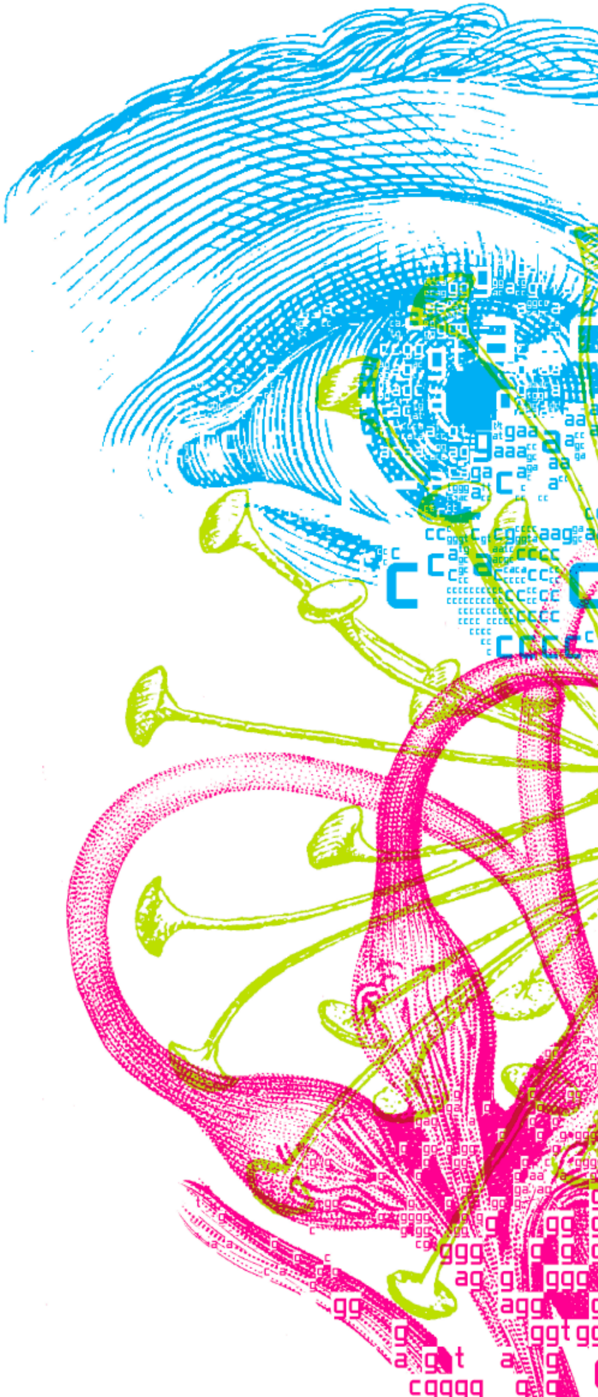
nbic Netherlands
Bioinformatics
Centre

Workflows Accessibility in Bioinformatics

Konstantinos Karasavvas

LUMC
Leiden University Medical Center



Workflows / Pipelines

- Large data sets
  - e.g. NGS
- Data analysis
  - multiple steps
    - automation
    - flexibility
    - reusability
- User groups
  - biologists
  - bioinformaticians

Simple Task

- List all enzymes catalyzing reactions involving a given compound

- Services on the Internet provide this functionality
  - e.g. KEGG services
  - http://www.genome.jp/kegg/

- Scripts
- Galaxy
- Taverna
- Taverna-Galaxy
- Taverna-Web

Script/Program (1)

```
#!/usr/bin/env ruby

require 'soap/wsdlDriver'

wsdl = "http://soap.genome.jp/KEGG.wsdl"
kegg_service = SOAP::WSDLDriverFactory.new(wsdl).create_rpc_driver

enzyme_classifications = []

begin
  reactions = kegg_service.get_reactions_by_compound(ARGV[0])
rescue => err
  puts err.message
end

reactions.each do |reaction|
  begin
     enzyme_classifications << kegg_service.get_enzymes_by_reaction(reaction)
  rescue => err
     puts err.message
  end
end

enzyme_classifications.uniq.each { |ec| puts ec }
```

Script/Program (2)

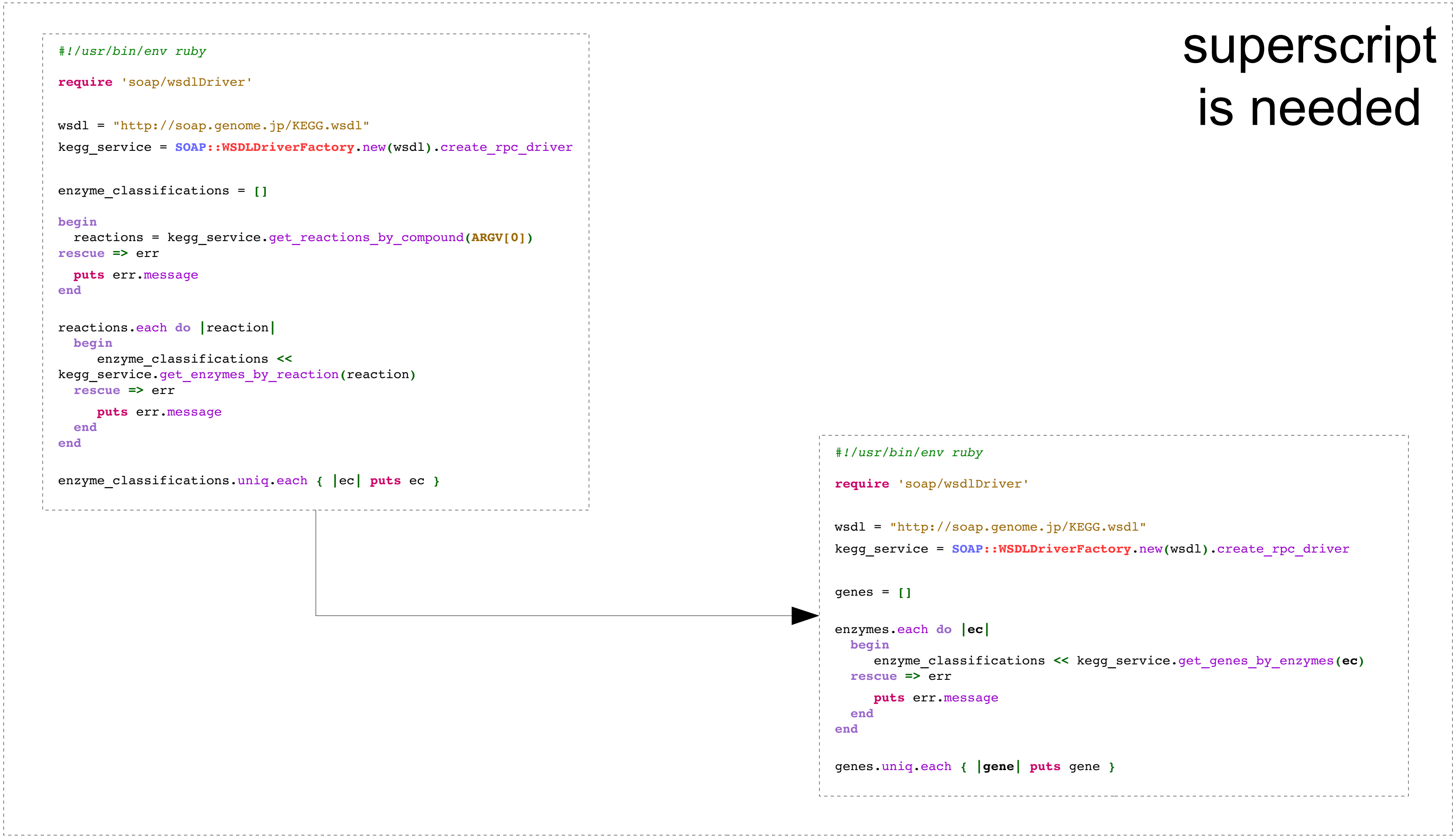
```
$ ./simpleGetEnzymesFromCompoundWkf.rb C15973

ec:2.3.1.12
ec:2.3.1.61
ec:2.3.1.61
ec:2.3.1.168
ec:2.3.1.168
ec:2.3.1.168
ec:1.8.1.4

$
```

Script/Program (2)

```
$ ./simpleGetEnzymesFromCompoundWkf.rb C15973

ec:2.3.1.12
ec:2.3.1.61
ec:2.3.1.61
ec:2.3.1.168
ec:2.3.1.168
ec:2.3.1.168
ec:1.8.1.4

$
```

Who does what?

Multiple Scripts

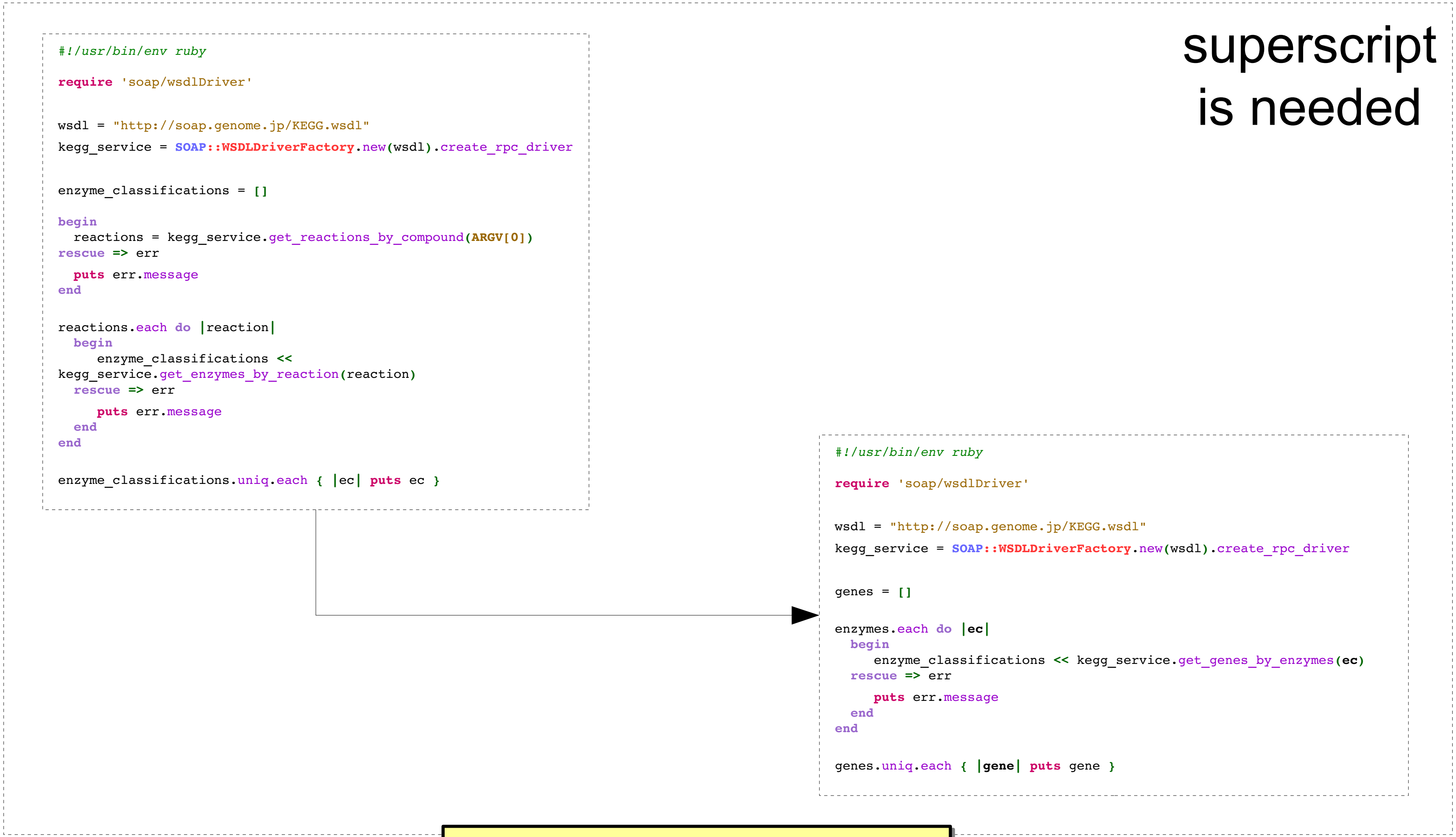


superscript
is needed

```
#!/usr/bin/env ruby

require 'soap/wsdlDriver'


wsdl = "http://soap.genome.jp/KEGG.wsdl"
kegg_service = SOAP::WSDLDriverFactory.new(wsdl).create_rpc_driver


enzyme_classifications = []

begin
  reactions = kegg_service.get_reactions_by_compound(ARGV[0])
rescue => err
  puts err.message
end

reactions.each do |reaction|
  begin
     enzyme_classifications <<
kegg_service.get_enzymes_by_reaction(reaction)
  rescue => err
     puts err.message
  end
end

enzyme_classifications.uniq.each { |ec| puts ec }
```

```
#!/usr/bin/env ruby

require 'soap/wsdlDriver'


wsdl = "http://soap.genome.jp/KEGG.wsdl"
kegg_service = SOAP::WSDLDriverFactory.new(wsdl).create_rpc_driver


genes = []

enzymes.each do |ec|
  begin
     enzyme_classifications << kegg_service.get_genes_by_enzymes(ec)
  rescue => err
     puts err.message
  end
end

genes.uniq.each { |gene| puts gene }
```

Multiple Scripts



```
#!/usr/bin/env ruby

require 'soap/wsdlDriver'


wsdl = "http://soap.genome.jp/KEGG.wsdl"
kegg_service = SOAP::WSDLDriverFactory.new(wsdl).create_rpc_driver


enzyme_classifications = []

begin
  reactions = kegg_service.get_reactions_by_compound(ARGV[0])
rescue => err
  puts err.message
end

reactions.each do |reaction|
  begin
     enzyme_classifications <<
kegg_service.get_enzymes_by_reaction(reaction)
  rescue => err
     puts err.message
  end
end

enzyme_classifications.uniq.each { |ec| puts ec }
```

```
#!/usr/bin/env ruby

require 'soap/wsdlDriver'


wsdl = "http://soap.genome.jp/KEGG.wsdl"
kegg_service = SOAP::WSDLDriverFactory.new(wsdl).create_rpc_driver


genes = []

enzymes.each do |ec|
  begin
     enzyme_classifications << kegg_service.get_genes_by_enzymes(ec)
  rescue => err
     puts err.message
  end
end

genes.uniq.each { |gene| puts gene }
```

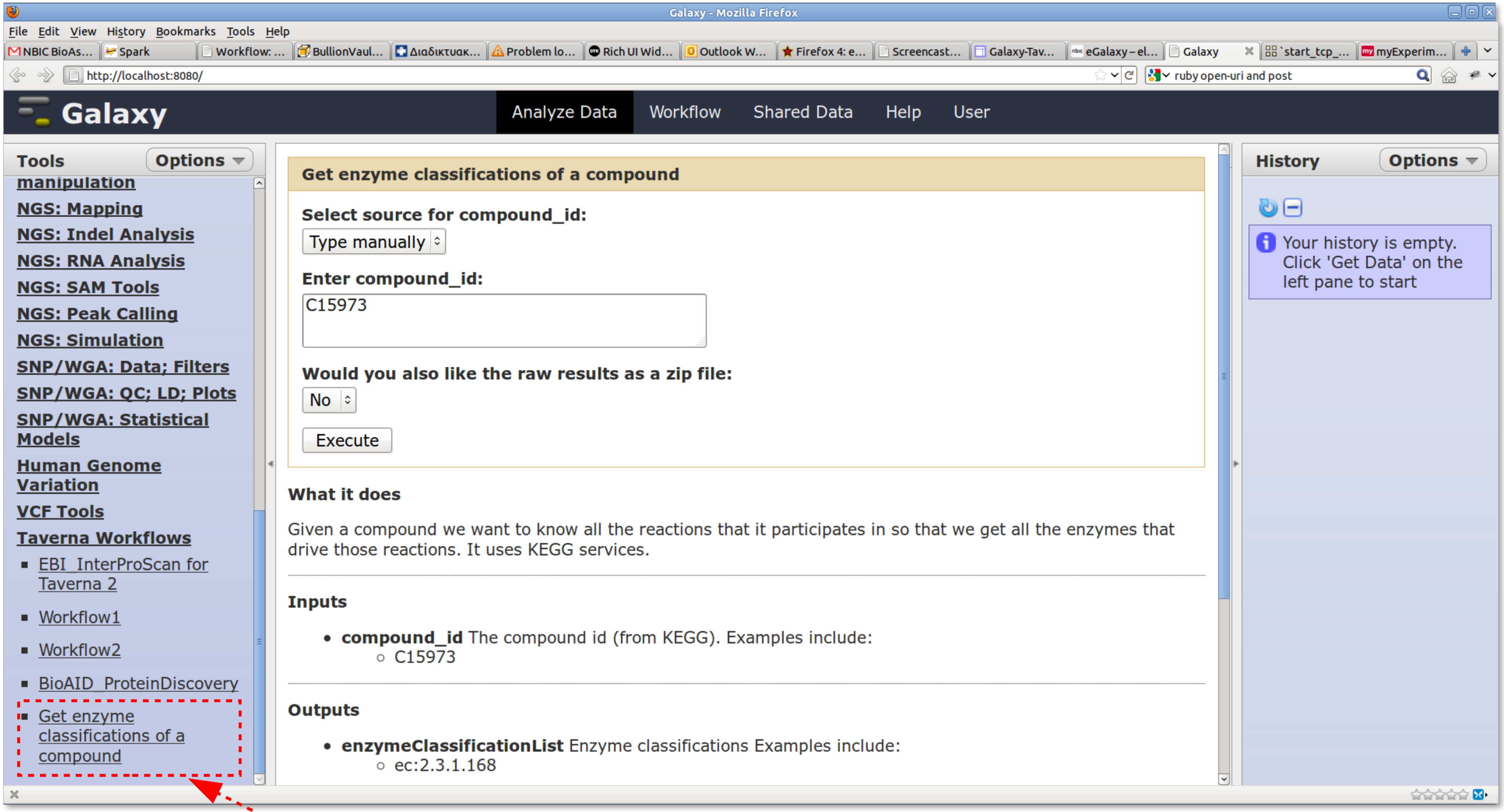
superscript
is needed

e.g. GAPSS pipeline

Multiple Scripts - Issues

- Difficult to manage
  - many scripts/programs
  - many arguments per script/program
  - many directories/files (scripts' I/Os)
- Difficult to reuse
  - due to low-level description
  - inconsistent program arguments

In Galaxy Web Portal (1)



Your history is empty. Click 'Get Data' on the left pane to start

Get enzyme classifications of a compound

Select source for compound_id: Type manually

Enter compound_id: C15973

Would you also like the raw results as a zip file: No

Execute

What it does

Given a compound we want to know all the reactions that it participates in so that we get all the enzymes that drive those reactions. It uses KEGG services.

Inputs

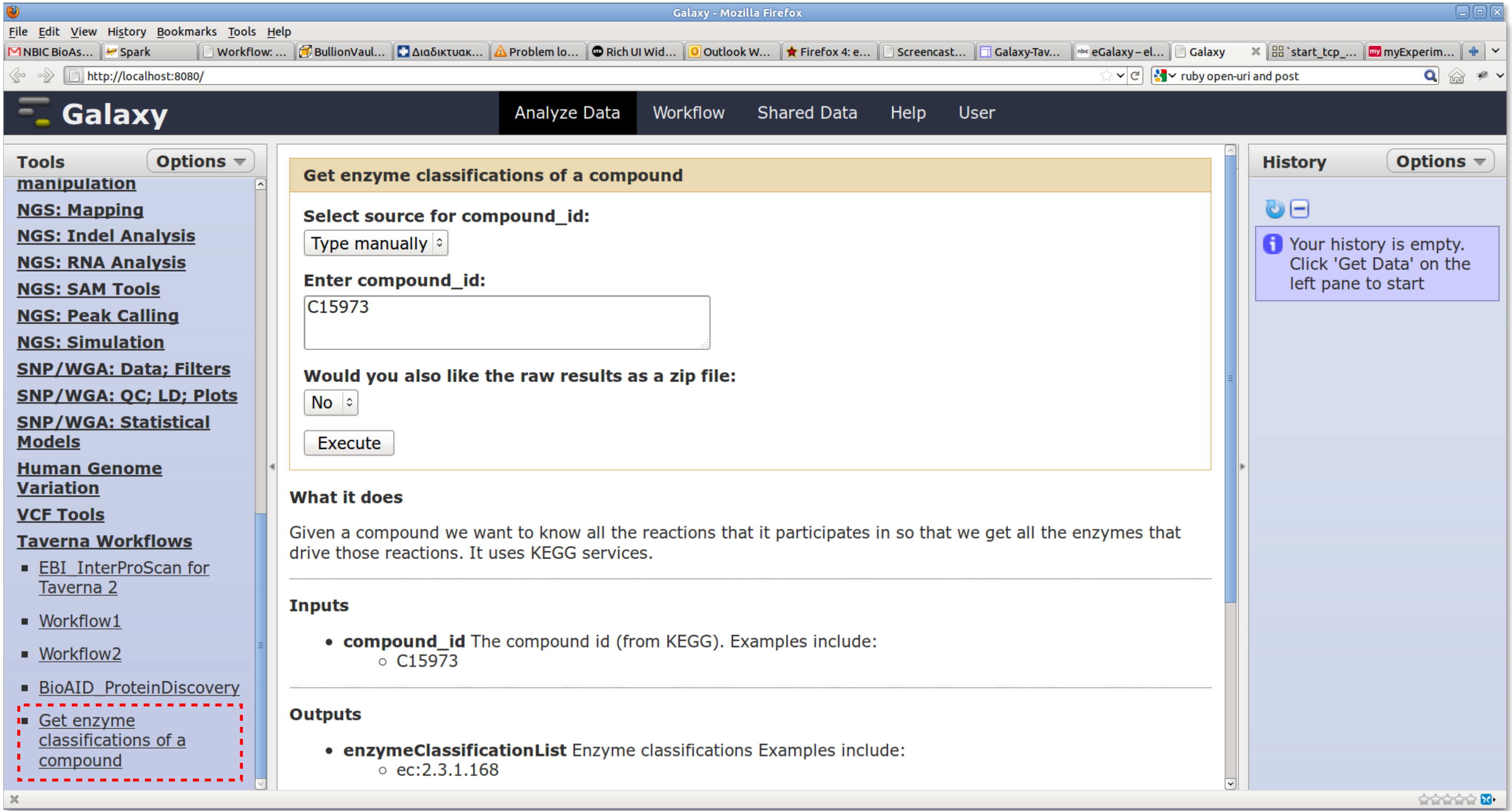
- **compound_id** The compound id (from KEGG). Examples include:
  - C15973

Outputs

- **enzymeClassificationList** Enzyme classifications Examples include:
  - ec:2.3.1.168

Tool

In Galaxy Web Portal (1)



Your history is empty. Click 'Get Data' on the left pane to start

Get enzyme classifications of a compound

Select source for compound_id: Type manually

Enter compound_id: C15973

Would you also like the raw results as a zip file: No

What it does

Given a compound we want to know all the reactions that it participates in so that we get all the enzymes that drive those reactions. It uses KEGG services.

Inputs

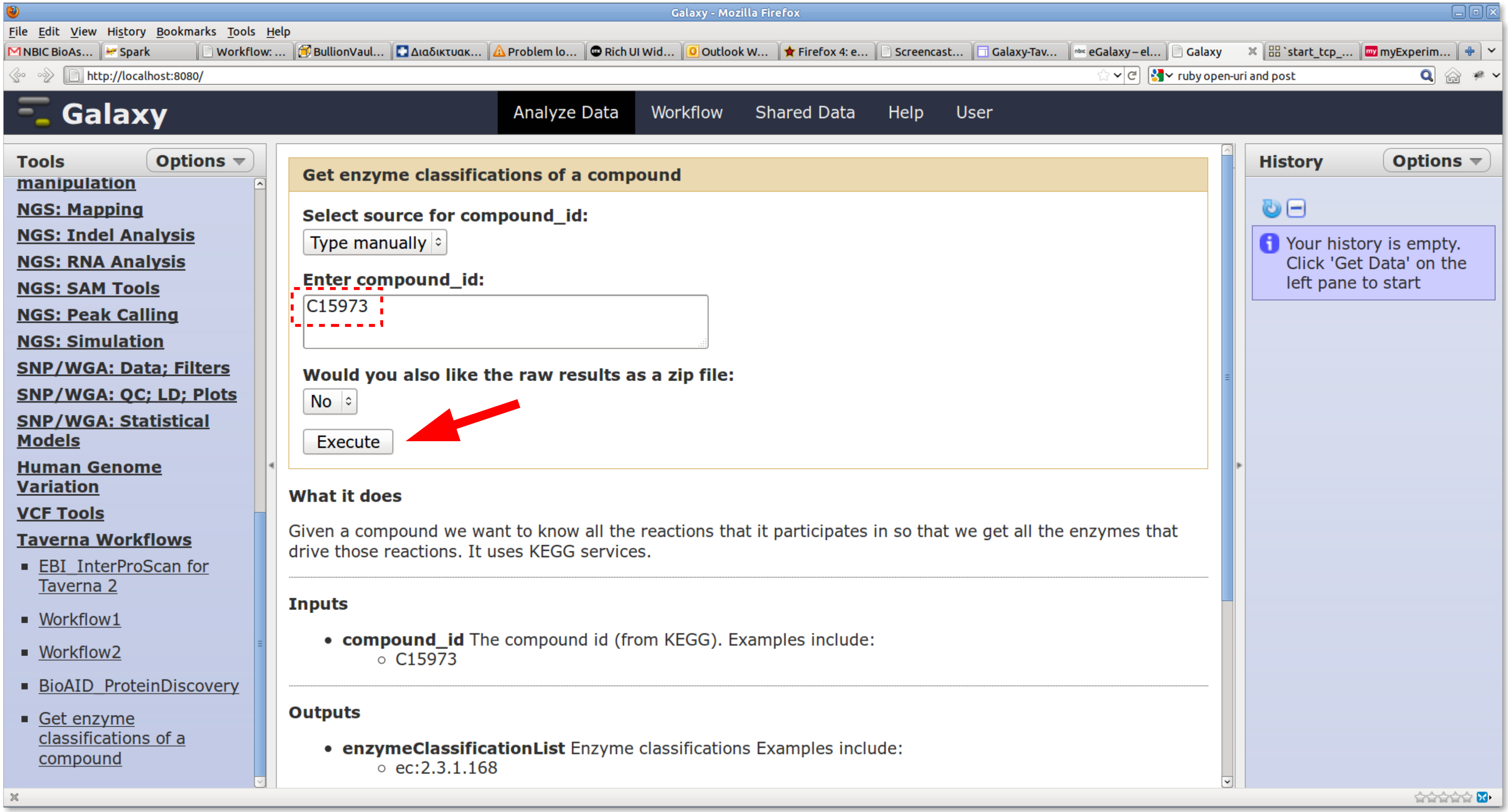
- **compound_id** The compound id (from KEGG). Examples include:
  - C15973

Outputs

- **enzymeClassificationList** Enzyme classifications Examples include:
  - ec:2.3.1.168

Who does what?

In Galaxy Web Portal (1)


Galaxy - Mozilla Firefox
File Edit View History Bookmarks Tools Help
NBIC BioAs… | Spark | Workflow: … | BullionVaul… | Διαδικτυακ… | Problem lo… | Rich UI Wid… | Outlook W… | Firefox 4: e… | Screencast… | Galaxy-Tav… | eGalaxy – el… | Galaxy | `start_tcp_… | myExperim…
http://localhost:8080/
ruby open-uri and post
Galaxy
Analyze Data
Workflow
Shared Data
Help
User
Tools
Options
History
Options

Tools

- manipulation
- NGS: Mapping
- NGS: Indel Analysis
- NGS: RNA Analysis
- NGS: SAM Tools
- NGS: Peak Calling
- NGS: Simulation
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- SNP/WGA: Statistical Models
- Human Genome Variation
- VCF Tools
- Taverna Workflows
  - EBI_InterProScan for Taverna 2
  - Workflow1
  - Workflow2
  - BioAID_ProteinDiscovery
  - Get enzyme classifications of a compound

Get enzyme classifications of a compound

Select source for compound_id: Type manually

Enter compound_id: C15973

Would you also like the raw results as a zip file: No

Execute

What it does

Given a compound we want to know all the reactions that it participates in so that we get all the enzymes that drive those reactions. It uses KEGG services.

Inputs

- **compound_id** The compound id (from KEGG). Examples include:
  - C15973

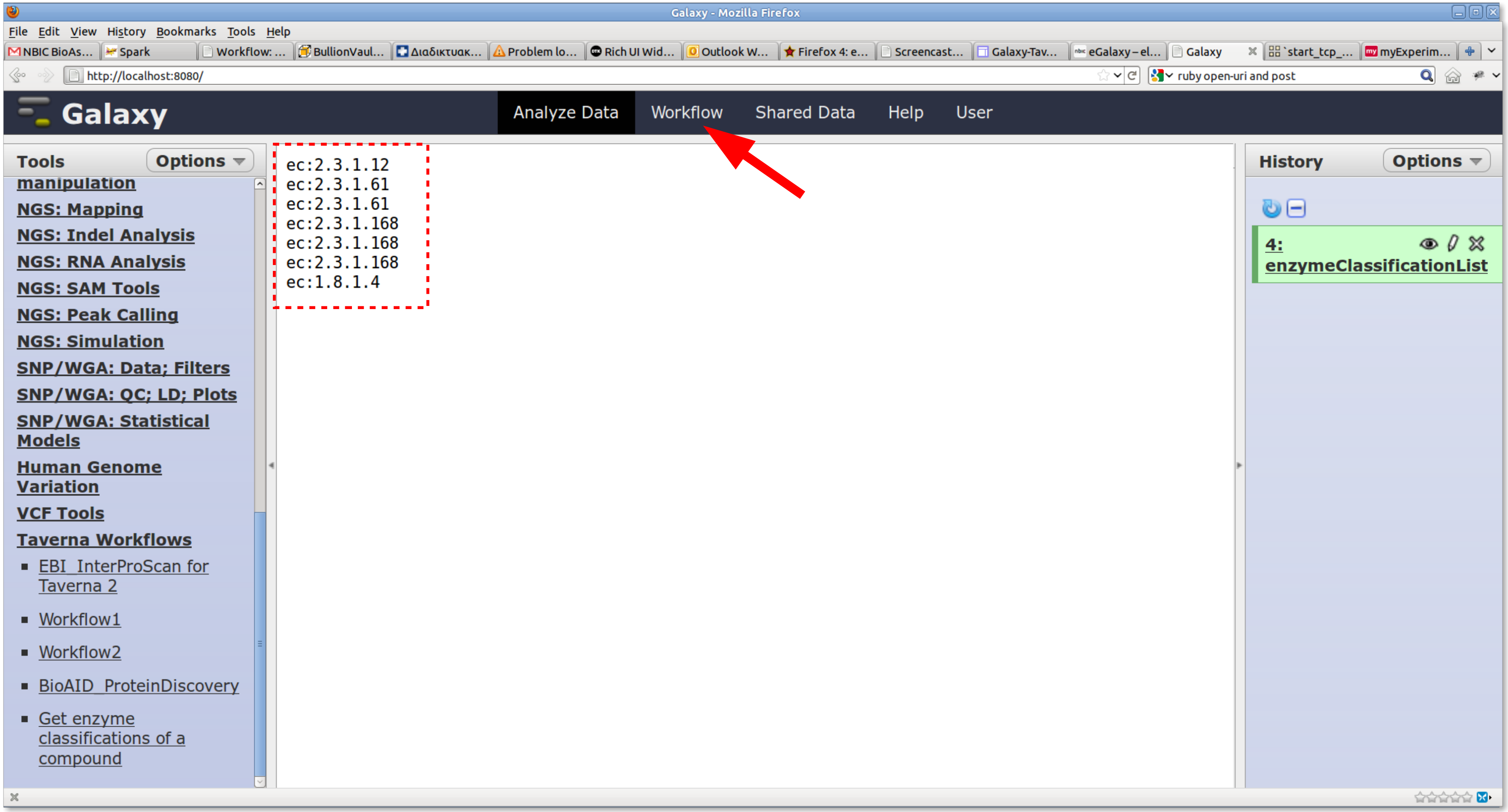
Outputs

- **enzymeClassificationList** Enzyme classifications Examples include:
  - ec:2.3.1.168

History

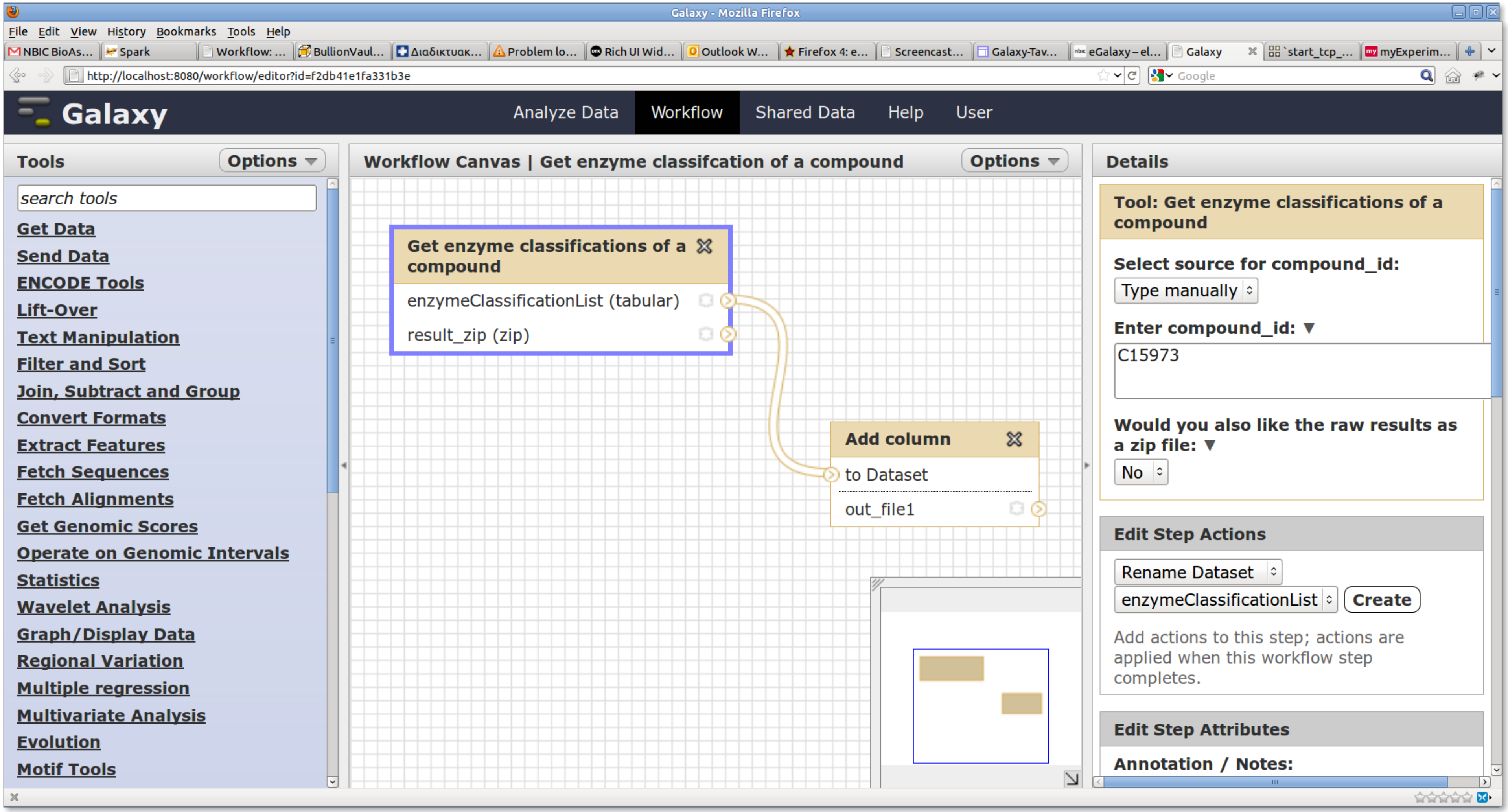
Your history is empty. Click 'Get Data' on the left pane to start

In Galaxy Web Portal (2)



```
ec:2.3.1.12
ec:2.3.1.61
ec:2.3.1.61
ec:2.3.1.168
ec:2.3.1.168
ec:2.3.1.168
ec:1.8.1.4
```

In Galaxy Web Portal (3)



http://localhost:8080/workflow/editor?id=f2db41e1fa331b3e

Workflow Canvas | Get enzyme classifcation of a compound

Get enzyme classifications of a compound
- enzymeClassificationList (tabular)
- result_zip (zip)

Add column
- to Dataset
- out_file1

Tool: Get enzyme classifications of a compound

Select source for compound_id: Type manually

Enter compound_id: ▼ C15973

Would you also like the raw results as a zip file: ▼ No

Edit Step Actions

Rename Dataset

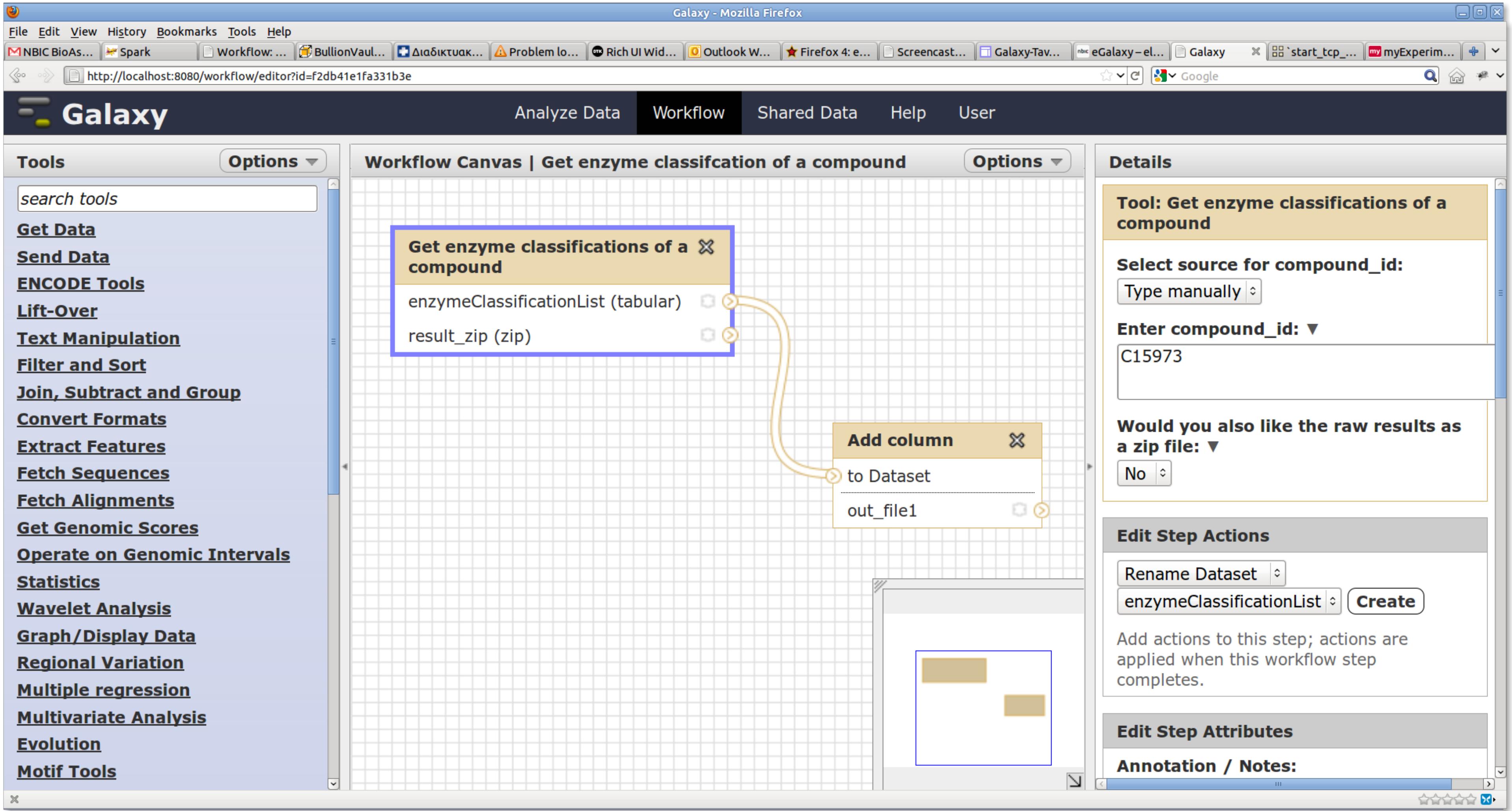
enzymeClassificationList

Add actions to this step; actions are applied when this workflow step completes.

Edit Step Attributes

Annotation / Notes:

In Galaxy Web Portal (3)



Tool: Get enzyme classifications of a compound

Select source for compound_id: Type manually

Enter compound_id: C15973

Would you also like the raw results as a zip file: No

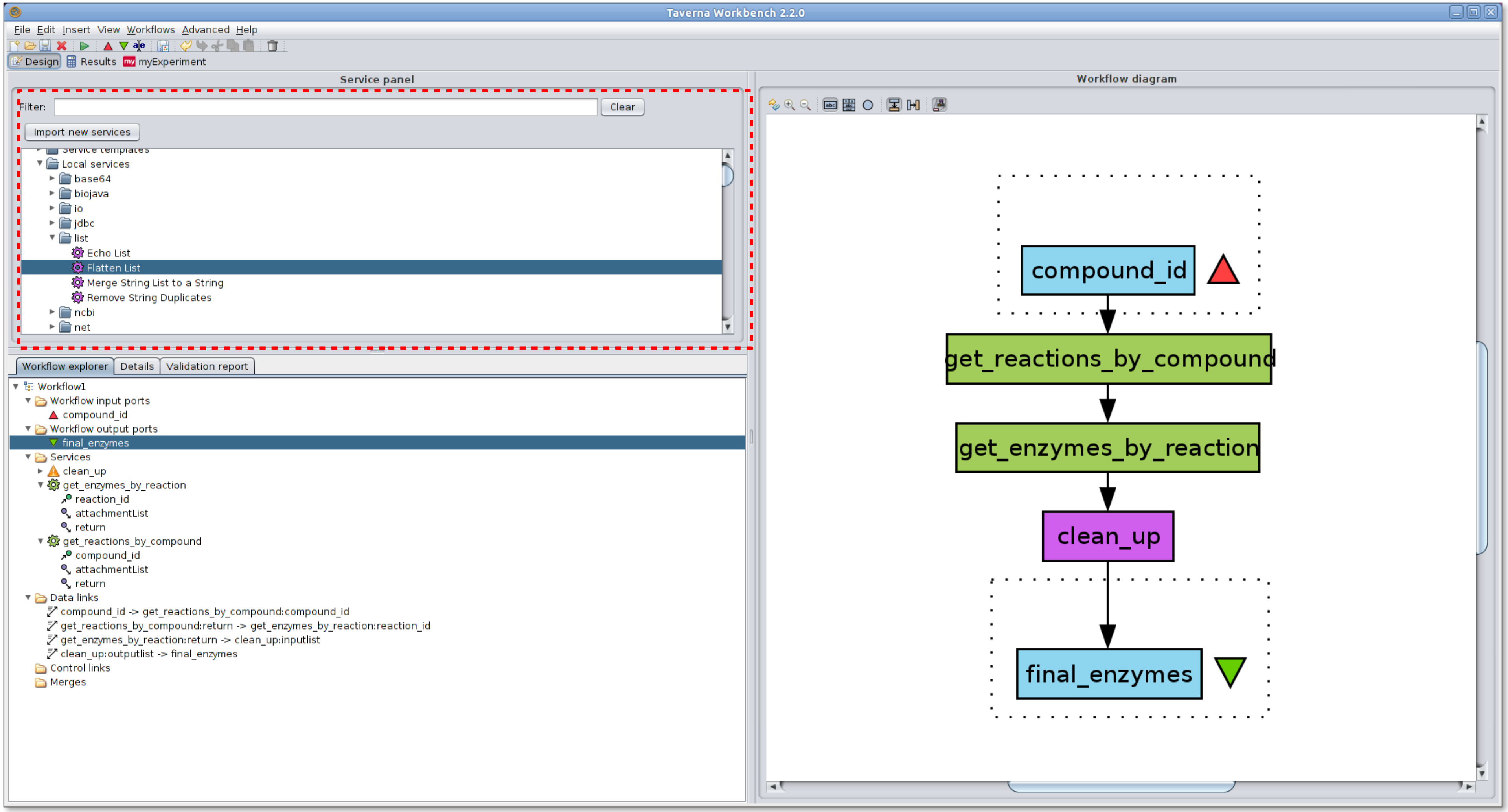
Add actions to this step; actions are applied when this workflow step completes.

e.g. GAPSS pipeline

Workflows Accessibility in Bioinformatics

Netherlands Bioinformatics Centre

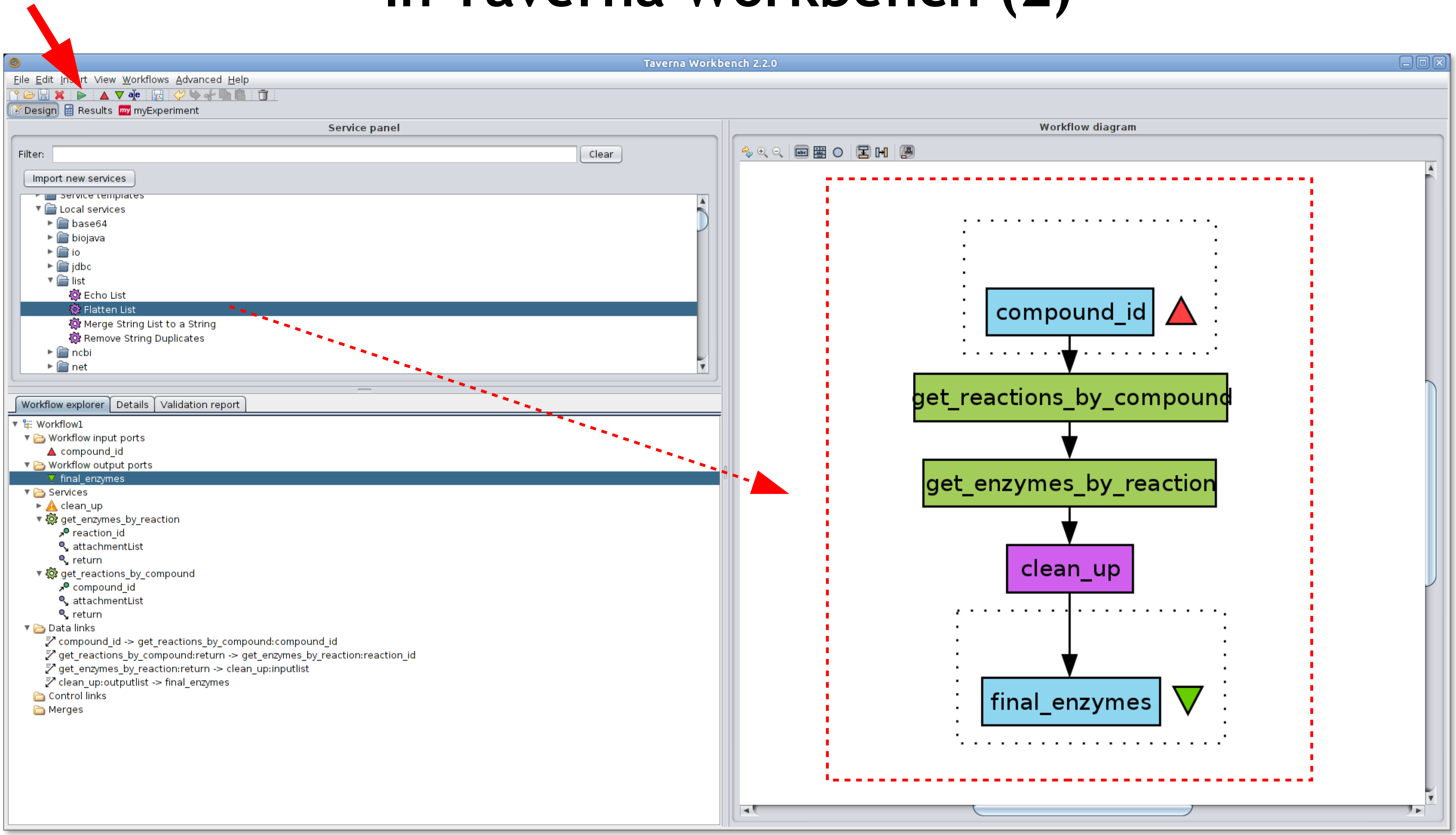
In Taverna Workbench (1)



```
▸ Service templates
▾ Local services
  ▸ base64
  ▸ biojava
  ▸ io
  ▸ jdbc
  ▾ list
      Echo List
      Flatten List
      Merge String List to a String
      Remove String Duplicates
  ▸ ncbi
  ▸ net
```

```
▾ Workflow1
  ▾ Workflow input ports
      compound_id
  ▾ Workflow output ports
      final_enzymes
  ▾ Services
    ▸ clean_up
    ▾ get_enzymes_by_reaction
        reaction_id
        attachmentList
        return
    ▾ get_reactions_by_compound
        compound_id
        attachmentList
        return
  ▾ Data links
      compound_id -> get_reactions_by_compound:compound_id
      get_reactions_by_compound:return -> get_enzymes_by_reaction:reaction_id
      get_enzymes_by_reaction:return -> clean_up:inputlist
      clean_up:outputlist -> final_enzymes
    Control links
    Merges
```

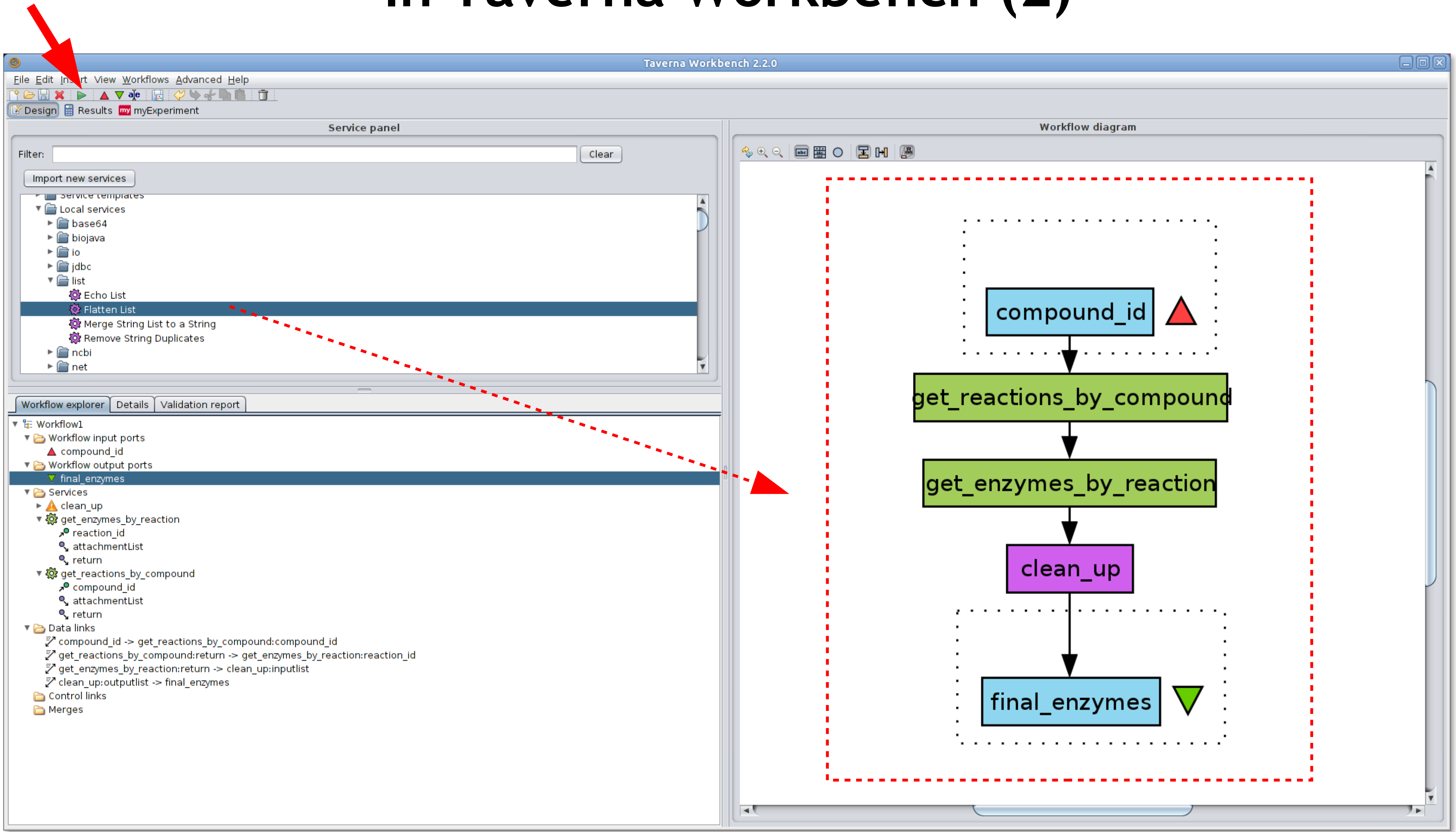
In Taverna Workbench (2)


Taverna Workbench 2.2.0
File Edit Insert View Workflows Advanced Help
Design Results myExperiment
Service panel
Filter:
Clear
Import new services
Workflow explorer Details Validation report
Workflow diagram

```
Service templates
Local services
  base64
  biojava
  io
  jdbc
  list
    Echo List
    Flatten List
    Merge String List to a String
    Remove String Duplicates
  ncbi
  net
```

```
Workflow1
  Workflow input ports
    compound_id
  Workflow output ports
    final_enzymes
  Services
    clean_up
    get_enzymes_by_reaction
      reaction_id
      attachmentList
      return
    get_reactions_by_compound
      compound_id
      attachmentList
      return
  Data links
    compound_id -> get_reactions_by_compound:compound_id
    get_reactions_by_compound:return -> get_enzymes_by_reaction:reaction_id
    get_enzymes_by_reaction:return -> clean_up:inputlist
    clean_up:outputlist -> final_enzymes
  Control links
  Merges
```

In Taverna Workbench (2)



- Service templates
- Local services
  - base64
  - biojava
  - io
  - jdbc
  - list
    - Echo List
    - Flatten List
    - Merge String List to a String
    - Remove String Duplicates
  - ncbi
  - net

- Workflow1
  - Workflow input ports
    - compound_id
  - Workflow output ports
    - final_enzymes
  - Services
    - clean_up
    - get_enzymes_by_reaction
      - reaction_id
      - attachmentList
      - return
    - get_reactions_by_compound
      - compound_id
      - attachmentList
      - return
  - Data links
    - compound_id -> get_reactions_by_compound:compound_id
    - get_reactions_by_compound:return -> get_enzymes_by_reaction:reaction_id
    - get_enzymes_by_reaction:return -> clean_up:inputlist
    - clean_up:outputlist -> final_enzymes
  - Control links
  - Merges

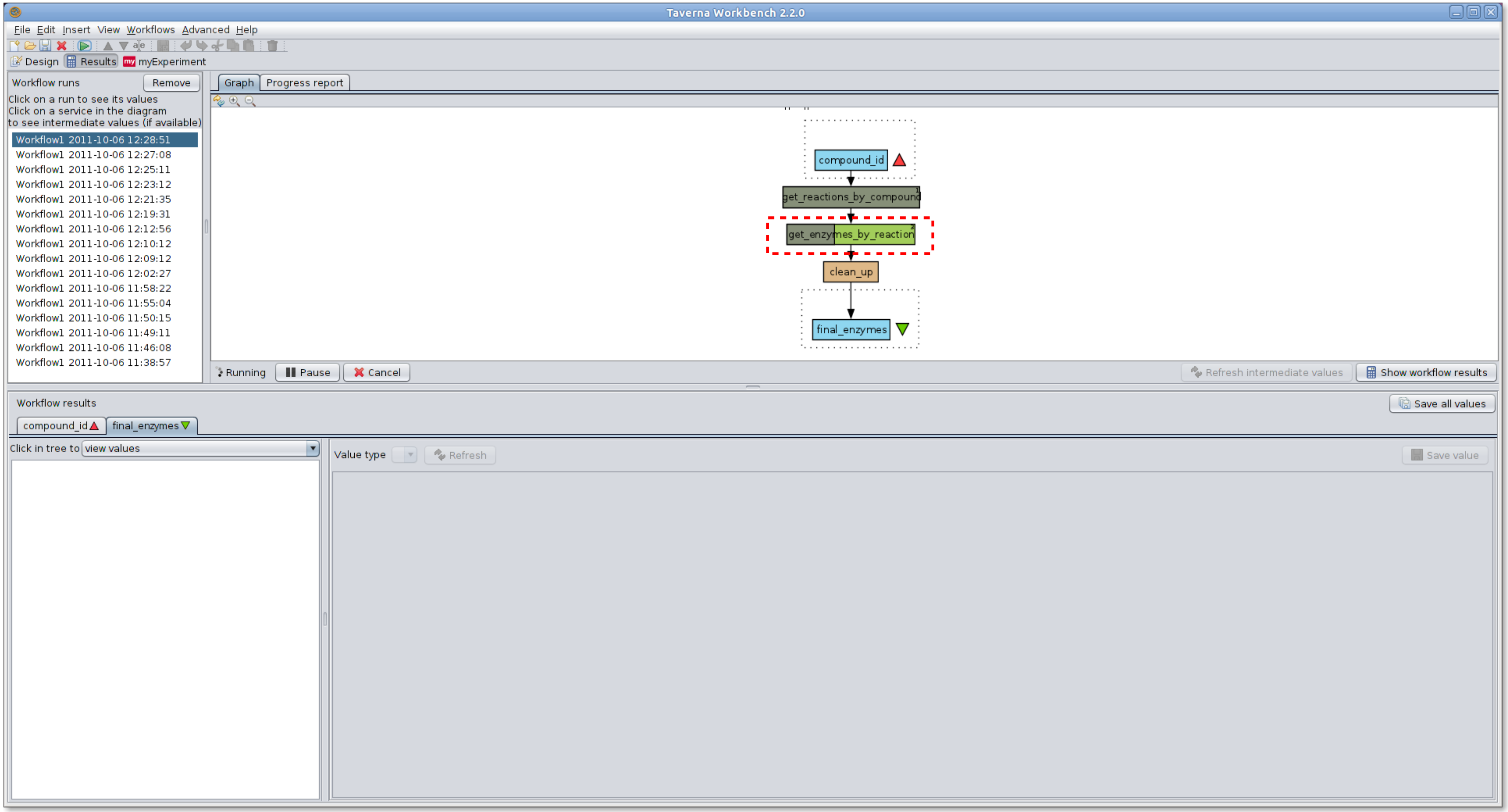
Who does what?

LUMC

Workflows Accessibility in Bioinformatics

nbic Netherlands Bioinformatics Centre

In Taverna Workbench (3)

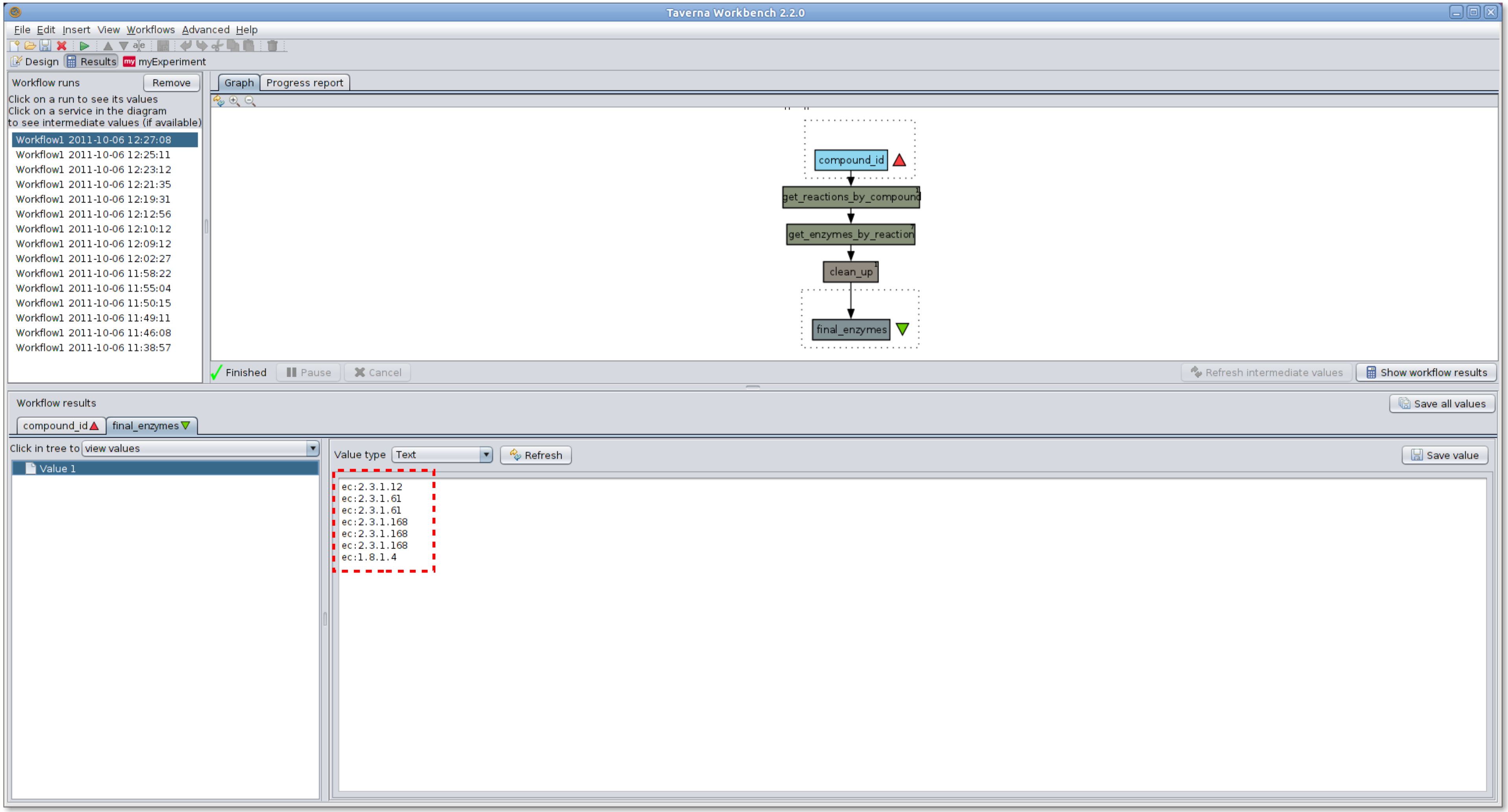


Click on a run to see its values
Click on a service in the diagram
to see intermediate values (if available)

Workflow1 2011-10-06 12:28:51
Workflow1 2011-10-06 12:27:08
Workflow1 2011-10-06 12:25:11
Workflow1 2011-10-06 12:23:12
Workflow1 2011-10-06 12:21:35
Workflow1 2011-10-06 12:19:31
Workflow1 2011-10-06 12:12:56
Workflow1 2011-10-06 12:10:12
Workflow1 2011-10-06 12:09:12
Workflow1 2011-10-06 12:02:27
Workflow1 2011-10-06 11:58:22
Workflow1 2011-10-06 11:55:04
Workflow1 2011-10-06 11:50:15
Workflow1 2011-10-06 11:49:11
Workflow1 2011-10-06 11:46:08
Workflow1 2011-10-06 11:38:57

Click in tree to view values

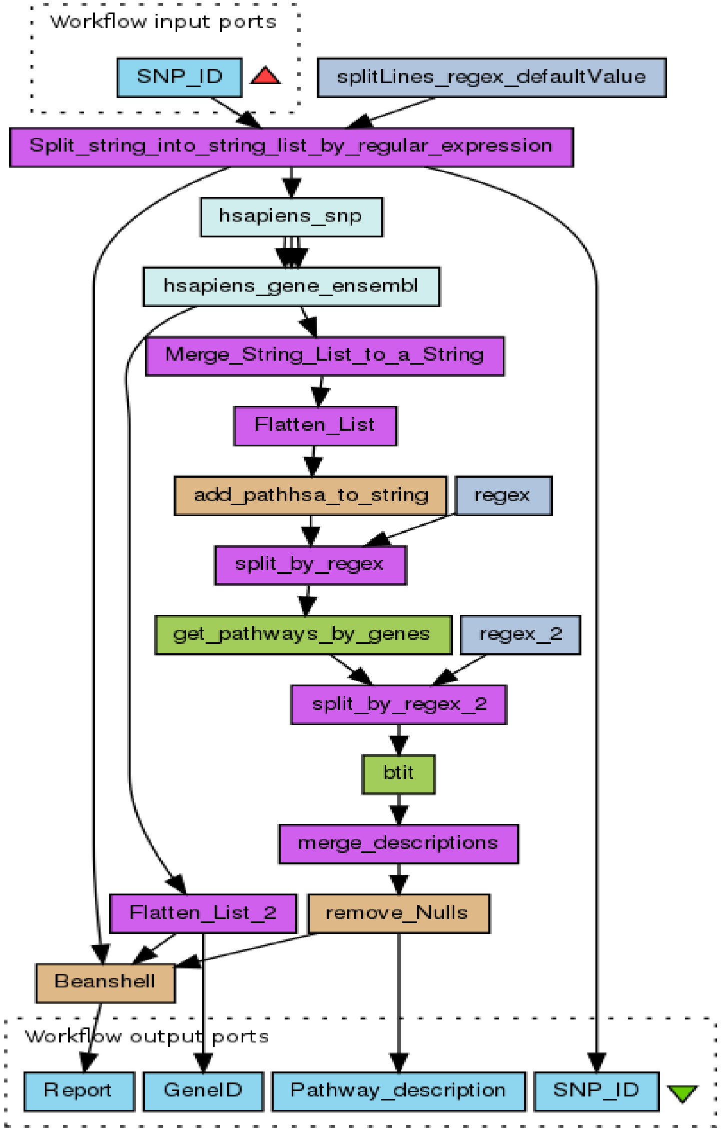
In Taverna Workbench (4)

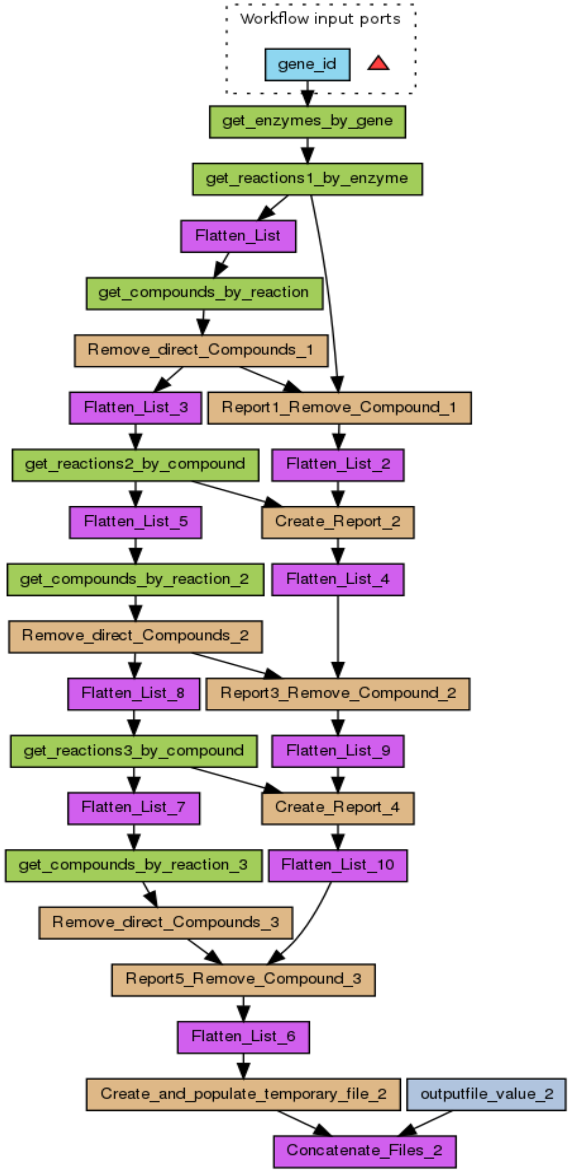

Taverna Workbench 2.2.0
File Edit Insert View Workflows Advanced Help
Design Results myExperiment
Workflow runs
Remove
Click on a run to see its values
Click on a service in the diagram
to see intermediate values (if available)
Graph
Progress report
compound_id
get_reactions_by_compound
get_enzymes_by_reaction
clean_up
final_enzymes
Finished
Pause
Cancel
Refresh intermediate values
Show workflow results
Workflow results
Save all values
compound_id
final_enzymes
Click in tree to view values
Value 1
Value type Text
Refresh
Save value

Workflow1 2011-10-06 12:27:08
Workflow1 2011-10-06 12:25:11
Workflow1 2011-10-06 12:23:12
Workflow1 2011-10-06 12:21:35
Workflow1 2011-10-06 12:19:31
Workflow1 2011-10-06 12:12:56
Workflow1 2011-10-06 12:10:12
Workflow1 2011-10-06 12:09:12
Workflow1 2011-10-06 12:02:27
Workflow1 2011-10-06 11:58:22
Workflow1 2011-10-06 11:55:04
Workflow1 2011-10-06 11:50:15
Workflow1 2011-10-06 11:49:11
Workflow1 2011-10-06 11:46:08
Workflow1 2011-10-06 11:38:57

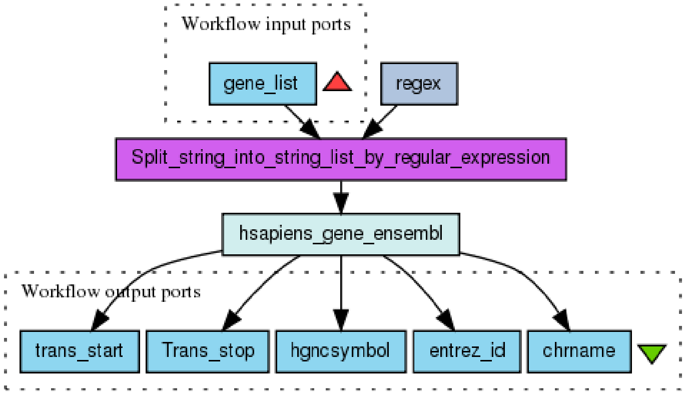
```
ec:2.3.1.12
ec:2.3.1.61
ec:2.3.1.61
ec:2.3.1.168
ec:2.3.1.168
ec:2.3.1.168
ec:1.8.1.4
```

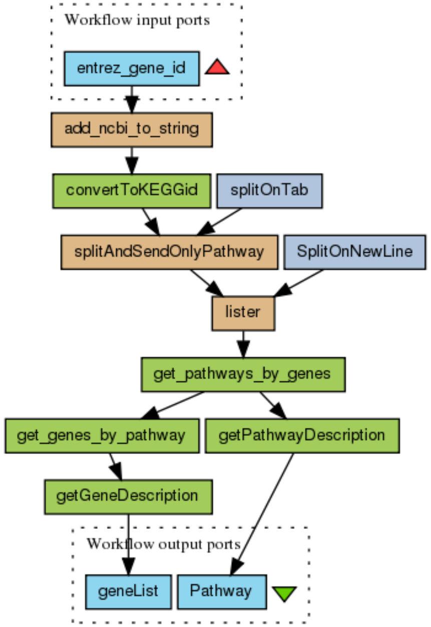
Harish's Workflows (GWAS: Metabolic Syndrome)

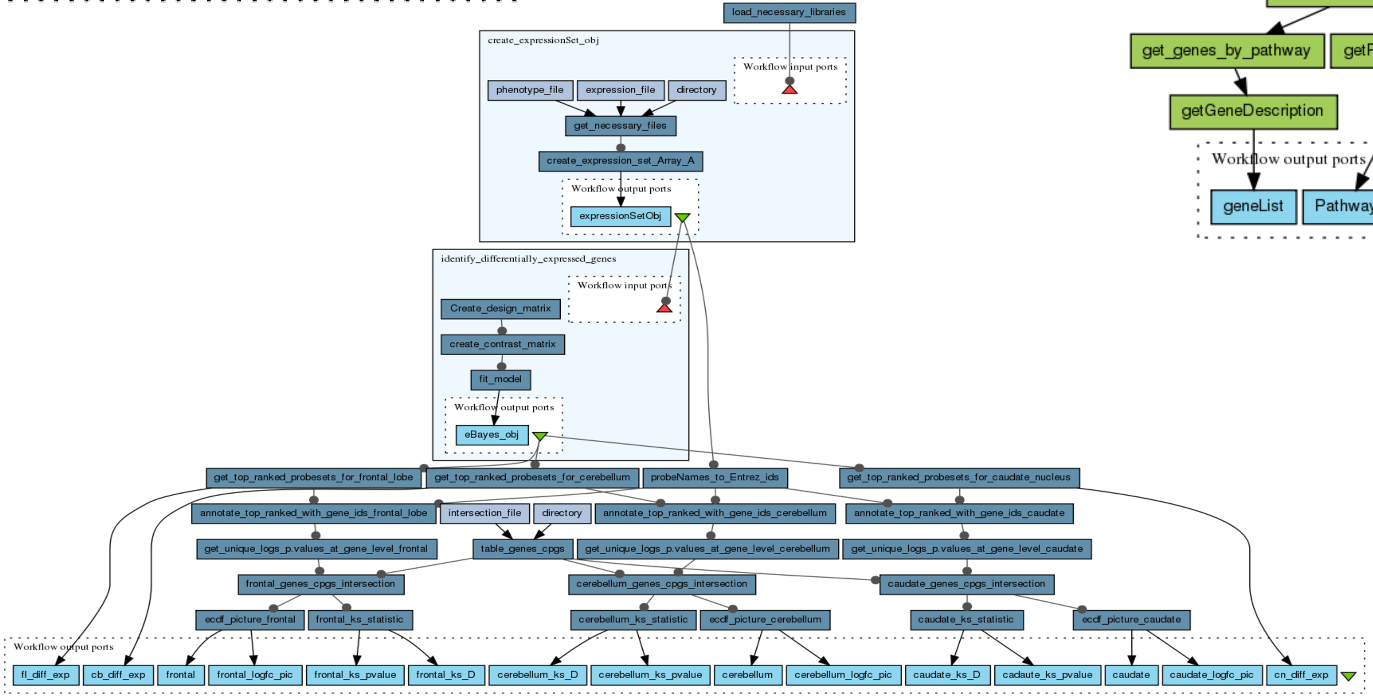



Workflow input ports
gene_id
get_enzymes_by_gene
get_reactions1_by_enzyme
Flatten_List
get_compounds_by_reaction
Remove_direct_Compounds_1
Flatten_List_3
Report1_Remove_Compound_1
get_reactions2_by_compound
Flatten_List_2
Flatten_List_5
Create_Report_2
get_compounds_by_reaction_2
Flatten_List_4
Remove_direct_Compounds_2
Flatten_List_8
Report3_Remove_Compound_2
get_reactions3_by_compound
Flatten_List_9
Flatten_List_7
Create_Report_4
get_compounds_by_reaction_3
Flatten_List_10
Remove_direct_Compounds_3
Report5_Remove_Compound_3
Flatten_List_6
Create_and_populate_temporary_file_2
outputfile_value_2
Concatenate_Files_2

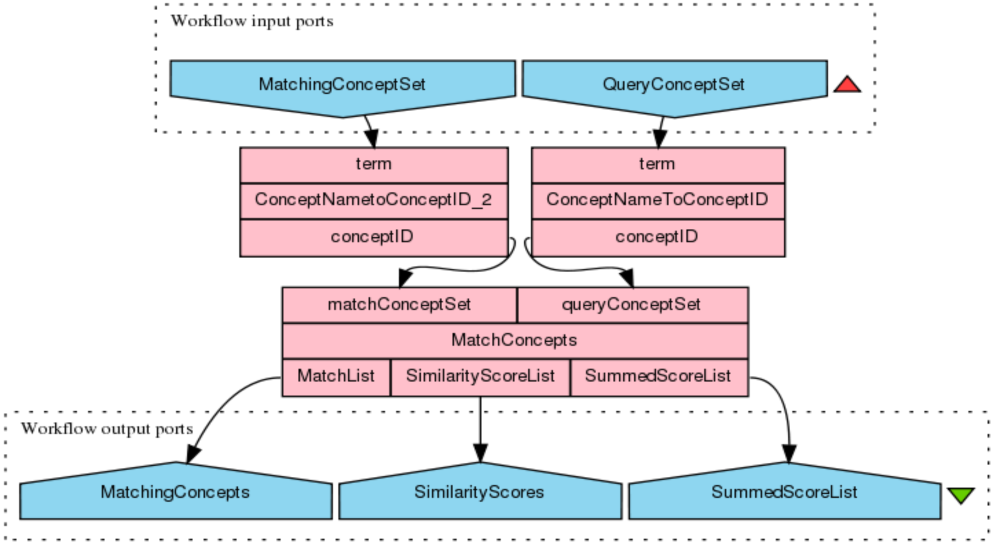
Eleni's Workflows (Huntington's Disease)

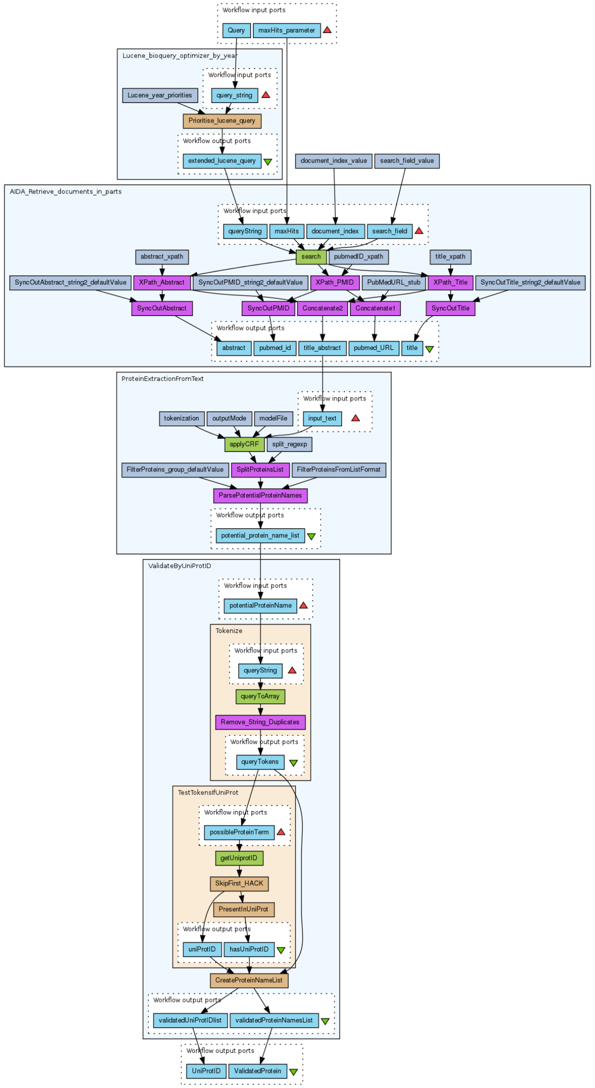






Marco's workflows (Semantic Text Mining)





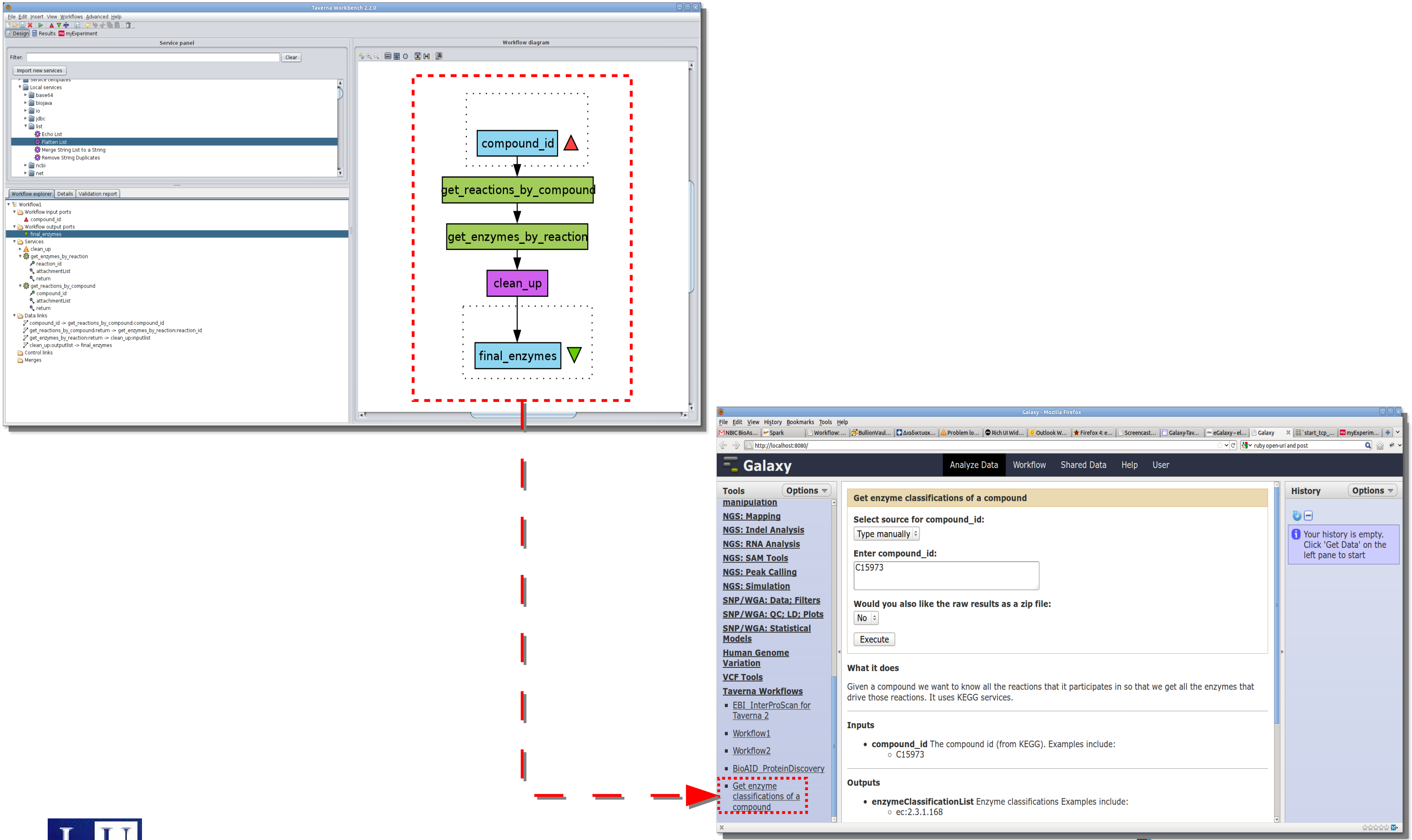
Taverna and Galaxy workflows

- The systems have a different focus
  - Some overlapping functionality but different strengths
  - Different fan clubs!

Galaxy	Taverna
straightforward workflows	very expressive workflows
exposing existing scripts	exposing existing web services
typically local tools	typically remote tools

- Do we have to make a choice?
- Does that limit our potential users?
- How to make them more interoperable?

Taverna Workflows in Galaxy (1)



Service templates
- Local services
  - base64
  - biojava
  - io
  - jdbc
  - list
    - Echo List
    - Flatten List
    - Merge String List to a String
    - Remove String Duplicates
  - ncbi
  - net

Workflow explorer | Details | Validation report

- Workflow1
  - Workflow input ports
    - compound_id
  - Workflow output ports
    - final_enzymes
  - Services
    - clean_up
    - get_enzymes_by_reaction
      - reaction_id
      - attachmentList
      - return
    - get_reactions_by_compound
      - compound_id
      - attachmentList
      - return
  - Data links
    - compound_id -> get_reactions_by_compound:compound_id
    - get_reactions_by_compound:return -> get_enzymes_by_reaction:reaction_id
    - get_enzymes_by_reaction:return -> clean_up:inputlist
    - clean_up:outputlist -> final_enzymes
  - Control links
  - Merges

manipulation

NGS: Mapping

NGS: Indel Analysis

NGS: RNA Analysis

NGS: SAM Tools

NGS: Peak Calling

NGS: Simulation

SNP/WGA: Data; Filters

SNP/WGA: QC; LD; Plots

SNP/WGA: Statistical Models

Human Genome Variation

VCF Tools

Taverna Workflows

- EBI_InterProScan for Taverna 2
- Workflow1
- Workflow2
- BioAID_ProteinDiscovery
- Get enzyme classifications of a compound

Get enzyme classifications of a compound

Select source for compound_id: Type manually

Enter compound_id: C15973

Would you also like the raw results as a zip file: No

Execute

What it does

Given a compound we want to know all the reactions that it participates in so that we get all the enzymes that drive those reactions. It uses KEGG services.

Inputs

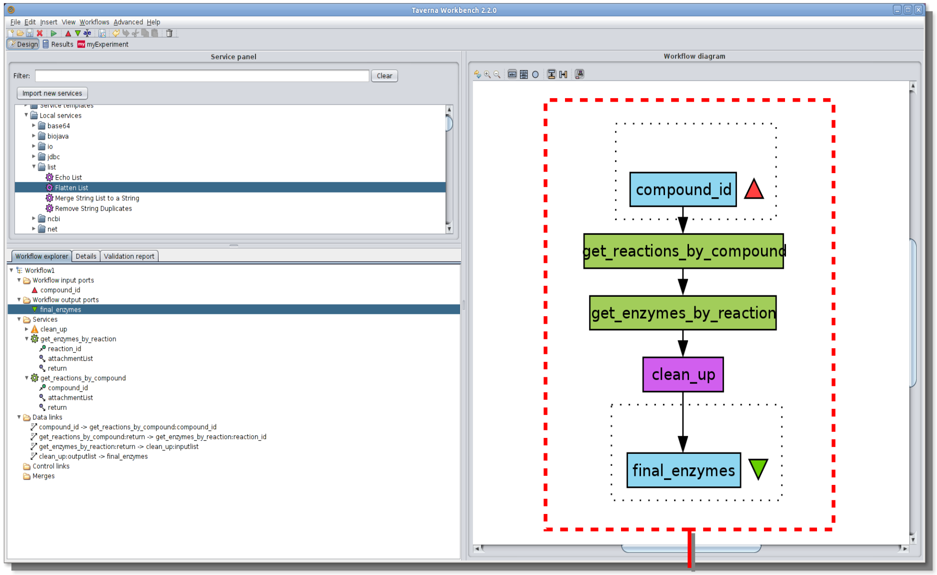
- **compound_id** The compound id (from KEGG). Examples include:
  - C15973

Outputs

- **enzymeClassificationList** Enzyme classifications Examples include:
  - ec:2.3.1.168

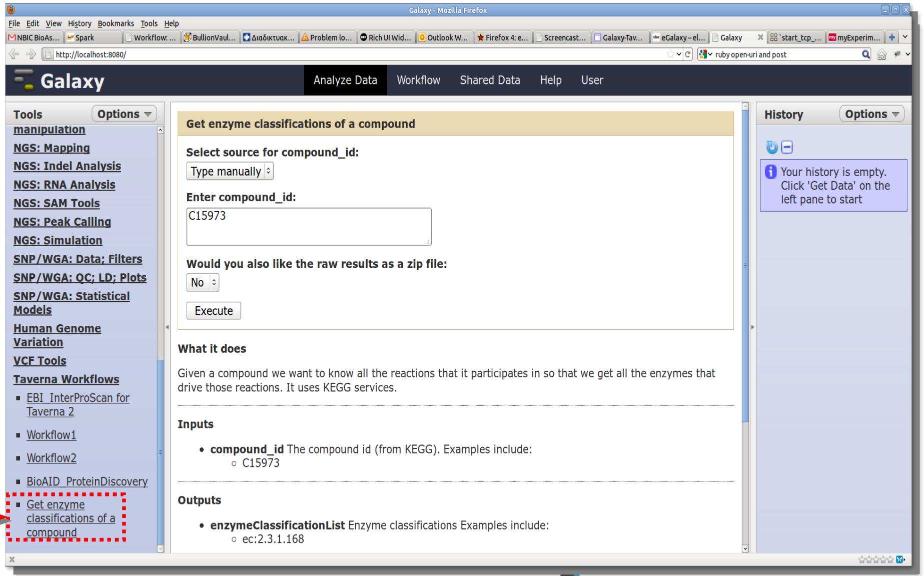
Your history is empty. Click 'Get Data' on the left pane to start

Taverna Workflows in Galaxy (1)


Taverna Workbench 2.2.0
File Edit Insert View Workflows Advanced Help
Design Results myExperiment
Service panel
Filter:
Clear
Import new services
Workflow diagram
Workflow explorer Details Validation report
compound_id
get_reactions_by_compound
get_enzymes_by_reaction
clean_up
final_enzymes

- Service templates
- Local services
  - base64
  - biojava
  - io
  - jdbc
  - list
    - Echo List
    - Flatten List
    - Merge String List to a String
    - Remove String Duplicates
  - ncbi
  - net

- Workflow1
  - Workflow input ports
    - compound_id
  - Workflow output ports
    - final_enzymes
  - Services
    - clean_up
    - get_enzymes_by_reaction
      - reaction_id
      - attachmentList
      - return
    - get_reactions_by_compound
      - compound_id
      - attachmentList
      - return
  - Data links
    - compound_id -> get_reactions_by_compound:compound_id
    - get_reactions_by_compound:return -> get_enzymes_by_reaction:reaction_id
    - get_enzymes_by_reaction:return -> clean_up:inputlist
    - clean_up:outputlist -> final_enzymes
  - Control links
  - Merges


Galaxy - Mozilla Firefox
File Edit View History Bookmarks Tools Help
NBIC BioAs... Spark Workflow: ... BullionVaul... Διαδικτυακ... Problem lo... Rich UI Wid... Outlook W... Firefox 4: e... Screencast... Galaxy-Tav... eGalaxy – el... Galaxy `start_tcp_... myExperim...
http://localhost:8080/
ruby open-uri and post
Galaxy
Analyze Data Workflow Shared Data Help User
Tools Options
History Options

manipulation
NGS: Mapping
NGS: Indel Analysis
NGS: RNA Analysis
NGS: SAM Tools
NGS: Peak Calling
NGS: Simulation
SNP/WGA: Data; Filters
SNP/WGA: QC; LD; Plots
SNP/WGA: Statistical Models
Human Genome Variation
VCF Tools
Taverna Workflows

- EBI_InterProScan for Taverna 2
- Workflow1
- Workflow2
- BioAID_ProteinDiscovery
- Get enzyme classifications of a compound

Get enzyme classifications of a compound

Select source for compound_id: Type manually

Enter compound_id: C15973

Would you also like the raw results as a zip file: No

Execute

What it does

Given a compound we want to know all the reactions that it participates in so that we get all the enzymes that drive those reactions. It uses KEGG services.

Inputs

- **compound_id** The compound id (from KEGG). Examples include:
  - C15973

Outputs

- **enzymeClassificationList** Enzyme classifications Examples include:
  - ec:2.3.1.168

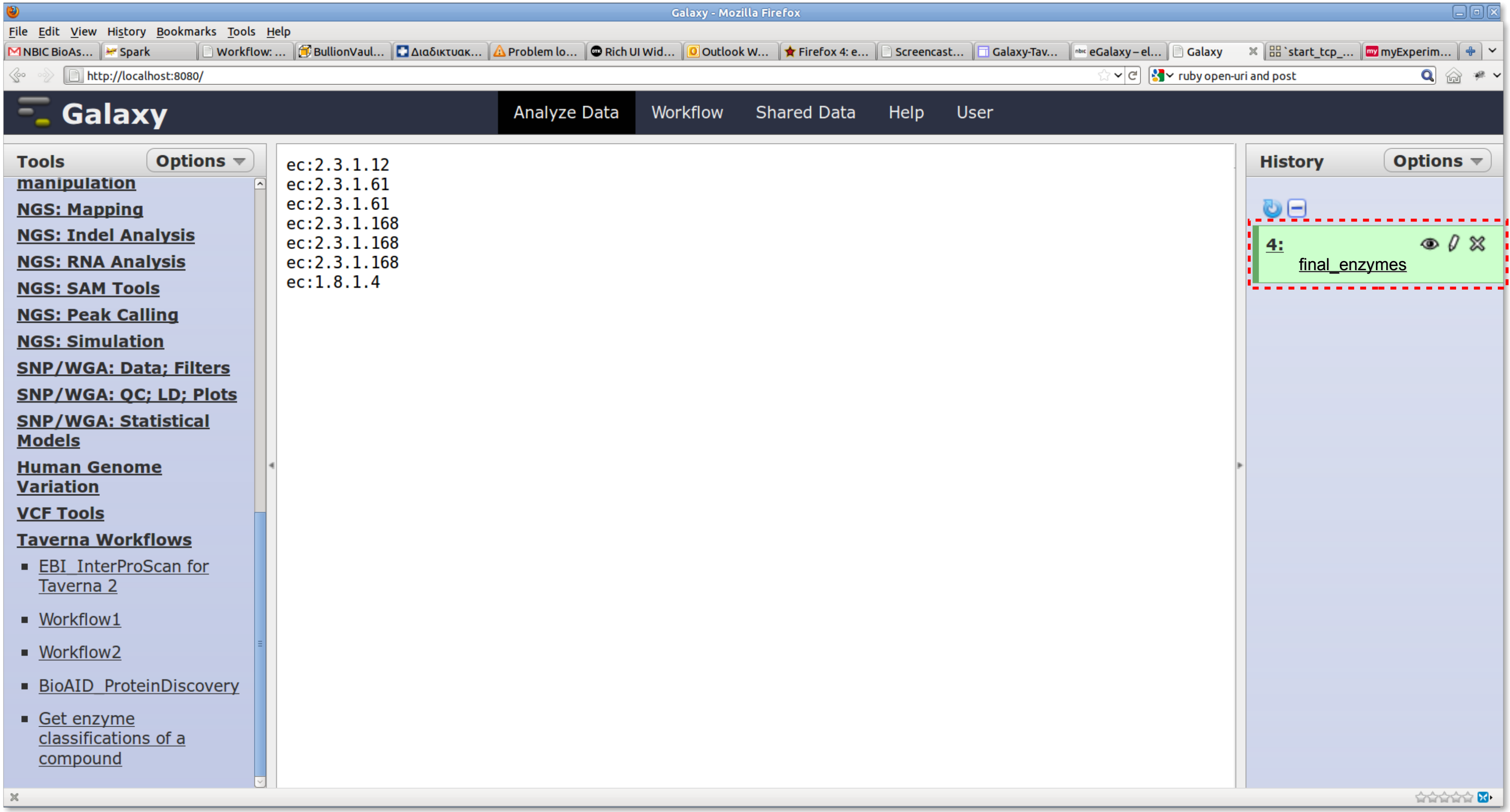
Your history is empty. Click 'Get Data' on the left pane to start

Who does what?

Workflows Accessibility in Bioinformatics

Netherlands Bioinformatics Centre

Taverna Workflows in Galaxy (2)



Galaxy - Mozilla Firefox
File Edit View History Bookmarks Tools Help
NBIC BioAs... Spark Workflow: ... BullionVaul... Διαδικτυακ... Problem lo... Rich UI Wid... Outlook W... Firefox 4: e... Screencast... Galaxy-Tav... eGalaxy – el... Galaxy `start_tcp_... myExperim...
http://localhost:8080/
ruby open-uri and post
Galaxy
Analyze Data
Workflow
Shared Data
Help
User
Tools
Options
History
Options

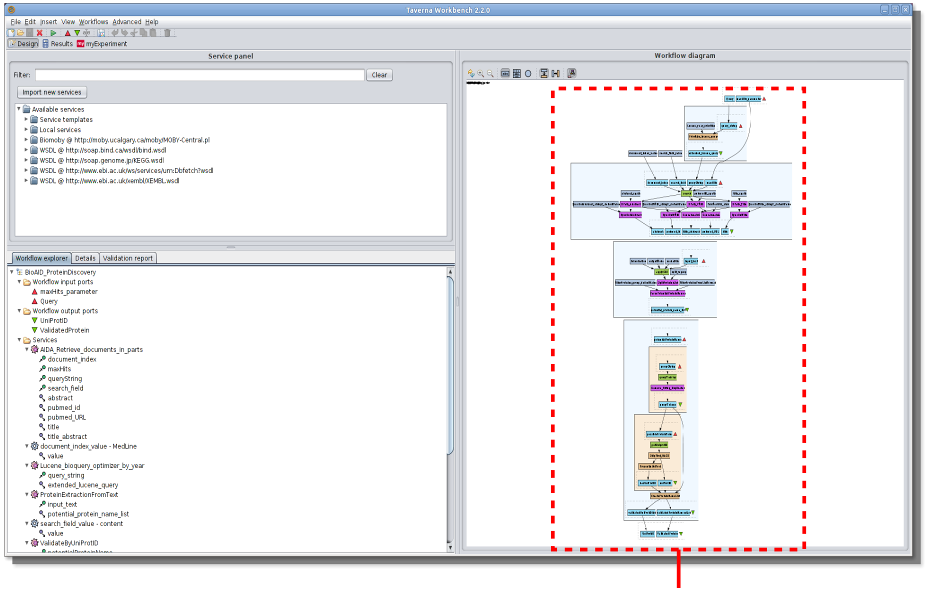
manipulation
NGS: Mapping
NGS: Indel Analysis
NGS: RNA Analysis
NGS: SAM Tools
NGS: Peak Calling
NGS: Simulation
SNP/WGA: Data; Filters
SNP/WGA: QC; LD; Plots
SNP/WGA: Statistical Models
Human Genome Variation
VCF Tools
Taverna Workflows

- EBI_InterProScan for Taverna 2
- Workflow1
- Workflow2
- BioAID_ProteinDiscovery
- Get enzyme classifications of a compound

```
ec:2.3.1.12
ec:2.3.1.61
ec:2.3.1.61
ec:2.3.1.168
ec:2.3.1.168
ec:2.3.1.168
ec:1.8.1.4
```

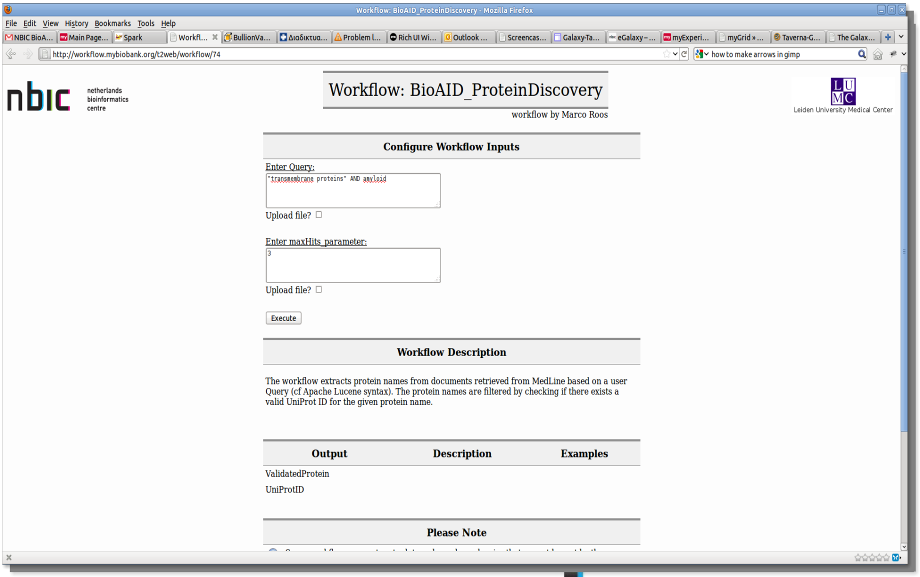
4: final_enzymes

Taverna Workflows on a Web Browser (1)



- Available services
  - Service templates
  - Local services
  - Biomoby @ http://moby.ucalgary.ca/moby/MOBY-Central.pl
  - WSDL @ http://soap.bind.ca/wsdl/bind.wsdl
  - WSDL @ http://soap.genome.jp/KEGG.wsdl
  - WSDL @ http://www.ebi.ac.uk/ws/services/urn:Dbfetch?wsdl
  - WSDL @ http://www.ebi.ac.uk/xembl/XEMBL.wsdl

- BioAID_ProteinDiscovery
  - Workflow input ports
    - maxHits_parameter
    - Query
  - Workflow output ports
    - UniProtID
    - ValidatedProtein
  - Services
    - AIDA_Retrieve_documents_in_parts
      - document_index
      - maxHits
      - queryString
      - search_field
      - abstract
      - pubmed_id
      - pubmed_URL
      - title
      - title_abstract
    - document_index_value - MedLine
      - value
    - Lucene_bioquery_optimizer_by_year
      - query_string
      - extended_lucene_query
    - ProteinExtractionFromText
      - input_text
      - potential_protein_name_list
    - search_field_value - content
      - value
    - ValidateByUniProtID



Workflow: BioAID_ProteinDiscovery - Mozilla Firefox
File Edit View History Bookmarks Tools Help
NBIC BioA...
Main Page...
Spark
Workfl...
BullionVa...
Διαδικτυα...
Problem l...
Rich UI Wi...
Outlook ...
Screencas...
Galaxy-Ta...
eGalaxy – ...
myExperi...
myGrid » ...
Taverna-G...
The Galax...
http://workflow.mybiobank.org/t2web/workflow/74
how to make arrows in gimp
nbic netherlands bioinformatics centre
Leiden University Medical Center

Workflow: BioAID_ProteinDiscovery

workflow by Marco Roos

Configure Workflow Inputs

Enter Query:
"transmembrane proteins" AND amyloid

Upload file? ☐

Enter maxHits_parameter:
3

Upload file? ☐

Execute

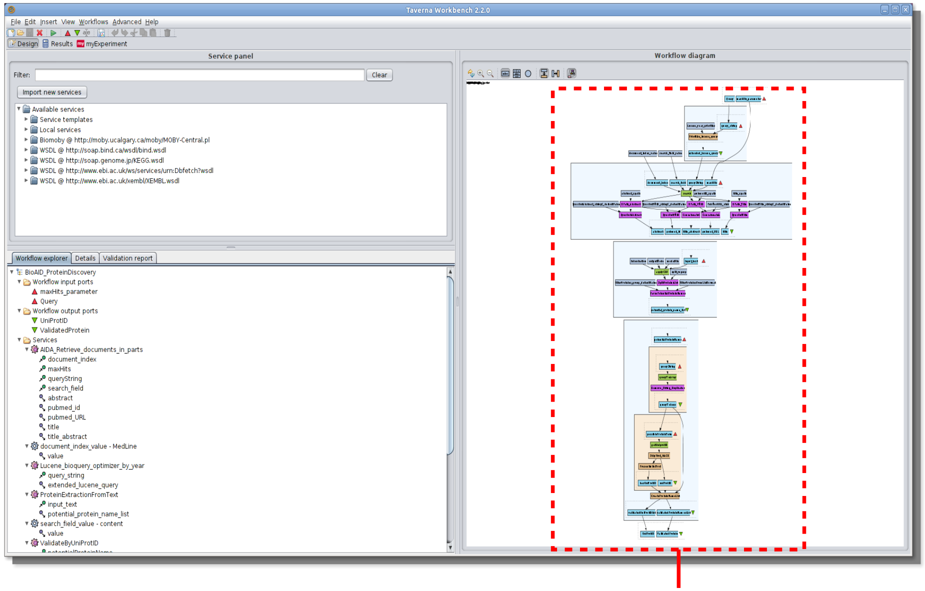
Workflow Description

The workflow extracts protein names from documents retrieved from MedLine based on a user Query (cf Apache Lucene syntax). The protein names are filtered by checking if there exists a valid UniProt ID for the given protein name.

Output	Description	Examples
ValidatedProtein		
UniProtID		

Please Note

Taverna Workflows on a Web Browser (1)



Taverna Workbench 2.2.0

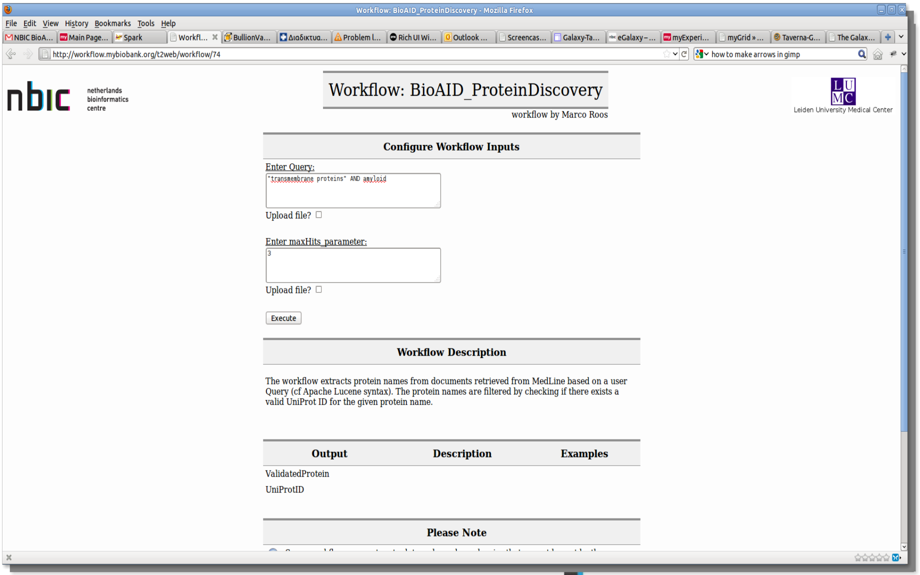
Service panel

Filter:

- Available services
  - Service templates
  - Local services
  - Biomoby @ http://moby.ucalgary.ca/moby/MOBY-Central.pl
  - WSDL @ http://soap.bind.ca/wsdl/bind.wsdl
  - WSDL @ http://soap.genome.jp/KEGG.wsdl
  - WSDL @ http://www.ebi.ac.uk/ws/services/urn:Dbfetch?wsdl
  - WSDL @ http://www.ebi.ac.uk/xembl/XEMBL.wsdl

Workflow explorer | Details | Validation report

- BioAID_ProteinDiscovery
  - Workflow input ports
    - maxHits_parameter
    - Query
  - Workflow output ports
    - UniProtID
    - ValidatedProtein
  - Services
    - AIDA_Retrieve_documents_in_parts
      - document_index
      - maxHits
      - queryString
      - search_field
      - abstract
      - pubmed_id
      - pubmed_URL
      - title
      - title_abstract
    - document_index_value - MedLine
      - value
    - Lucene_bioquery_optimizer_by_year
      - query_string
      - extended_lucene_query
    - ProteinExtractionFromText
      - input_text
      - potential_protein_name_list
    - search_field_value - content
      - value
    - ValidateByUniProtID



Workflow: BioAID_ProteinDiscovery - Mozilla Firefox

http://workflow.mybiobank.org/t2web/workflow/74

Workflow: BioAID_ProteinDiscovery

workflow by Marco Roos

Configure Workflow Inputs

Enter Query:

"transmembrane proteins" AND amyloid

Upload file? □

Enter maxHits_parameter:

3

Upload file? □

Workflow Description

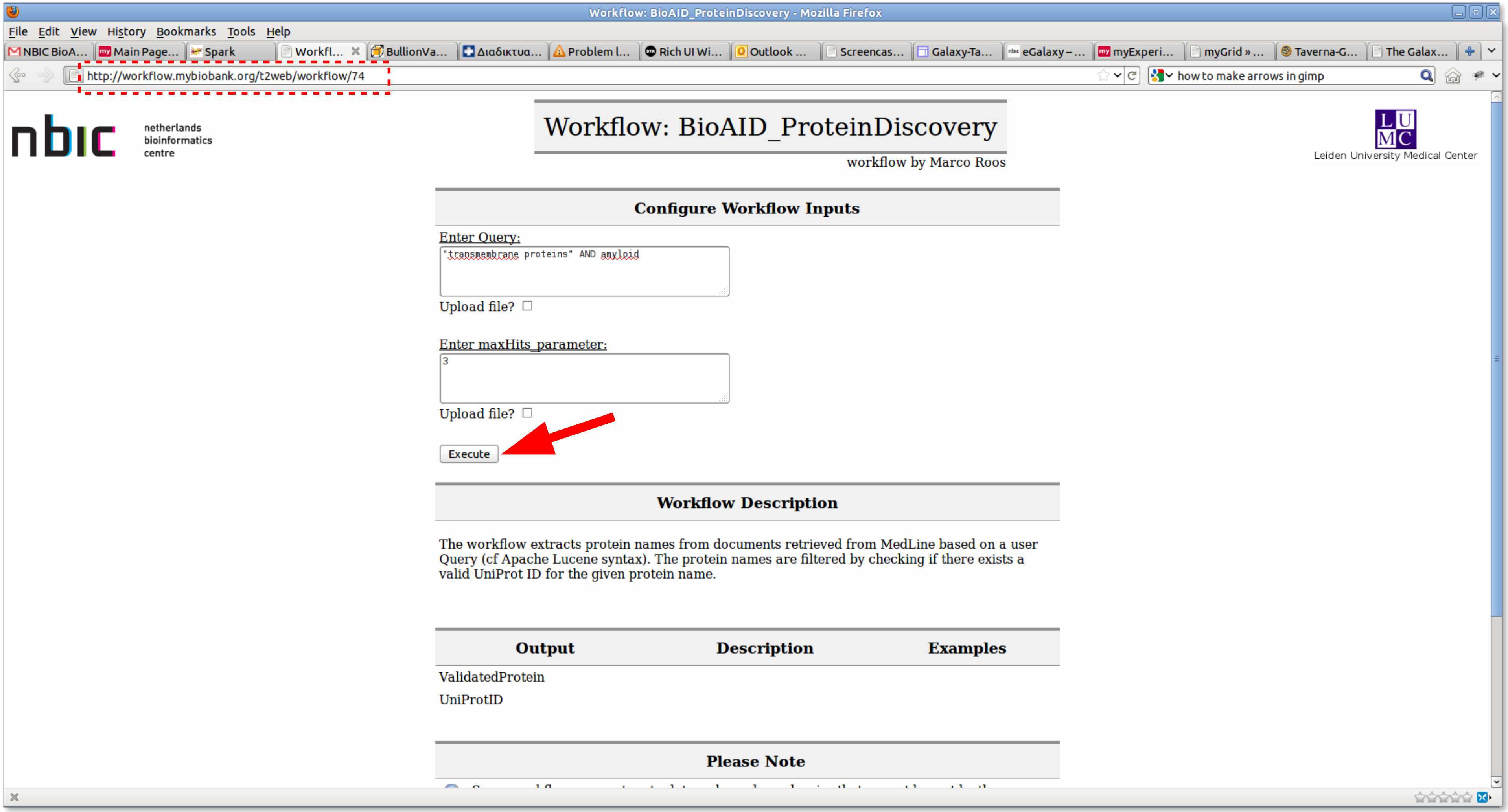
The workflow extracts protein names from documents retrieved from MedLine based on a user Query (cf Apache Lucene syntax). The protein names are filtered by checking if there exists a valid UniProt ID for the given protein name.

Output	Description	Examples
ValidatedProtein		
UniProtID		

Please Note

Who does what?

Taverna Workflows on a Web Browser (2)


Workflow: BioAID_ProteinDiscovery - Mozilla Firefox
File Edit View History Bookmarks Tools Help
NBIC BioA… Main Page… Spark Workfl… BullionVa… Διαδικτυα… Problem l… Rich UI Wi… Outlook … Screencas… Galaxy-Ta… eGalaxy – … myExperi… myGrid » … Taverna-G… The Galax…
how to make arrows in gimp
netherlands bioinformatics centre
Leiden University Medical Center

http://workflow.mybiobank.org/t2web/workflow/74

Workflow: BioAID_ProteinDiscovery

workflow by Marco Roos

Configure Workflow Inputs

Enter Query:

"transmembrane proteins" AND amyloid

Upload file? □

Enter maxHits_parameter:

3

Upload file? □

Execute

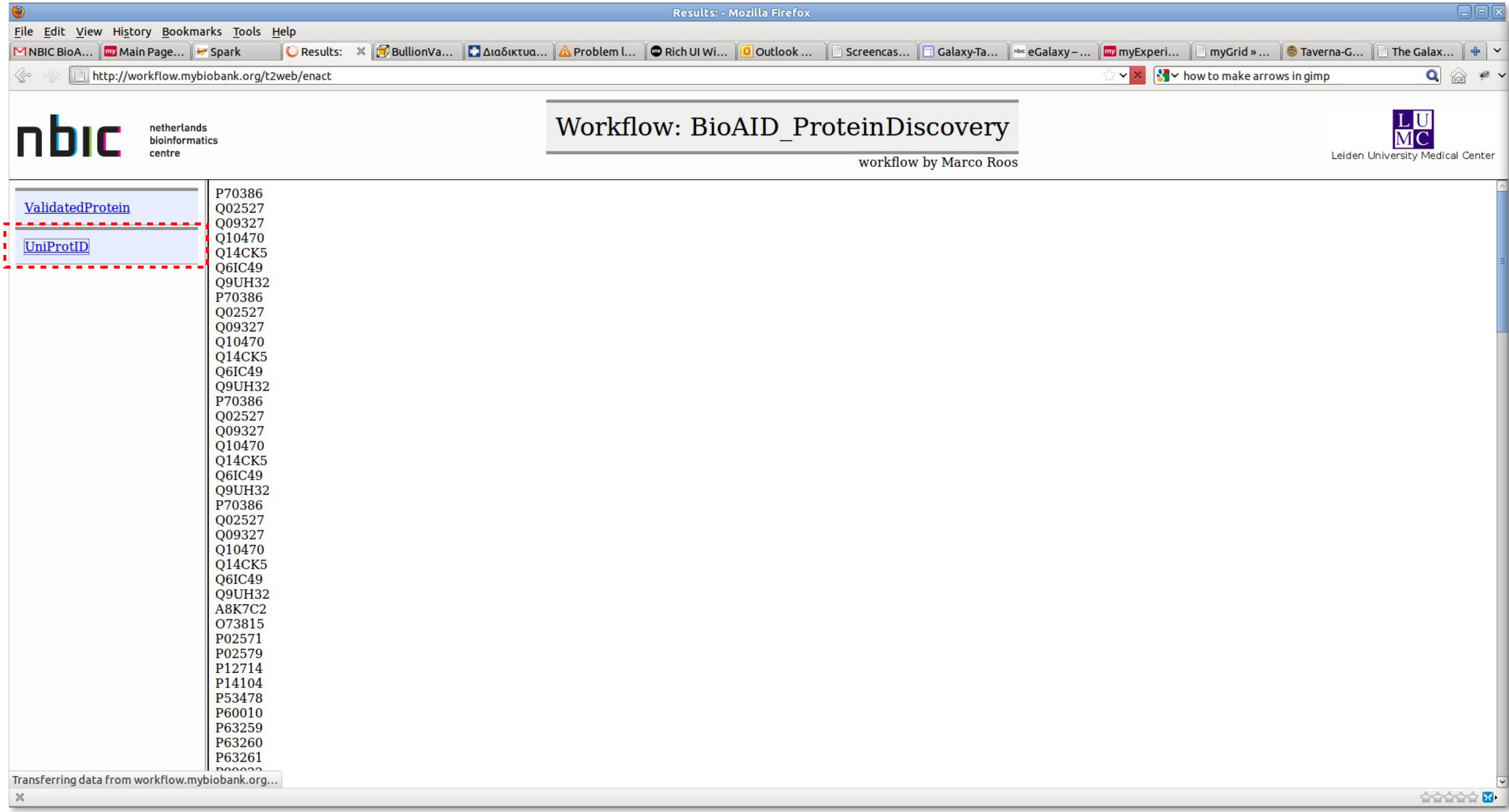
Workflow Description

The workflow extracts protein names from documents retrieved from MedLine based on a user Query (cf Apache Lucene syntax). The protein names are filtered by checking if there exists a valid UniProt ID for the given protein name.

Output	Description	Examples
ValidatedProtein		
UniProtID		

Please Note

Taverna Workflows on a Web Browser (3)


Results: - Mozilla Firefox
File Edit View History Bookmarks Tools Help
NBIC BioA… Main Page… Spark Results: BullionVa… Διαδικτυα… Problem l… Rich UI Wi… Outlook … Screencas… Galaxy-Ta… eGalaxy – … myExperi… myGrid » … Taverna-G… The Galax…
http://workflow.mybiobank.org/t2web/enact
how to make arrows in gimp
nbic netherlands bioinformatics centre
Leiden University Medical Center
Transferring data from workflow.mybiobank.org…

Workflow: BioAID_ProteinDiscovery

workflow by Marco Roos

ValidatedProtein

UniProtID

P70386
Q02527
Q09327
Q10470
Q14CK5
Q6IC49
Q9UH32
P70386
Q02527
Q09327
Q10470
Q14CK5
Q6IC49
Q9UH32
P70386
Q02527
Q09327
Q10470
Q14CK5
Q6IC49
Q9UH32
P70386
Q02527
Q09327
Q10470
Q14CK5
Q6IC49
Q9UH32
A8K7C2
O73815
P02571
P02579
P12714
P14104
P53478
P60010
P63259
P63260
P63261

Summary

- Workflow example using several approaches

- Taverna workfkows can be accessed in Galaxy
  - Bioinformatician *creates/finds* appropriate workflow
  - He uses Taverna-Galaxy to create new tool and installs it
  - … biologist will see the new tool in the Galaxy server
    - The taverna workflow can now take part in a Galaxy workflow

- Taverna workflows can be accessed via the web
  - Bioinformatician *creates/finds* appropriate workflow
  - … sends the URL to biologist

- More information
  - http://galaxy.psu.edu/
  - http://www.taverna.org.uk/
  - https://trac.nbic.nl/elabfactory/wiki/eGalaxy

- Questions?
  - kostas.karasavvas@nbic.nl