

Workflows in Astrogrid with Taverna

Nicholas Walton (IoA, Cambridge)
and
Kevin Benson (MSSL, UCL)

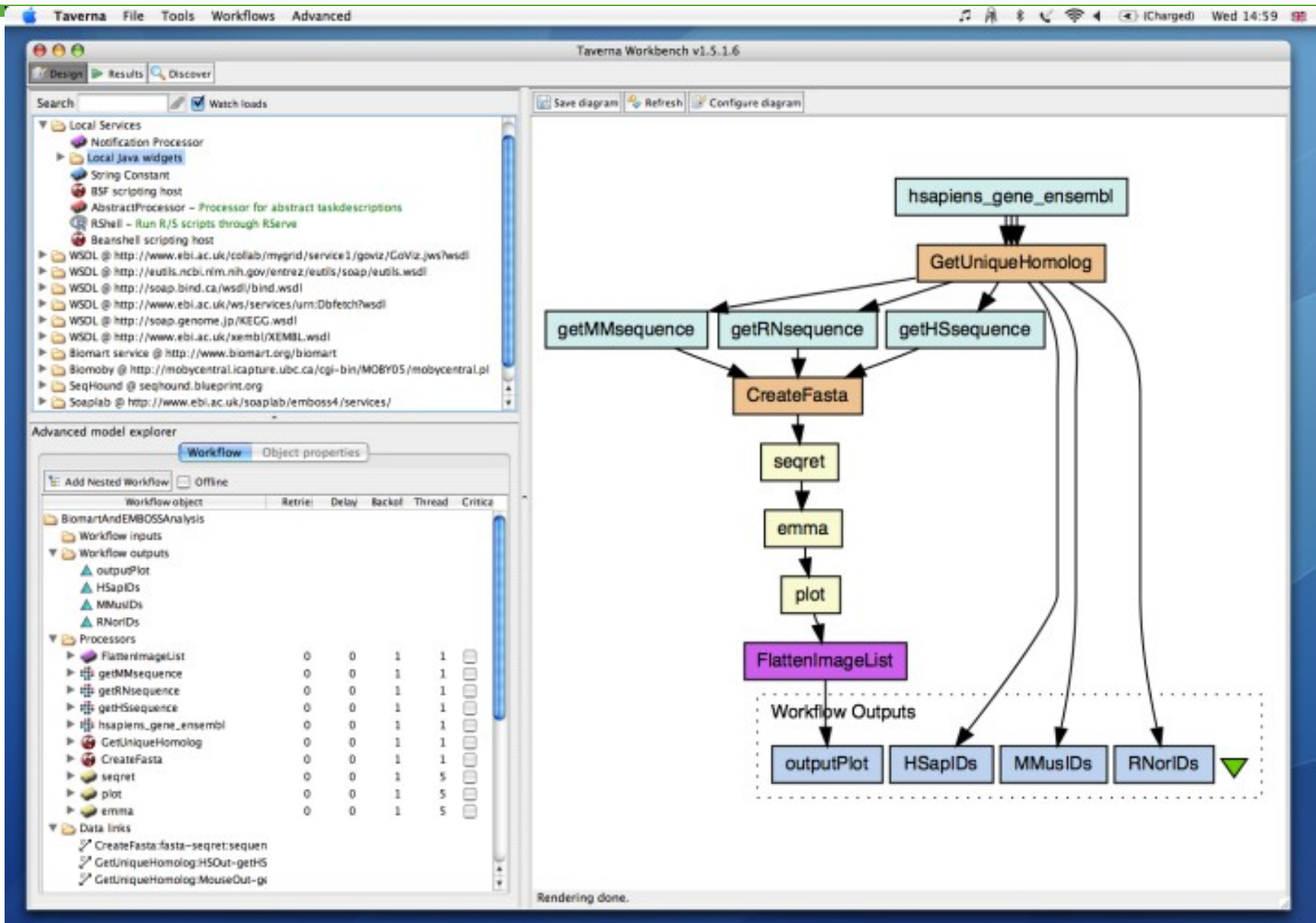
- The Taverna workbench is free software tool for designing and executing workflows, created by the myGrid project. <http://taverna.sourceforge.net/>
- Allows a biologist or bioinformatician with limited computing background and limited technical resources and support to construct highly complex analyses over public and private data and computational resources
 - all from a standard PC, UNIX box or Apple computer.
- The myExperiment social web site supports finding and sharing of workflows. <http://www.myexperiment.org/>

Taverna Architecture



- Taverna Plug-in architecture allows simple connectivity to standard Virtual Observatory services
 - ESO uses a tool called Reflex in implementing data reduction pipelines via Taverna.
 - Japanese VO will have a plugin to Taverna
- AstroGrid have developed a plug-in adaptor for Taverna, thereby exposing the standard range of IVOA standard services, thus those for images (SIAP), spectra (SSAP), tables (TAP), and applications (CEA-Common Execution Architecture) to Taverna.

Examples Taverna workflow



The screenshot displays the Taverna Workbench v1.5.1.6 interface. The main window shows a workflow diagram with the following components and flow:

- Input:** `hsapiens_gene_ensembl` (blue box)
- Process:** `GetUniqueHomolog` (orange box)
- Parallel Processes:** `getMMsequence`, `getRNsequence`, and `getHSsequence` (light blue boxes) receive input from `GetUniqueHomolog`.
- Process:** `CreateFasta` (orange box) receives input from the three parallel processes.
- Process:** `seqret` (yellow box) receives input from `CreateFasta`.
- Process:** `emma` (yellow box) receives input from `seqret`.
- Process:** `plot` (yellow box) receives input from `emma`.
- Process:** `FlattenImageList` (purple box) receives input from `plot`.
- Workflow Outputs:** A dashed box contains `outputPlot`, `HSapiIDs`, `MMusIDs`, and `RNorIDs` (blue boxes).
 - `FlattenImageList` outputs to `outputPlot`.
 - `GetUniqueHomolog` outputs to `HSapiIDs`, `MMusIDs`, and `RNorIDs`.

The **Advanced model explorer** on the left shows a table of workflow objects:

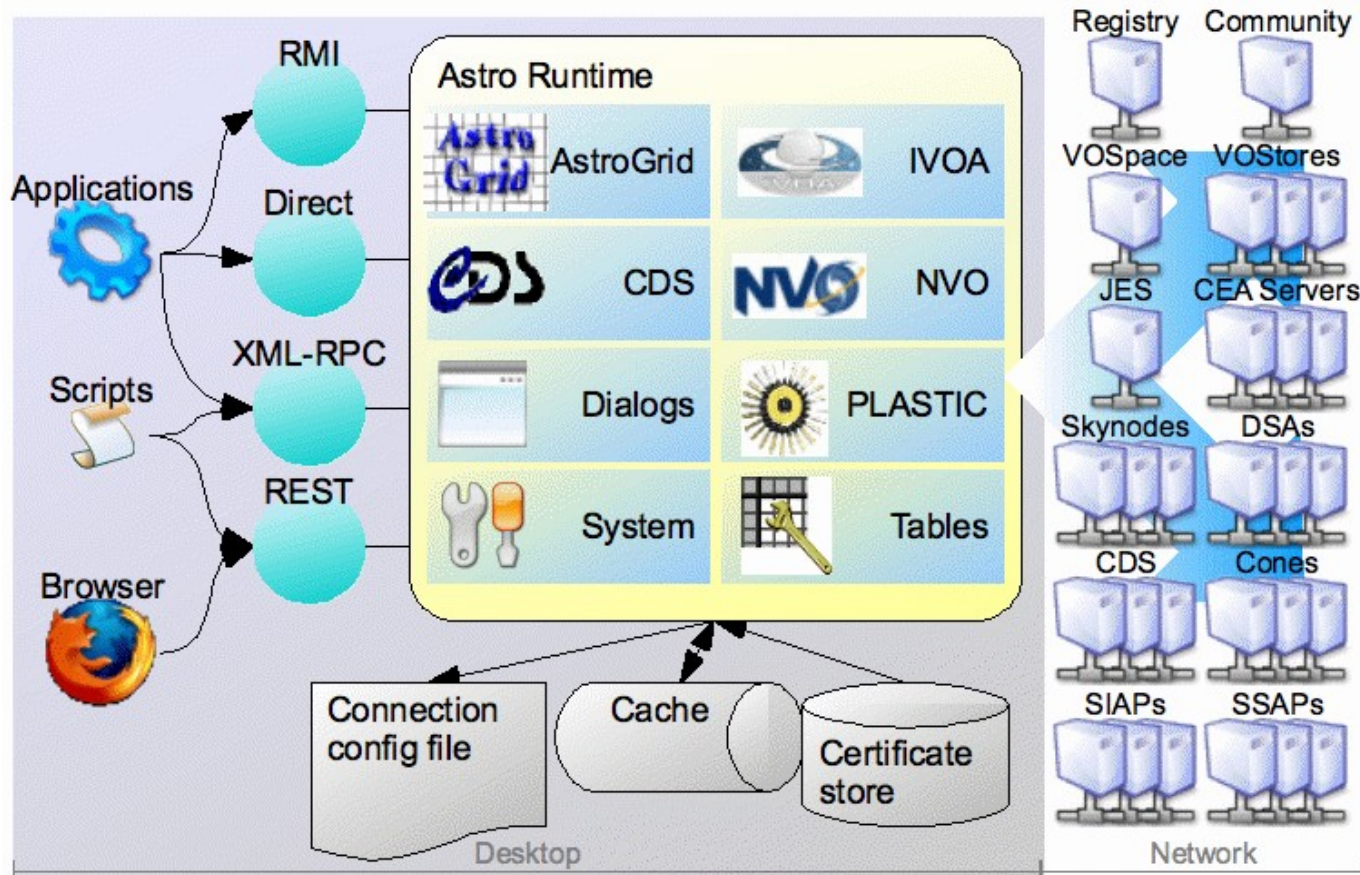
Workflow object	Retrie	Delay	Backof	Thread	Critica
BiomartAndEMBOSSAnalysis					
Workflow inputs					
Workflow outputs					
outputPlot					
HSapiIDs					
MMusIDs					
RNorIDs					
Processors					
FlattenImageList	0	0	1	1	
getMMsequence	0	0	1	1	
getRNsequence	0	0	1	1	
getHSsequence	0	0	1	1	
hsapiens_gene_ensembl	0	0	1	1	
GetUniqueHomolog	0	0	1	1	
CreateFasta	0	0	1	1	
seqret	0	0	1	5	
emma	0	0	1	5	
plot	0	0	1	5	
emma	0	0	1	5	
Data links					
CreateFasta.fasta-seqret:sequen					
GetUniqueHomolog:HSOut-getHS					
GetUniqueHomolog:MouseOut-gi					

AstroGrid Astro Runtime



- A client-side library of functions to access VO
 - scope: integrate all VO standards, popular ad-hoc services, suitable helper functions
- Library is exposed as a desktop service
 - accessible from almost all programming languages
 - XMLRPC, HTTP, RMI, C-binding
- API designed to use consistent abstractions and types
 - shallow learning curve
 - Insulates from change.

Astro Runtime - Schematic



<http://www2.astrogrid.org/desktop/astro-runtime>
<http://www.astrogrid.org>

Client-side Science Scripting



- Command-line access to VO
 - through Python scripts (or perl, or other language) calling AstroRuntime routines.
 - astronomer can integrate VO services with their existing toolset.
- Power of the 'for loop'.
 - remote app. execution
 - sextractor script, montage mosaic tool
 - image search at multiple positions
 - cone search over multiple archives

Client-side Workflow

- Taverna workflow engine
 - Graphical interface
 - Step-through execution
- Connects to Astro Runtime to call Virtual Observatory services.

Example AstroGrid workflow

Taverna Workbench v1.7.1.0

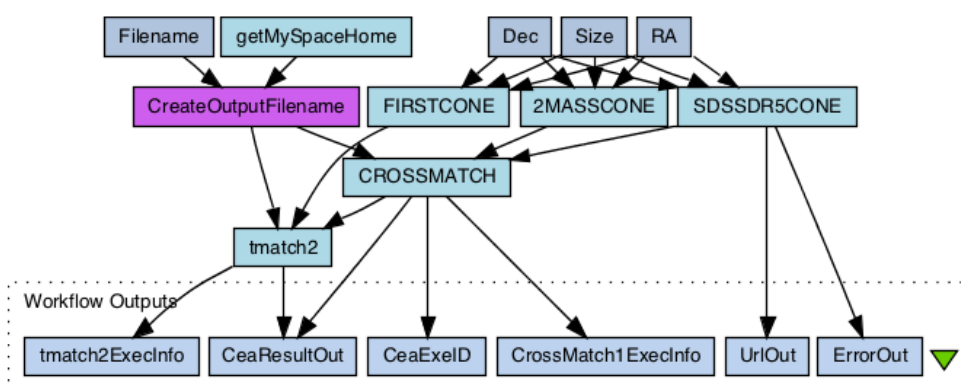
Design Results Taverna 2 preview

Search Watch loads

- Available Processors
 - Local Services
 - Astro Runtime CEA
 - Astro Runtime
 - Astro Runtime DSA
 - Astro Runtime VOHTTP
 - Astro Runtime Myspace

Graphical Interactive (experimental)

Save diagram Refresh Configure diagram



Workflow Outputs

tmatch2ExecInfo CeaResultOut CeaExelD CrossMatch1 ExecInfo UrlOut ErrorOut

Rendering done.

Advanced model explorer

Workflow Object properties

Add Nested Workflow Offline

Workflow object	Retries	Delay	Backoff	Threads	Critical
NewConeQuerySdssdr5_1					
Workflow inputs					

Example AstroGrid workflow

Taverna Workbench v1.7.1.0

Design Results Taverna 2 preview

NewConeQuerySdssdr5_TwoMass 3:03 PM NewConeQuerySdssdr5_TwoMass 3:06 PM

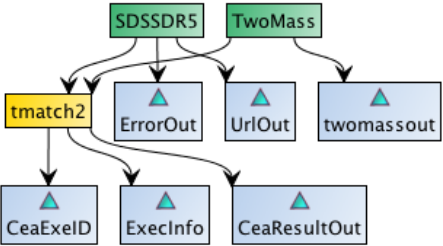
Workflow Status : **Running**

Status

Processor statuses

Ty...	Name	Last event	Event timestamp	Event detail	Breakpoint
AG	TwoMass	ProcessComplete	Jun 4, 2008 3:06:05 PM	.	.
AG	SDSSDR5	ProcessComplete	Jun 4, 2008 3:06:05 PM	.	.
AG	tmatch2	InvokingWithIteration	Jun 4, 2008 3:06:05 PM	IterationNumber='1' IterationTotal=...	.

Graph Intermediate inputs Intermediate outputs



```
graph TD; SDSSDR5[SDSSDR5] --> tmatch2[tmatch2]; TwoMass[TwoMass] --> tmatch2; tmatch2 --> ErrorOut[ErrorOut]; tmatch2 --> UrlOut[UrlOut]; tmatch2 --> twomassout[twomassout]; ErrorOut --> CeaExelD[CeaExelD]; UrlOut --> ExecInfo[ExecInfo]; twomassout --> CeaResultOut[CeaResultOut];
```

Links

- AstroGrid: <http://www.astrogrid.org>
- Astro-Runtime:
<http://www2.astrogrid.org/desktop/astro-runtime>
- Taverna: <http://taverna.sourceforge.net>
- MyExperiment: <http://www.myexperiment.org/>

OGF IPR Policies Apply

- “I acknowledge that participation in this meeting is subject to the OGF Intellectual Property Policy.”
- **Intellectual Property Notices Note Well:** All statements related to the activities of the OGF and addressed to the OGF are subject to all provisions of Appendix B of GFD-C.1, which grants to the OGF and its participants certain licenses and rights in such statements. Such statements include verbal statements in OGF meetings, as well as written and electronic communications made at any time or place, which are addressed to:
 - the OGF plenary session,
 - any OGF working group or portion thereof,
 - the OGF Board of Directors, the GFSG, or any member thereof on behalf of the OGF,
 - the ADCOM, or any member thereof on behalf of the ADCOM,
 - any OGF mailing list, including any group list, or any other list functioning under OGF auspices,
 - the OGF Editor or the document authoring and review process
- Statements made outside of a OGF meeting, mailing list or other function, that are clearly not intended to be input to an OGF activity, group or function, are not subject to these provisions.
- Excerpt from Appendix B of GFD-C.1: “Where the OGF knows of rights, or claimed rights, the OGF secretariat shall attempt to obtain from the claimant of such rights, a written assurance that upon approval by the GFSG of the relevant OGF document(s), any party will be able to obtain the right to implement, use and distribute the technology or works when implementing, using or distributing technology based upon the specific specification(s) under openly specified, reasonable, non-discriminatory terms. The working group or research group proposing the use of the technology with respect to which the proprietary rights are claimed may assist the OGF secretariat in this effort. The results of this procedure shall not affect advancement of document, except that the GFSG may defer approval where a delay may facilitate the obtaining of such assurances. The results will, however, be recorded by the OGF Secretariat, and made available. The GFSG may also direct that a summary of the results be included in any GFD published containing the specification.”
- OGF Intellectual Property Policies are adapted from the IETF Intellectual Property Policies that support the Internet Standards Process.

Full Copyright Notice

Copyright (C) Open Grid Forum (- 2008). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works.

The limited permissions granted above are perpetual and will not be revoked by the OGF or its successors or assignees.