_map.t2flow
obis_legend_graphic.t2flow
obis_map.t2flow
obis_occurrences_for_biovel.t2flow
obis_reflect.t2flow
Species_chooser.t2flow
Species_Occurrence.t2flow
species1.txt
species2.txt

File name: species1.txt

Files of type: All Files

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/1051/versions/1/download/species1.txt



www.myexperiment.o

: http://www.myexperiment.org/files/1051/versions/1/download/species

Add an http URL

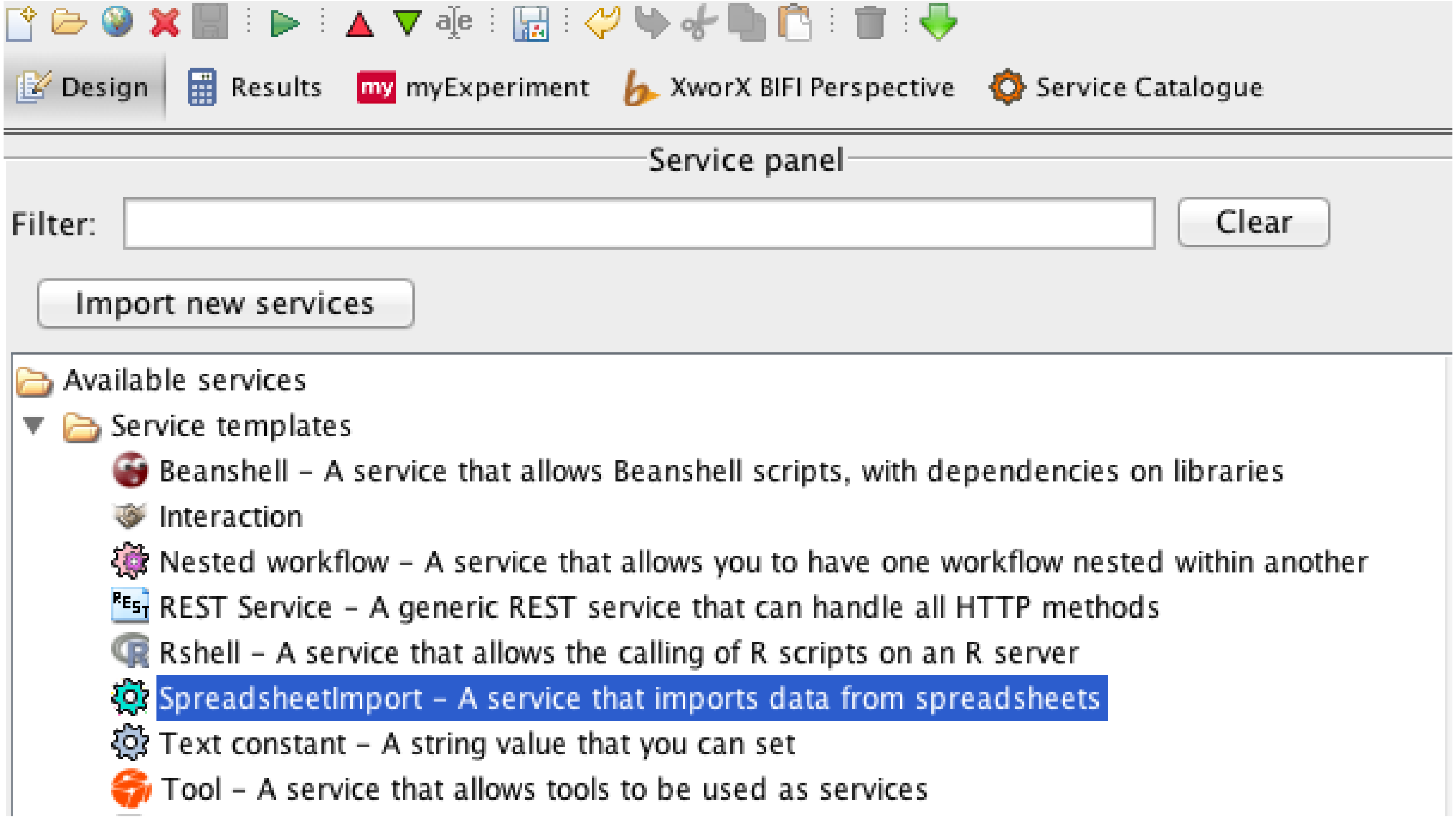
Set the URL.

Name: rg/files/1051/versions/1/download/species1.txt

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 *Spreadsheet file with data for the species workflow species-list.xls*
    - *See: http://www.myexperiment.org/files/1064.html*
- *O*pen it on your machine and see what it contains (the list of the species name is in cells B3 to B6)
- From the **Service Templates** select **Spreadsheet import** right-click on it and add it to the workflow

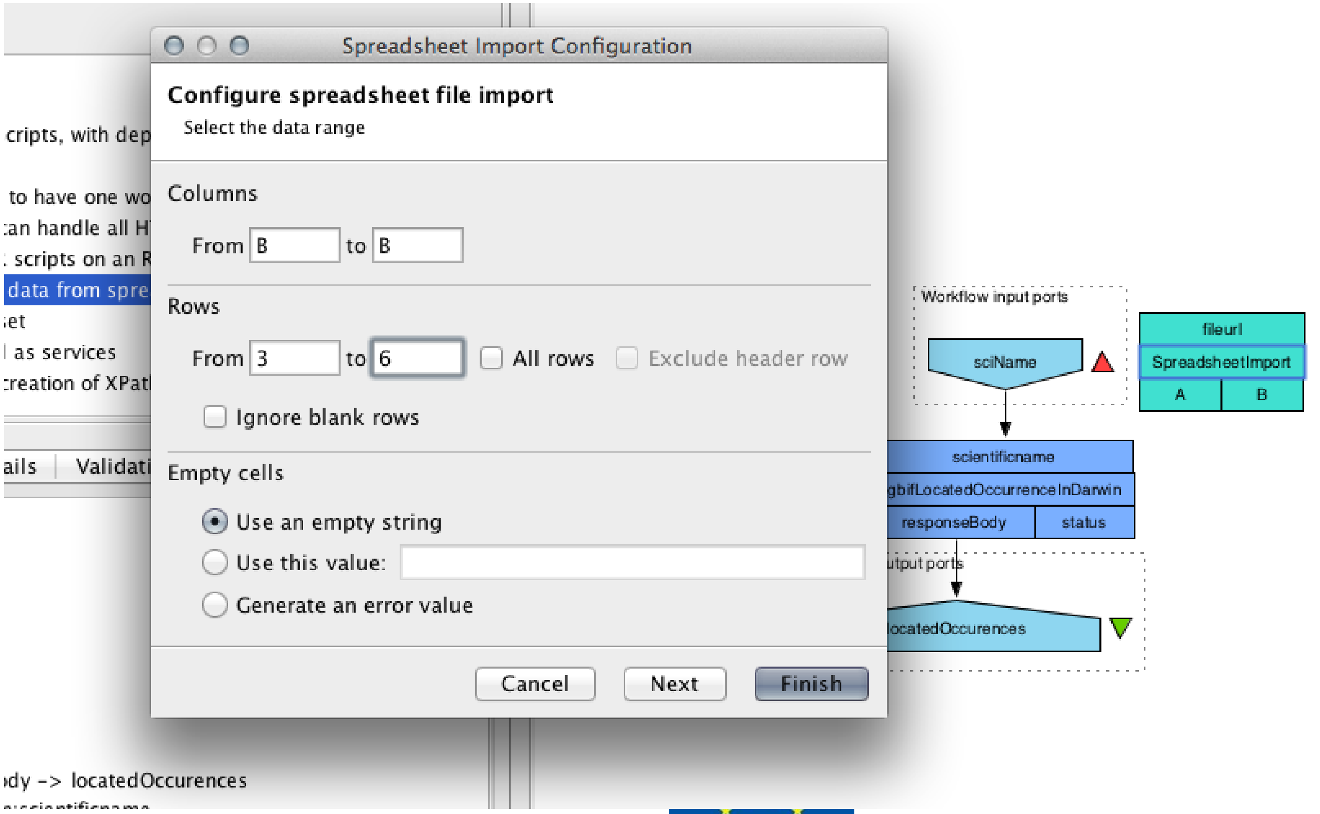
Entering data from different sources



Available services
- Service templates
  - Beanshell – A service that allows Beanshell scripts, with dependencies on libraries
  - Interaction
  - Nested workflow – A service that allows you to have one workflow nested within another
  - REST Service – A generic REST service that can handle all HTTP methods
  - Rshell – A service that allows the calling of R scripts on an R server
  - SpreadsheetImport – A service that imports data from spreadsheets
  - Text constant – A string value that you can set
  - Tool – A service that allows tools to be used as services

Entering data from different sources

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



Spreadsheet Import Configuration

Configure spreadsheet file import

Select the data range

Columns

From B to B

Rows

From 3 to 6 ☐ All rows ☐ Exclude header row

☐ Ignore blank rows

Empty cells

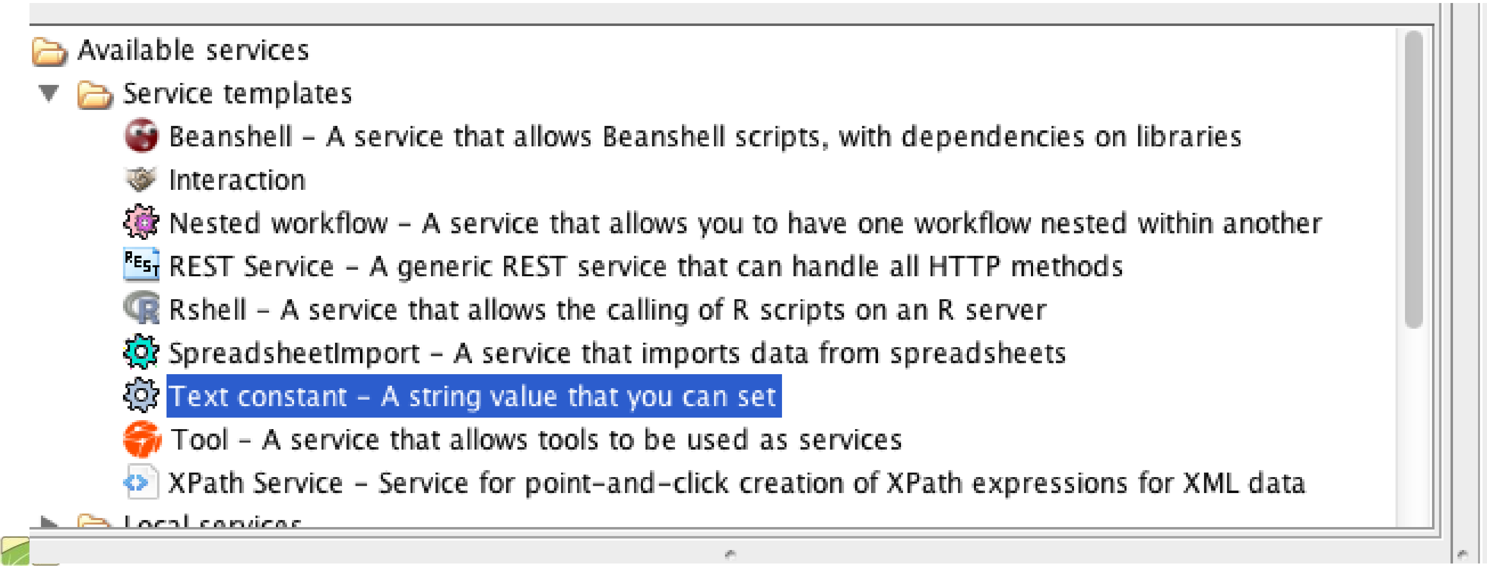
● Use an empty string

○ Use this value:

○ Generate an error value

Entering data from different sources

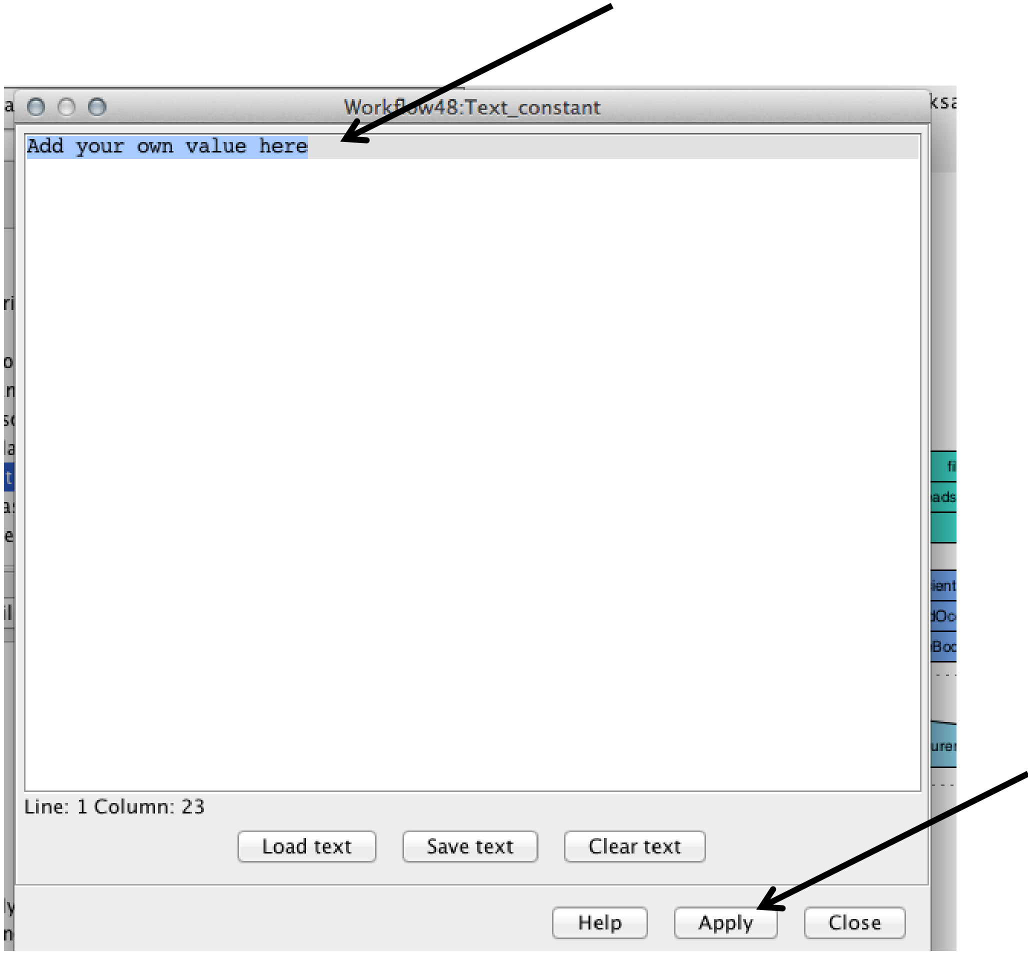
- We need to delete the input port for the workflow (right click on it and select Delete)
- The Spreadsheet tool expects as an input the URL (or path) to the file. The best way to feed in that URL/path is to add a service called “Text constant”



Available services
- Service templates
  - Beanshell – A service that allows Beanshell scripts, with dependencies on libraries
  - Interaction
  - Nested workflow – A service that allows you to have one workflow nested within another
  - REST Service – A generic REST service that can handle all HTTP methods
  - Rshell – A service that allows the calling of R scripts on an R server
  - SpreadsheetImport – A service that imports data from spreadsheets
  - Text constant – A string value that you can set
  - Tool – A service that allows tools to be used as services
  - XPath Service – Service for point-and-click creation of XPath expressions for XML data

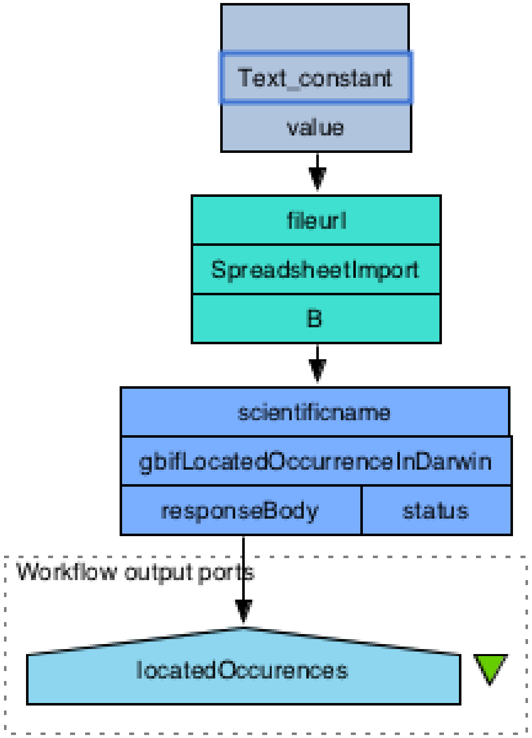
Entering data from different sources

- Where is says "Add your own value here" enter: http://www.myexperiment.org/files/1064/versions/1/download/species-list.xls (or if you prefer the full path to your local file), then Apply and Close



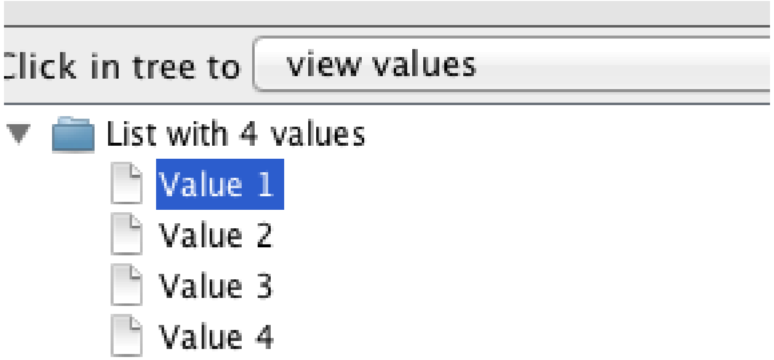
Entering data from different sources

- Connect the Text constant with the Spreadsheet tool and the Spreadsheet tool with the input to the GBIF service



Entering data from different sources

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

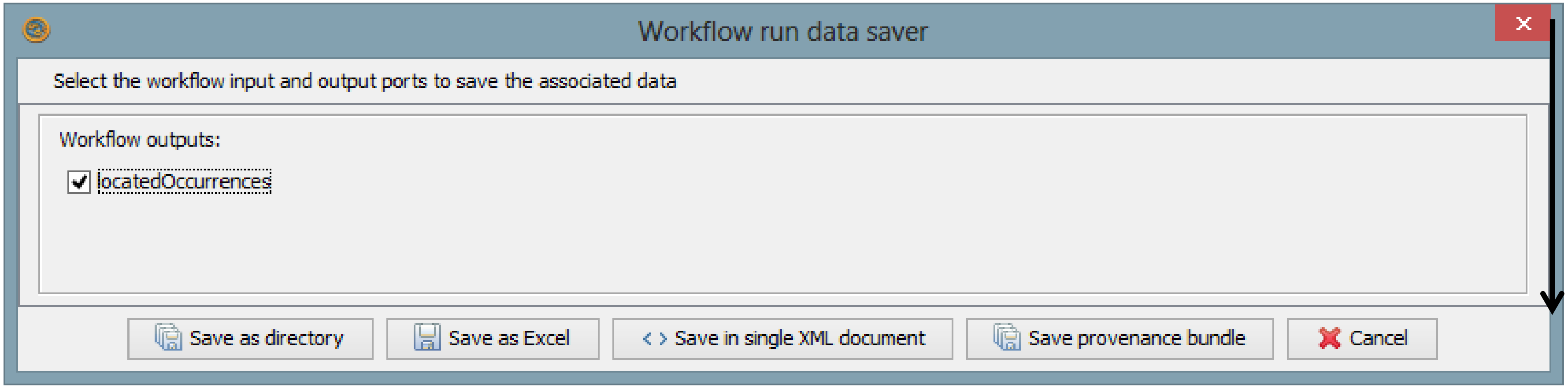


Click in tree to view values

List with 4 values
- Value 1
- Value 2
- Value 3
- Value 4

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:



Workflow run data saver

Select the workflow input and output ports to save the associated data

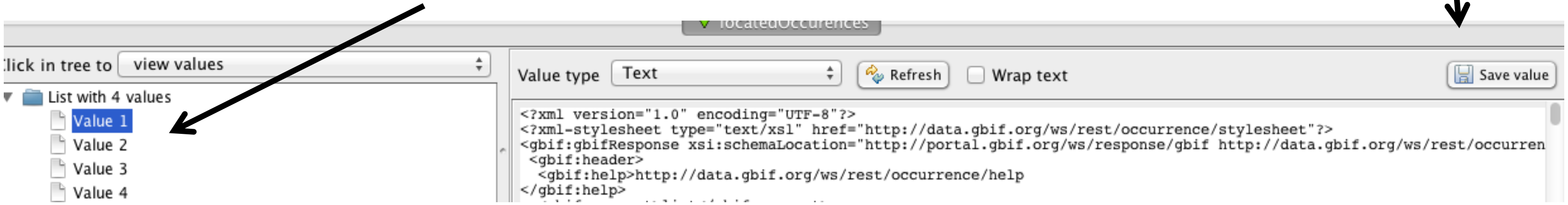
Workflow outputs:

☑ locatedOccurrences

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

Saving workflow results

- You can also save each single value separately:



lick in tree to: view values

Value type: Text

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="http://data.gbif.org/ws/rest/occurrence/stylesheet"?>
<gbif:gbifResponse xsi:schemaLocation="http://portal.gbif.org/ws/response/gbif http://data.gbif.org/ws/rest/occurren
 <gbif:header>
  <gbif:help>http://data.gbif.org/ws/rest/occurrence/help
</gbif:help>
```

- 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