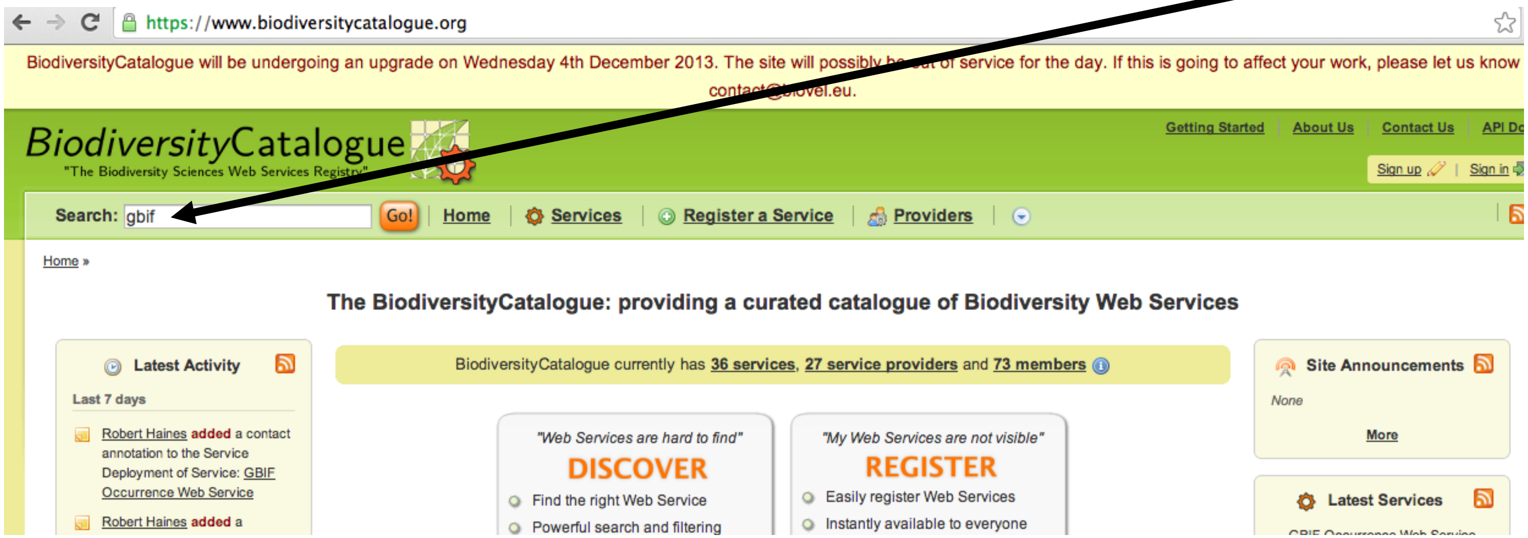
Using BiodiversityCatalogue

Aleksandra Pawlik
myGrid Team
University of Manchester

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

Building a biodiversity Workflow

- We will start with something easy – we will use a GBIF service to retrieve information about the occurrences of a species which name we will provide
- Go to the www.biodiversitycatalogue.org and search for “gbif”



https://www.biodiversitycatalogue.org

BiodiversityCatalogue will be undergoing an upgrade on Wednesday 4th December 2013. The site will possibly be out of service for the day. If this is going to affect your work, please let us know contact@biovel.eu.

The BiodiversityCatalogue: providing a curated catalogue of Biodiversity Web Services

Last 7 days

Robert Haines **added** a contact annotation to the Service Deployment of Service: GBIF Occurrence Web Service

Robert Haines **added** a

BiodiversityCatalogue currently has 36 services, 27 service providers and 73 members

"Web Services are hard to find"

DISCOVER

- Find the right Web Service
- Powerful search and filtering

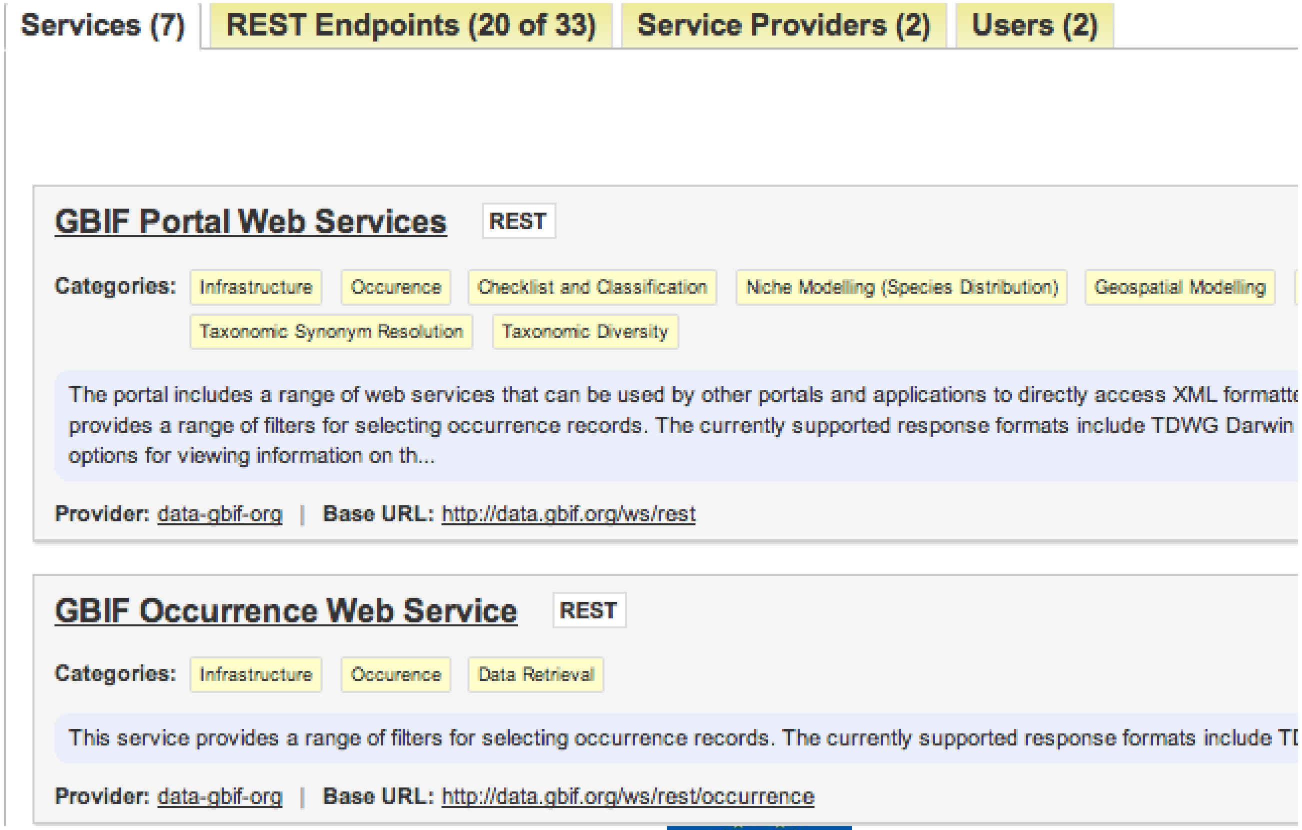
"My Web Services are not visible"

REGISTER

- Easily register Web Services
- Instantly available to everyone

Service selection

- From the results select *GBIF Occurrence Web Service*



GBIF Portal Web Services REST

Categories: Infrastructure | Occurence | Checklist and Classification | Niche Modelling (Species Distribution) | Geospatial Modelling | Taxonomic Synonym Resolution | Taxonomic Diversity

The portal includes a range of web services that can be used by other portals and applications to directly access XML formatte provides a range of filters for selecting occurrence records. The currently supported response formats include TDWG Darwin options for viewing information on th...

Provider: data-gbif-org | **Base URL:** http://data.gbif.org/ws/rest

GBIF Occurrence Web Service REST

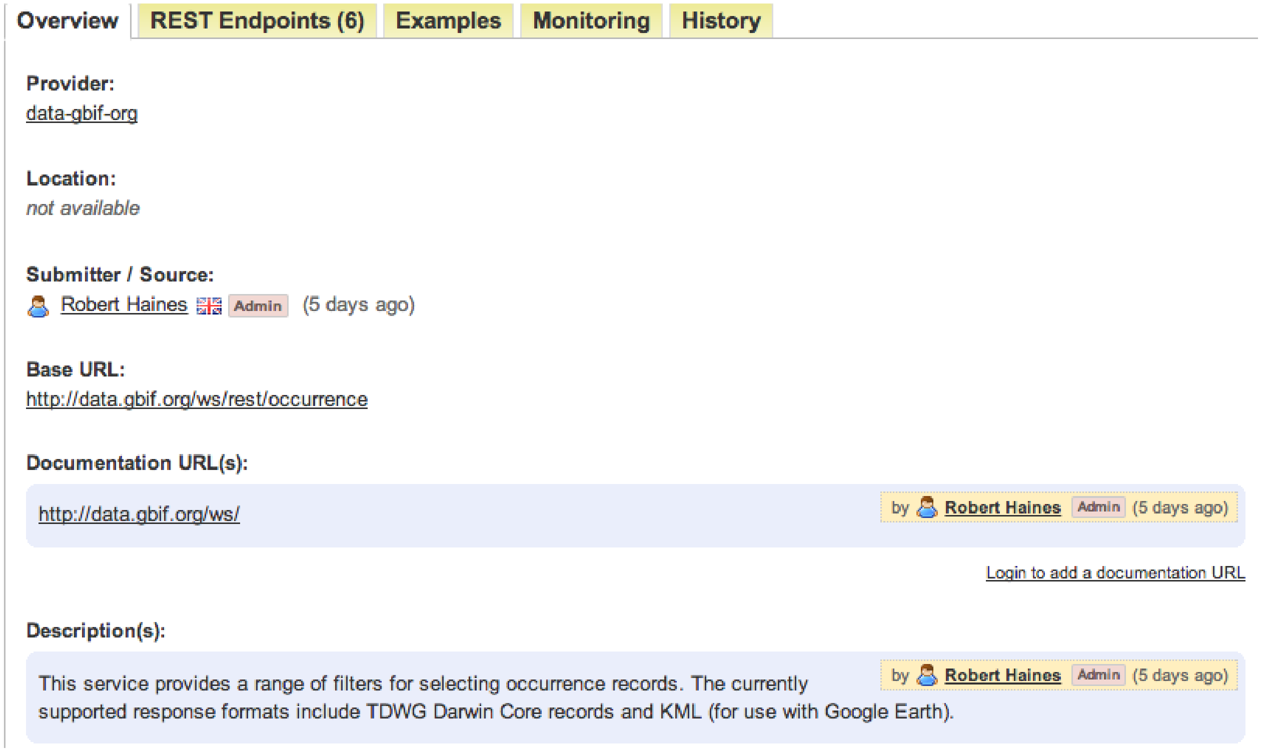
Categories: Infrastructure | Occurence | Data Retrieval

This service provides a range of filters for selecting occurrence records. The currently supported response formats include TI

Provider: data-gbif-org | **Base URL:** http://data.gbif.org/ws/rest/occurrence

Service description

- Have a look at the service description



Provider:
data-gbif-org

Location:
not available

Submitter / Source:
Robert Haines Admin (5 days ago)

Base URL:
http://data.gbif.org/ws/rest/occurrence

Documentation URL(s):

http://data.gbif.org/ws/ by Robert Haines Admin (5 days ago)

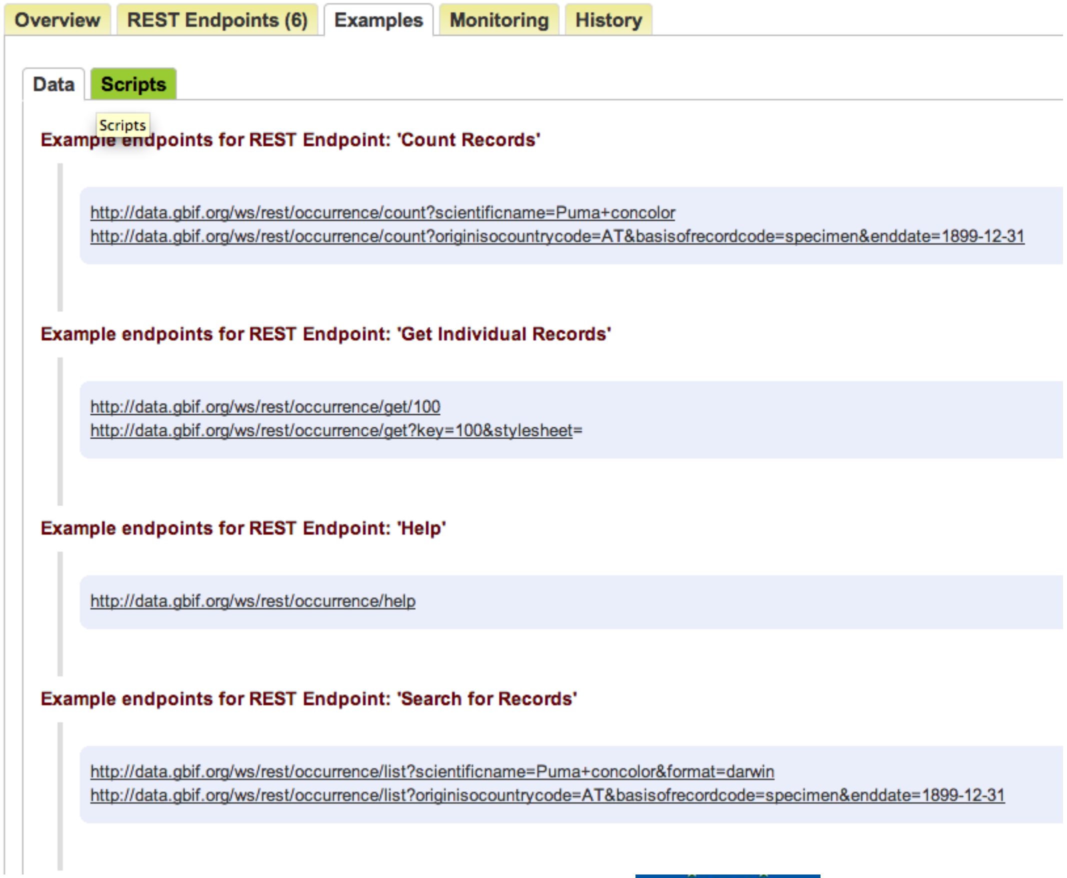
Login to add a documentation URL

Description(s):

This service provides a range of filters for selecting occurrence records. The currently supported response formats include TDWG Darwin Core records and KML (for use with Google Earth). by Robert Haines Admin (5 days ago)

Examples

- Select the **Examples** tab and see how the service can be used



Example endpoints for REST Endpoint: 'Count Records'

http://data.gbif.org/ws/rest/occurrence/count?scientificname=Puma+concolor
http://data.gbif.org/ws/rest/occurrence/count?originisocountrycode=AT&basisofrecordcode=specimen&enddate=1899-12-31

Example endpoints for REST Endpoint: 'Get Individual Records'

http://data.gbif.org/ws/rest/occurrence/get/100
http://data.gbif.org/ws/rest/occurrence/get?key=100&stylesheet=

Example endpoints for REST Endpoint: 'Help'

http://data.gbif.org/ws/rest/occurrence/help

Example endpoints for REST Endpoint: 'Search for Records'

http://data.gbif.org/ws/rest/occurrence/list?scientificname=Puma+concolor&format=darwin
http://data.gbif.org/ws/rest/occurrence/list?originisocountrycode=AT&basisofrecordcode=specimen&enddate=1899-12-31

Using a REST template

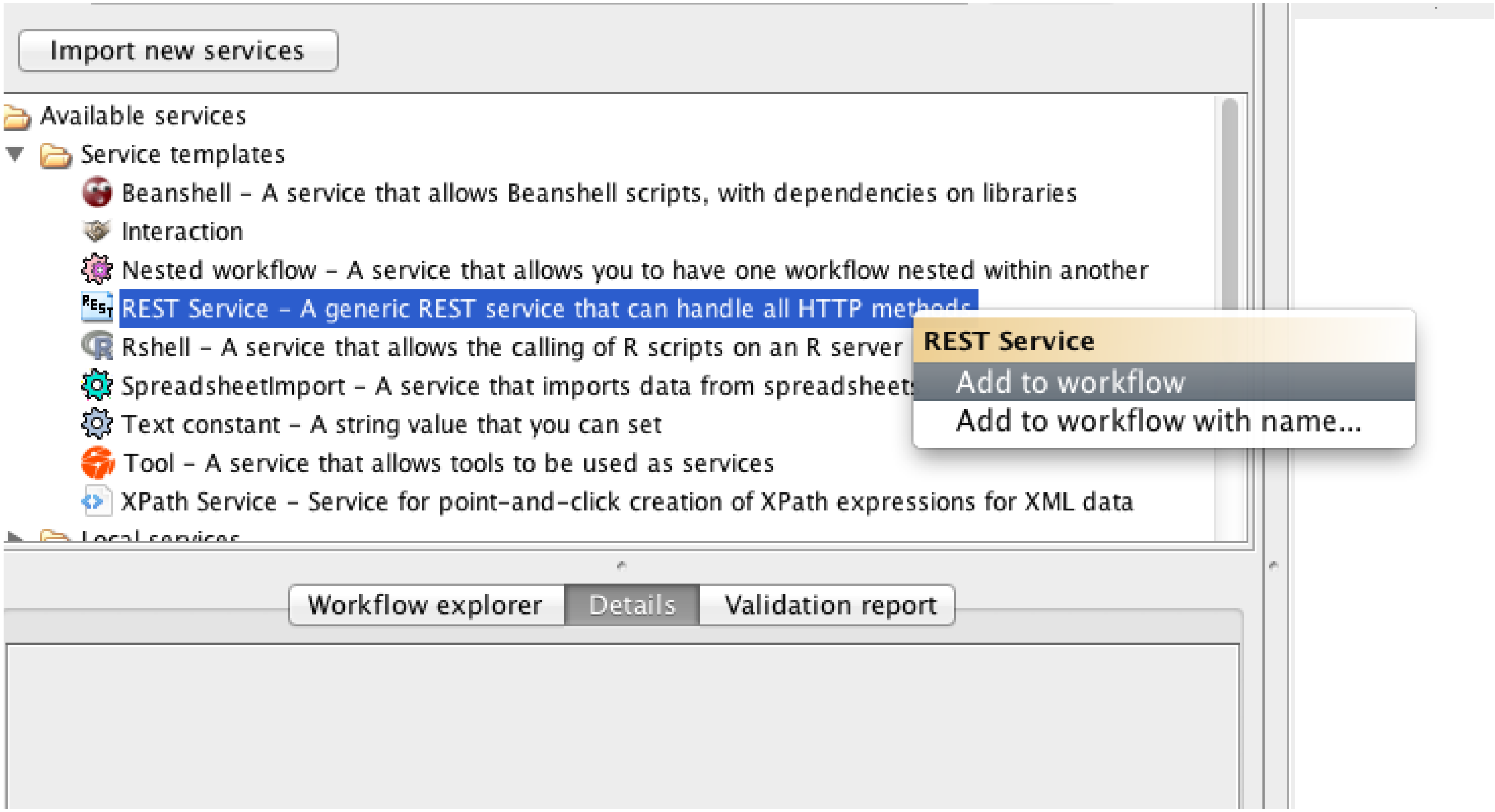
- We want the service to return the results for a species which name we will provide in darwin format, that have coordinates included and we want to limit the number of results to 100 so our REST configuration will look like this:

  http://data.gbif.org/ws/rest/occurrence/list?scientificname={scientificname}&format=darwin&coordinatestatus=true&maxresults=100

❑ The *{scientificname}* means the REST service has a parameter called *scientificname*. Within Taverna, the parameter value can be passed into the service call.

- In Taverna Workbench go to the Services Panel
- From the Available Services select Services Template and REST
- Right-click on it as select Add to workflow (see the next slide)

REST template selection



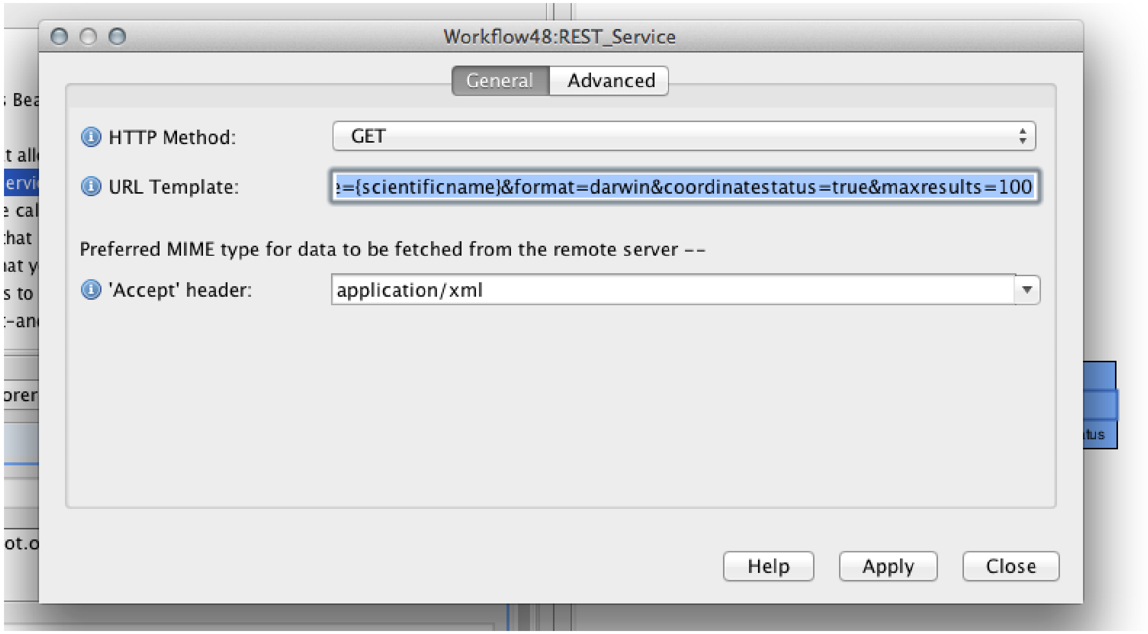
Available services

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 spreadsheet
- 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

Local services

REST template configuration



HTTP Method: GET

URL Template: ={scientificname}&format=darwin&coordinatestatus=true&maxresults=100

Preferred MIME type for data to be fetched from the remote server --

'Accept' header: application/xml

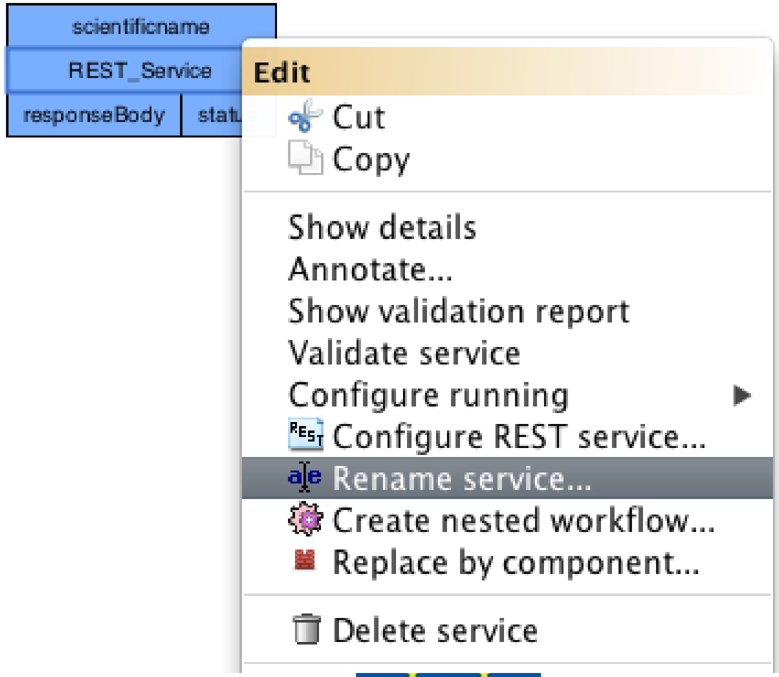
- Enter the following into the URL template field:

http://data.gbif.org/ws/rest/occurrence/list?scientificname={scientificname}&format=darwin&coordinatestatus=true&maxresults=100

- Click **Apply** and **Close** to save the configuration

Service renaming

- Let’s change the name of the service to: gbifLocatedOccurrenceInDarwin



Show ports

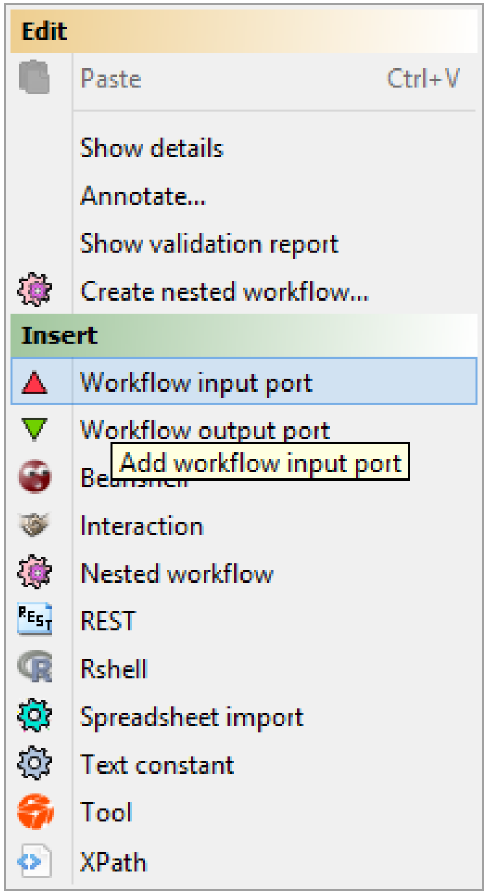
- At the top of the workflow diagram panel, change the view to show all ports by clicking on the icon shown below

Show all ports icon

- This view allows you to see any data input/output or parameter value options for your chosen service
- The REST service should have an input port called *scientificname*

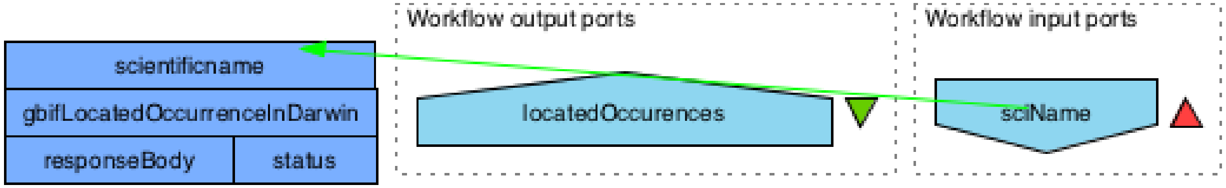
Workflow port creation

- 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. sciName) and click “ok”



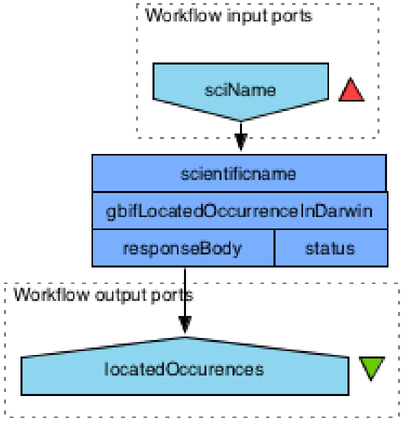
- Do the same to create a new workflow output. Call this output “locatedOccurences”

Workflow connection



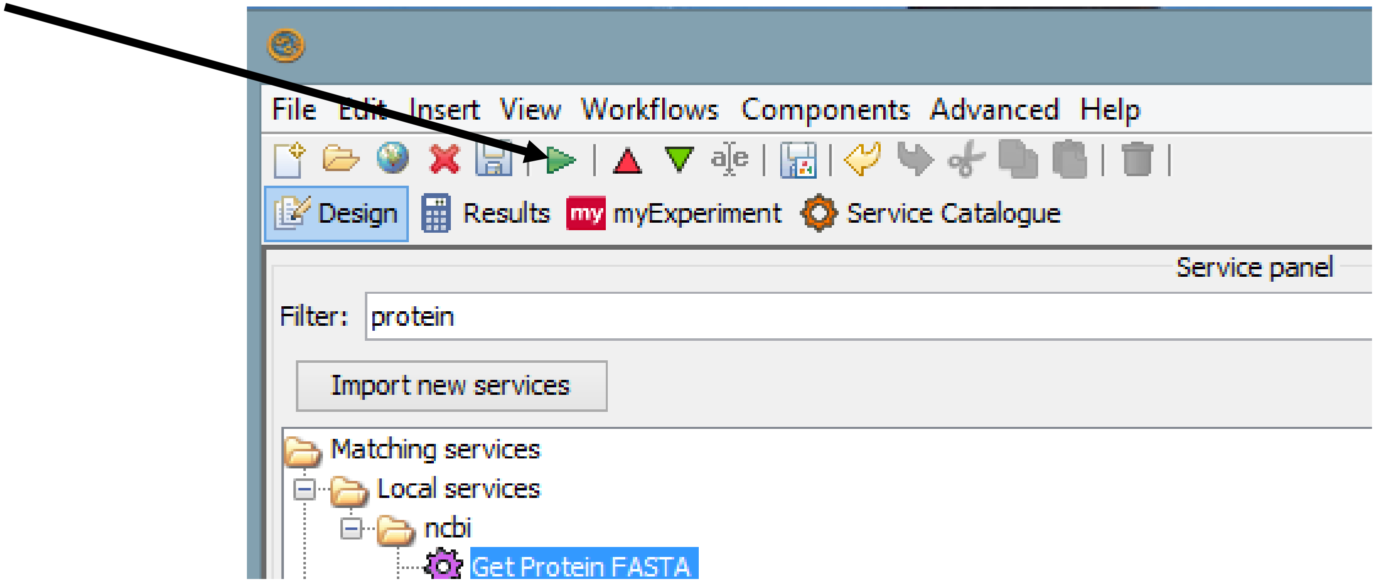
- Connect the input and output ports

- Your workflow should look like this



Running the workflow

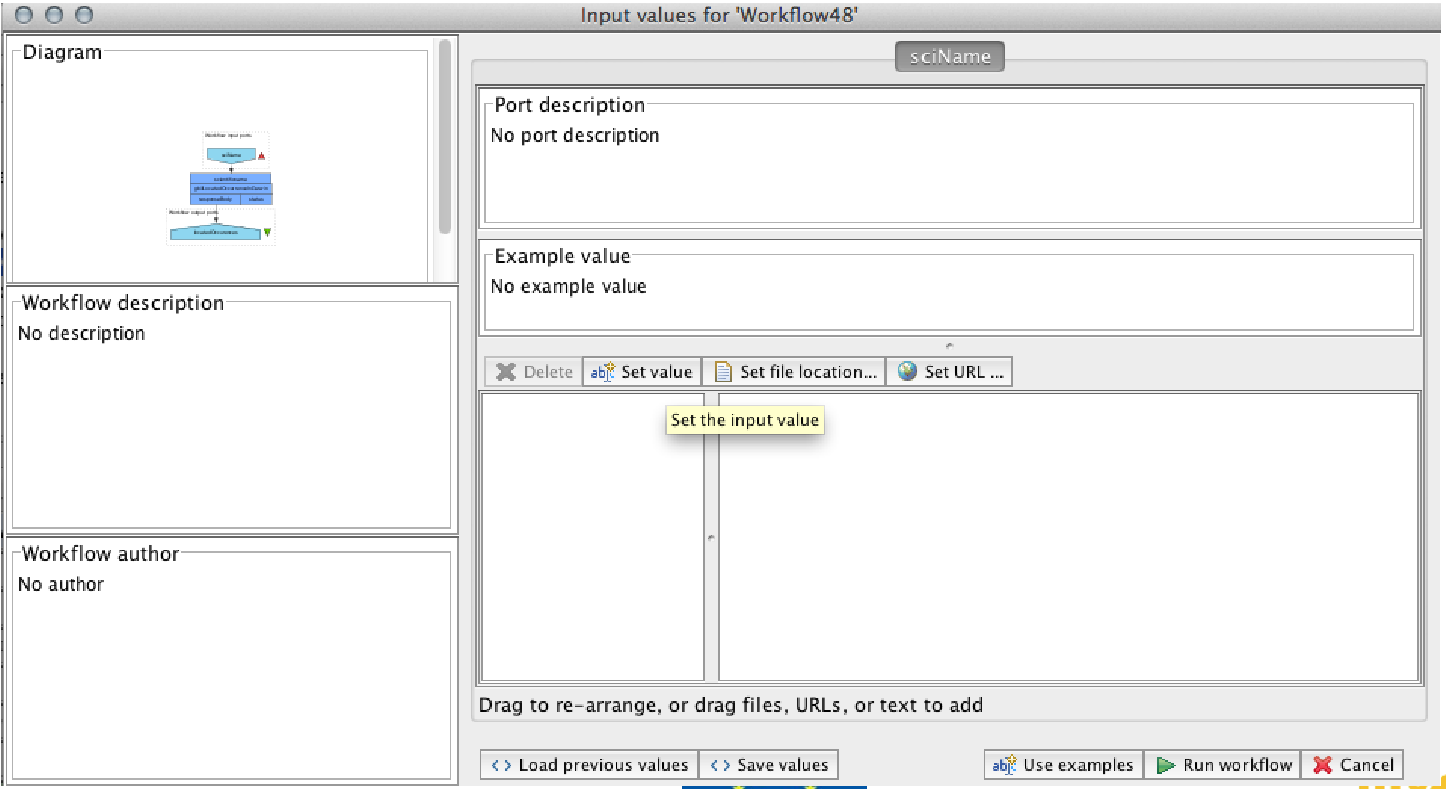
- Run the workflow by selecting “file -> run workflow”, or by clicking on the play button at the top of the workbench



Filter: protein

Setting input values

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



Input values for 'Workflow48'

Diagram

Workflow description

No description

Workflow author

No author

sciName

Port description

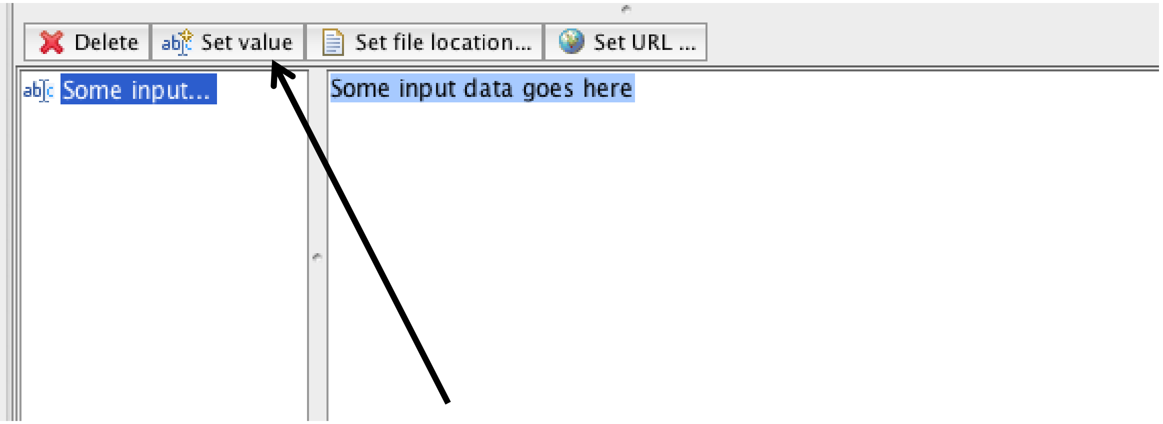
No port description

Example value

No example value

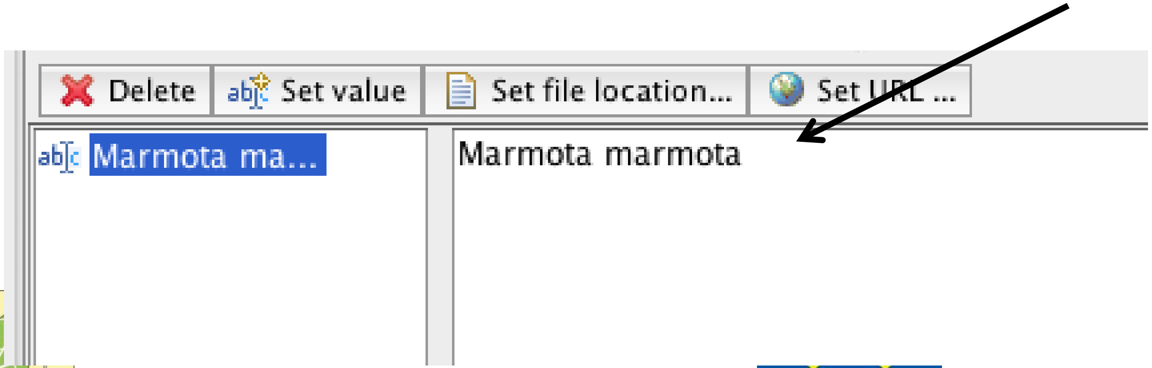
Drag to re-arrange, or drag files, URLs, or text to add

Setting input values



Some input...

Some input data goes here

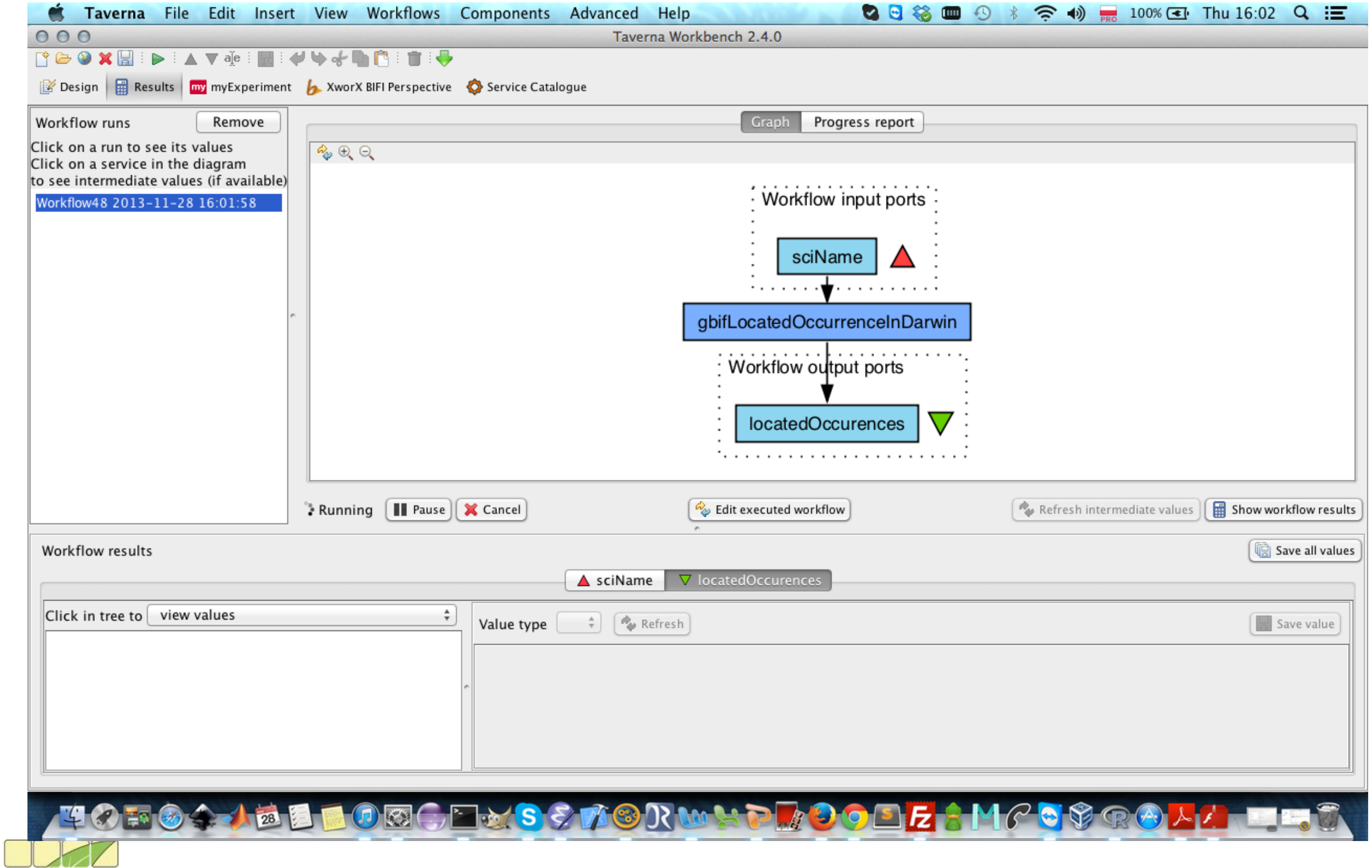


Marmota ma...

Marmota marmota

- Click “Select value” and enter “Marmota marmota” and then at the bottom of the window “Run workflow”

REST workflow run



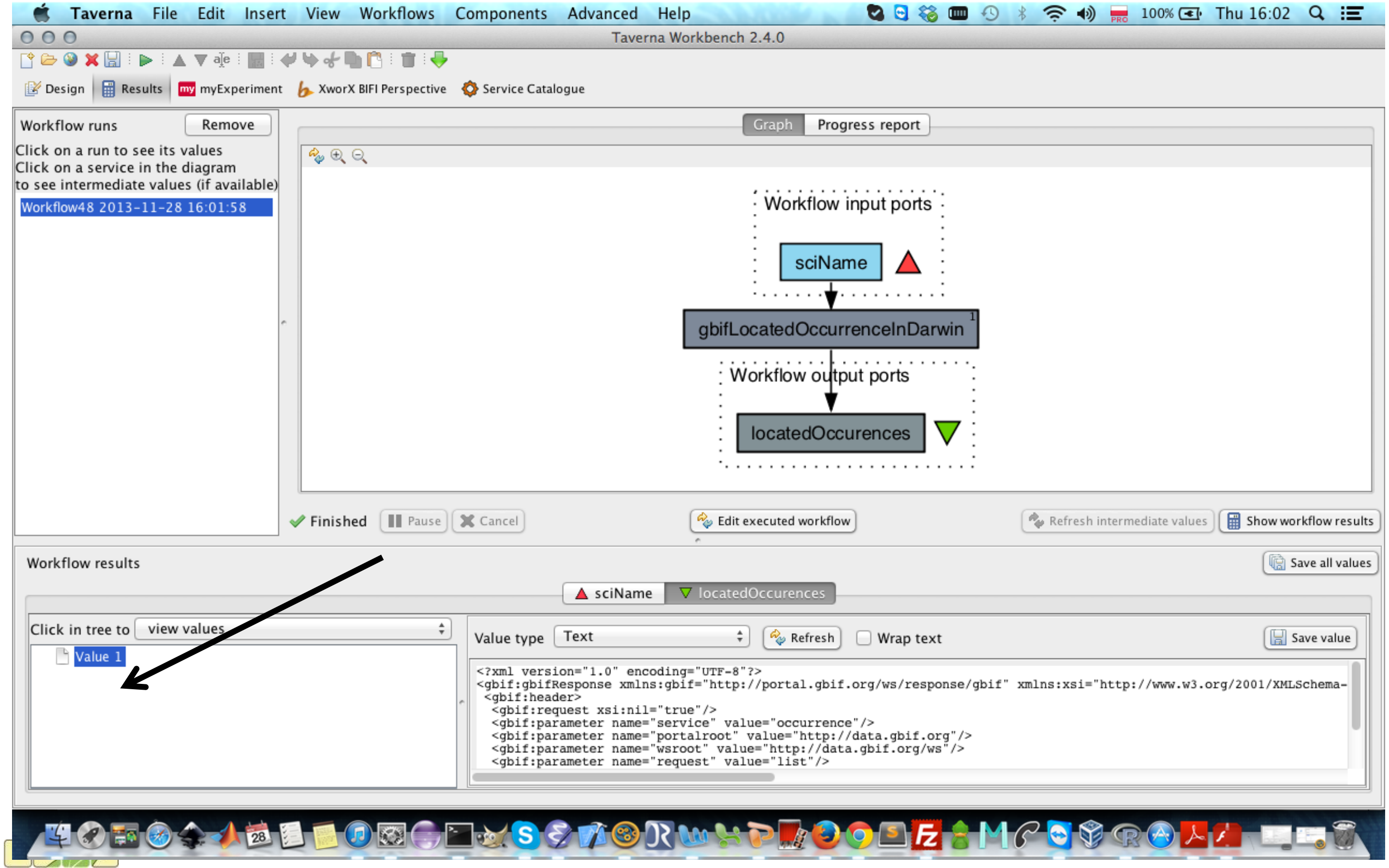
Click on a run to see its values
Click on a service in the diagram
to see intermediate values (if available)

Workflow48 2013-11-28 16:01:58

- You should see the workflow running

REST workflow results

- Once the workflow finished running click on “Value” to see the results



Click on a run to see its values
Click on a service in the diagram
to see intermediate values (if available)

Workflow48 2013-11-28 16:01:58

Click in tree to view values

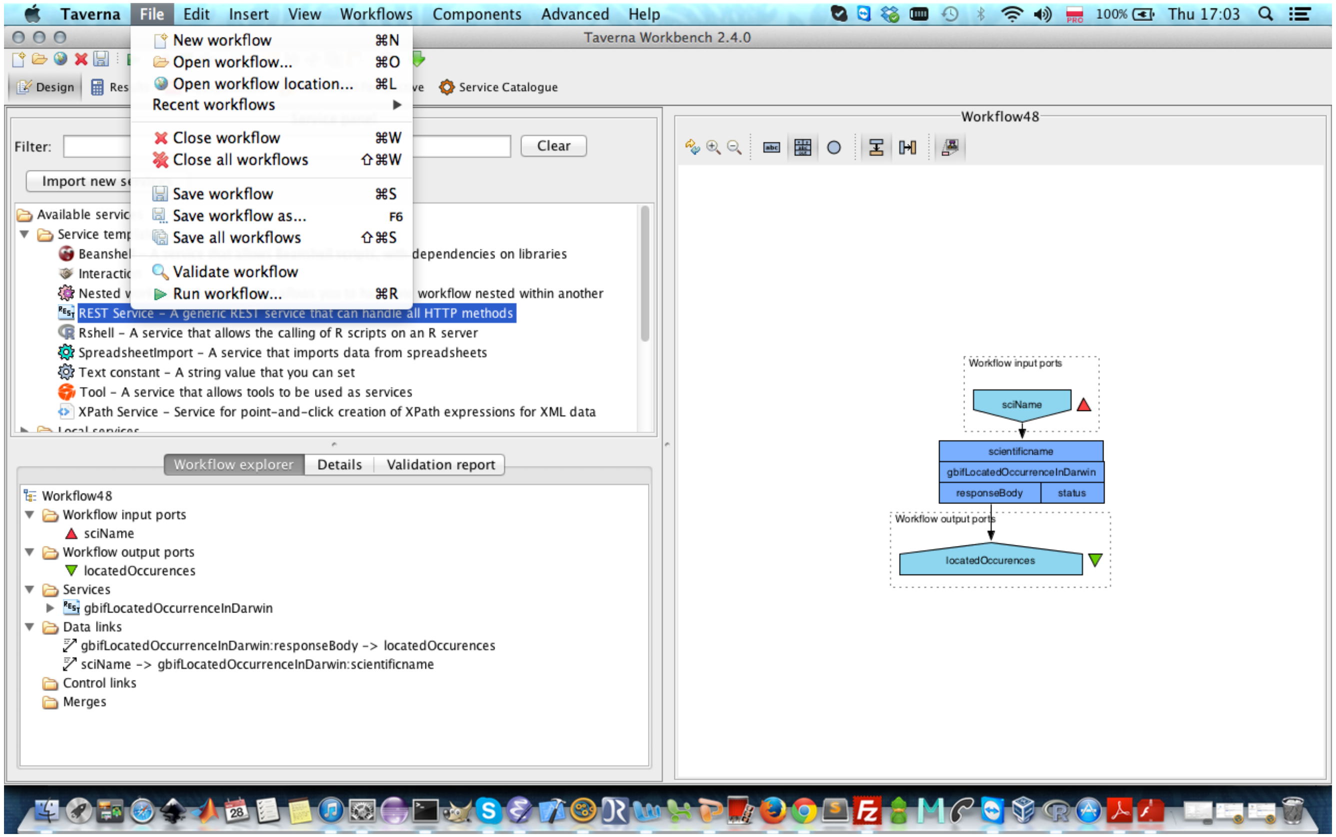
Value type Text

☐ Wrap text

```
<?xml version="1.0" encoding="UTF-8"?>
<gbif:gbifResponse xmlns:gbif="http://portal.gbif.org/ws/response/gbif" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
 <gbif:header>
  <gbif:request xsi:nil="true"/>
  <gbif:parameter name="service" value="occurrence"/>
  <gbif:parameter name="portalroot" value="http://data.gbif.org"/>
  <gbif:parameter name="wsroot" value="http://data.gbif.org/ws"/>
  <gbif:parameter name="request" value="list"/>
```

Saving a workflow

- Let’s save the workflow now as “Species_Occurrence”



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 BiodiversityCatalogue 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

change orientation

Adding a Workflow Description

- 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 does it do
- You can also add examples and descriptions for the workflow inputs by selecting them and selecting “Annotate”
- Add an example for the species “Marmota maromta”
- Save the workflow by going to “File -> save workflow”
- Run the workflow again and look at the results