

Taverna: The ^{my}Grid Workflow Enactment Environment

Matthew Pocock

School of Computer Science

Taverna Overview

- Becoming the ^{my}Grid Front End
- Concentrates on producing useable workflow tools, object models, APIs
- Intended to be used by non computer scientists, particularly life science domain experts
- LGPL, hosted at <http://taverna.sf.net>
 - User and developer mailing lists
 - Documentation, demonstration workflows and tutorials
 - Download it and try!

Workflows

- Description of an in-silico experiment
- Processors that manipulate data
- Pipe the output of one Processor to the input of the next one
- Iteration over multiple inputs
- Branching and conditions

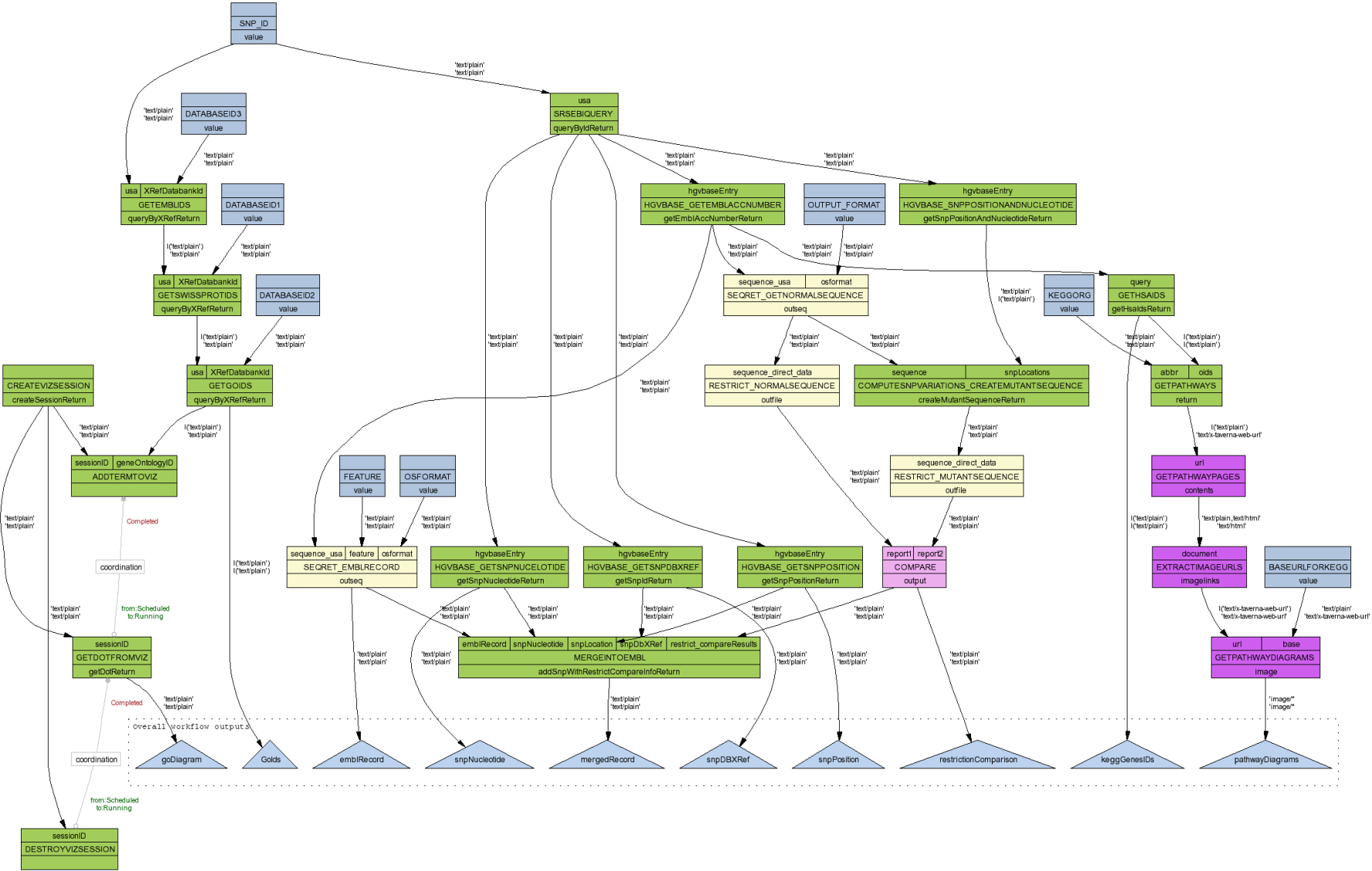
Taverna Processors

- SOAP/WSDL
- SoapLab
- BioMoby
- Internal Java Processors
- Others coming
 - Shims
 - BJV2 Utilities
 - ...

Some ^{my}Grid Middleware Features Used by Taverna

- Freefluo workflow enactment engine (embedded or remote)
- LSID assignment and resolution
- MIR (^{my}Grid information repository)
- Notification

Real life science workflow



Taverna Running

OldDemo.avi - Camtasia Player

File Play Help

Scufl Workbench

Tools and Workflow Invocation

Taverna
Tom Oinn, Matthew Pocock, Justus Tim Carver, Mark Greenwood, Pe myGrid team.

Advanced model explorer

Workflow: Object properties

Load Load from web Save New subworkflow Reset

Workflow object	Retries	Delay	Backoff	Critical
▲ Pepstats				
▶ Processors				
▶ Databaseld1 : GO	0	0	1	<input type="checkbox"/>
▶ KEGGBASEURL : http	0	0	1	<input type="checkbox"/>
▶ Databaseld2 : MEDLI	0	0	1	<input type="checkbox"/>
▶ ProbeSetId : 38276_ε	0	0	1	<input type="checkbox"/>
▶ Databaseld3 : pdlb	0	0	1	<input type="checkbox"/>
▶ REGEX_GROUP : 1	0	0	1	<input type="checkbox"/>
▶ GETIMAGE	0	0	1	<input type="checkbox"/>
▶ INTERPRO_REGEX : I	0	0	1	<input type="checkbox"/>
▶ INTERPRO_URL : http	0	0	1	<input type="checkbox"/>
▶ KEGGORG : hsa	0	0	1	<input type="checkbox"/>

Workflow diagram

Save as DOT PNG SVG Show types Bound ports Vertical Fit to window

Available services

Search list Watch loads

- Filter list of strings by regex
- String list union
- Split string into string list by regula
- Concatenate two strings
- Filter list of strings extracting matc
- String list intersection
- base64
 - Encode byte[] to base64
 - Decode base64 to byte[]
- moby
 - Create moby data
 - Parse moby data
- test
 - Create lots of strings
 - TEST - always fails
- Notification Processor
- String Constant
- WSDL @ http://www.ebi.ac.uk/collab/mygrid/se
- WSDL @ http://soap.genome.ad.jp/KEGG.wsdl
- WSDL @ http://mygrid.ncl.ac.uk/axis/services/
- porttype: AffymetrixMapper [RPC]
 - getSequence
 - getSwissProtId
 - getAccessionNumbers
 - getAccessionNumber
 - getOmidId
 - getUnigenId
 - getLocusLinkId
 - getECNumber
 - getMolecularFunctionGOIds
 - getCellularComponentGOIds
 - getBiologicalProcessGOIds
 - getInterProIds
- WSDL @ http://mygrid.ncl.ac.uk/axis/services/K
- WSDL @ http://mygrid.ncl.ac.uk/axis/services/S
- Soaplab @ http://industry.ebi.ac.uk/soap/soap/
- Soaplab @ http://www.ebi.ac.uk/collab/mygrid/

Stopped 01:02/05:50

Acknowledgements

myGrid is an EPSRC funded UK eScience Program Pilot Project



EMBL
European Bioinformatics Institute



Particular thanks to Tom Oinn and the other members of the Taverna project, <http://taverna.sf.net>