

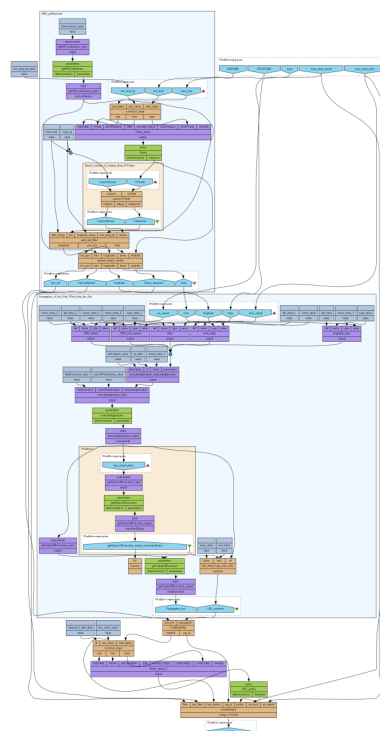
Taverna, myExperiment and HELIO services

Anja Le Blanc
Stian Soiland-Reyes
Alan Willams
University of Manchester



What is a Taverna Workflow?

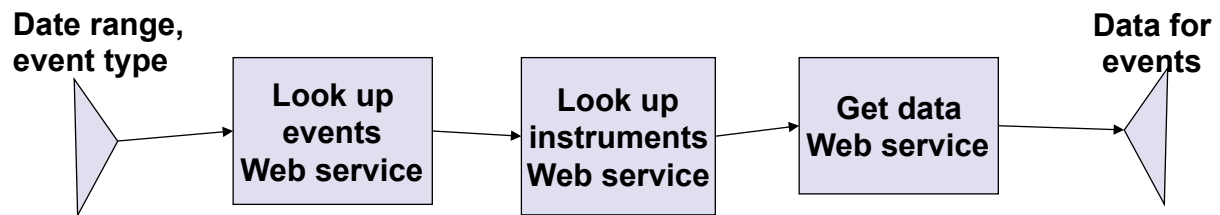
- Workflow management system
- Sophisticated analysis pipelines
- A set of **services** to analyze or manage data (either local or remote)
- Data flow through services
- Control of service invocation



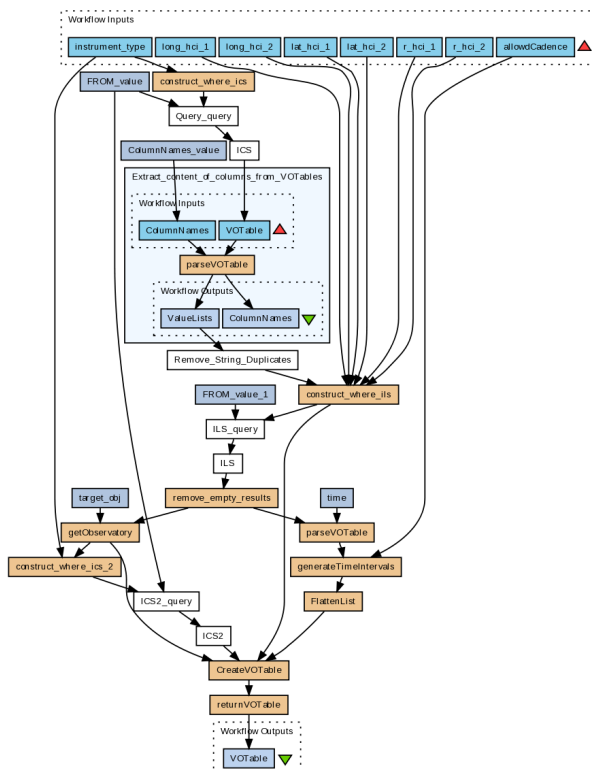
What is a Workflow?

Describes *what* you want to do, rather than focusing on *how* you want to do it

Simple dataflow language specifies how processes fit together



Taverna Workflows



- Interoperability, Integration and Collaboration
- Access to distributed and local resources
- Iteration over data sets
- Automation of **data flow**
- Provenance
- Extensible code
- Experimental science protocols



Workflows are ideal for...

- High throughput analysis
 - Find all events which fit a pattern
- Generalizing analyses
 - Move from analysis of one event to class of events
- Data integration, data interoperation
 - Combine web services from different providers
- Data management
 - Combine data from different services

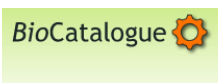


myGrid project

- Part of UK E-Science Programme
- myGrid started in 2001 – Universities Manchester, Newcastle, Sheffield, Nottingham, Southampton, Edinburgh, Oxford, EBI
- Director Prof Carole Goble, Manchester
- Now: Manchester, EBI, Oxford/Southampton
- Makes Open Source software
- Community building/community support



A Collection of Components



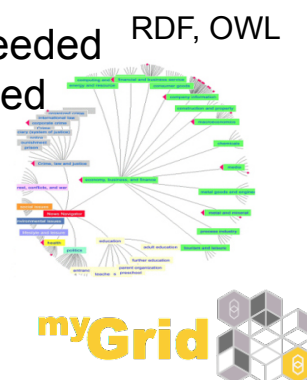
Discover and reuse services

Share, discover and reuse workflows **my experiment**

Create and run workflows

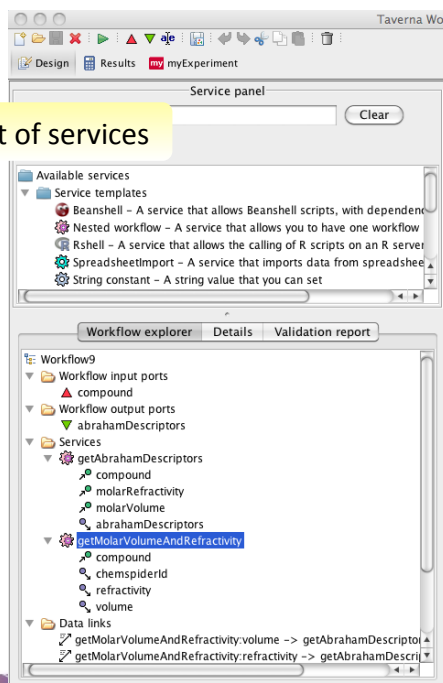


Manage the metadata needed and generated RDF, OWL



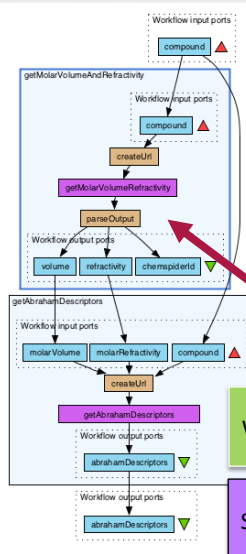
Taverna Workbench

List of services



Workflow engine to run workflows

Construct and visualise workflows



Web Services

e.g. HEC

Scripts

e.g. beanshell, R

Programming libraries

e.g. STIL library



What are Web Services?

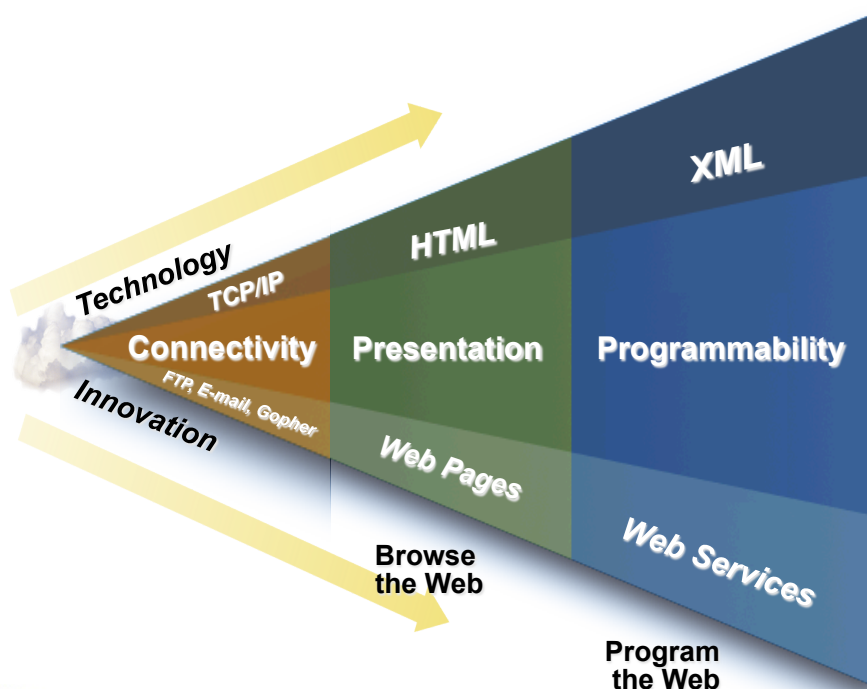
NOT the same as services on the web (i.e. web forms)

Web services support machine-to-machine interaction over a network

- SOAP/WSDL
- REST



Web Evolution



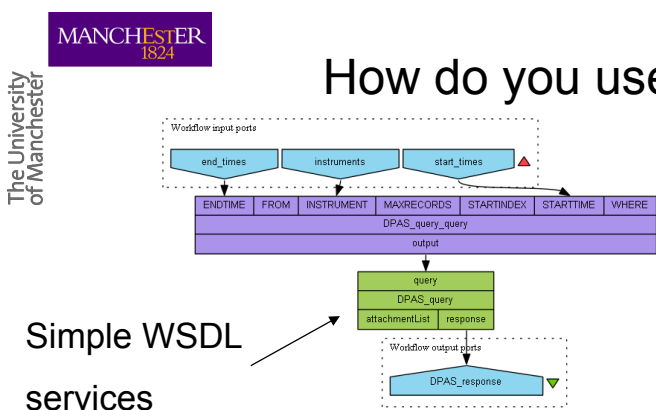
Taken from :<http://www.softstar-inc.com/>

What types of service?

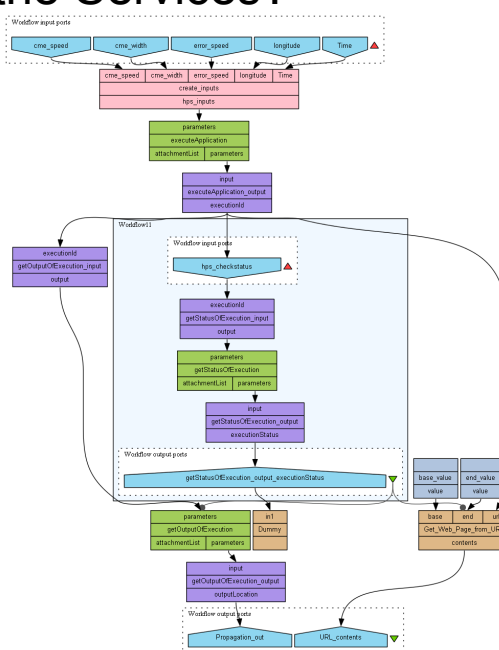
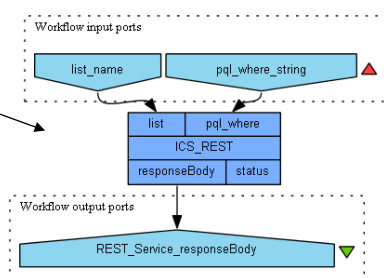
- WSDL Web Services
- REST
- R scripts
- Command line tools (local, ssh, grid)
- Java libraries
- Beanshell scripts (common scripts built-in)
- Nested workflows
- ... *your plugin here!*



How do you use the Services?



Simple REST service



Asynchronous services



How to find HELIO services

verity
chester

MANCHESTER
1824

How to find HELIO services

Browse Registered Resources - Mozilla Firefox

File Edit View History Bookmarks Tools Help

StarTable (STIL) Overview (Java Pla... http://www...ents.xsd STILTS coordinates.pdf (a... http://vo...vice/wsdll fits jar - Google Se... HELIO_pm.pdf (ap... Rserve installation ... Browse Registre... x

http://mskww.mssl.ucl.ac.uk:8080/helio_registry/browse.jsp

Most Visited Getting Started Latest Headlines BBC NEWS | News Fr...

HELIO

Helio Registry

Home Pages
Welcome
Setup & Admin
Investigation
View Resource
Browse Registry
Keyword Search
Registry XQuery
Registry Tree
Harvest Status
Registry
Create Entry
Edit Entry
Override Entry
Administration
Harvest Registry
Harvest VOSI
External Registry
Add Authority
Remove Resource
Edit Properties
Change Status
System Info
Database
Client & WebDav
Document
Installation
Configuration
FAQ
eXist Reference
Registry Reference
Uninstall
Upgrading
Resource Help

Browse Registry

Find IVORNs including: List

Title	Type	Authority ID	Resource Key	Updated	Actions
Helio Test Registry IVO	vg:Registry	helio_ivo.mssl.ucl.ac.uk	org.astrogrid.registry.RegistryService	2011-06-24T13:30:01	View, XML, Edit
Helio Test Registry IVO	vr:Organisation	helio_ivo.mssl.ucl.ac.uk	organisation	2009-08-11T14:03:40	View, XML, Edit
Registry	vg:Authority	helio_ivo.mssl.ucl.ac.uk	null?!	2011-07-01T17:54:10	View, XML, Edit
Heliophysics Event Catalogue	vs:CatalogService	helio-vo.eu	hec	2011-10-24T14:00:11	View, XML, Edit
Helio VO Project	vg:Authority	helio-vo.eu	null?!	2011-03-03T16:26:43	View, XML, Edit
Unified Observing Catalogue	vs:CatalogService	helio-vo.eu	uoc	2011-07-29T19:00:22	View, XML, Edit
Heliophysics Feature Catalogue	vs:CatalogService	helio-vo.eu	hfc	2011-07-28T10:58:40	View, XML, Edit
Instrument Capability Service	vs:CatalogService	helio-vo.eu	ics	2011-09-09T10:29:38	View, XML, Edit
Instrument Location Service	vr:Service	helio-vo.eu	ils	2011-09-09T10:23:17	View, XML, Edit

verity
chester

MANCHESTER
1824

How to find HELIO services cont ...

VAMDC Registry Access Pages - Mozilla Firefox

File Edit View History Bookmarks Tools Help

StarTable (STIL) Overview (Java Pla... http://www...ents.xsd STILTS coordinates.pdf (a... http://vo...vice/wsdll fits jar - Google Se... HELIO_pm.pdf (ap... Rserve installation ... VAMDC Registry ... x

http://mskww.mssl.ucl.ac.uk:8080/helio_registry/viewResourceEntry.jsp?IVORN=ivo%3A%2F%2Fhelio-vo.eu%2Fhec

Most Visited Getting Started Latest Headlines BBC NEWS | News Fr...

HELIO

Helio Registry

Home Pages
Welcome
Setup & Admin
Investigation
View Resource
Browse Registry
Keyword Search
Registry XQuery
Registry Tree
Harvest Status
Registry
Create Entry
Edit Entry
Override Entry
Administration
Harvest Registry
Harvest VOSI
External Registry
Add Authority
Remove Resource
Edit Properties
Change Status
System Info
Database
Client & WebDav
Document
Installation
Configuration

```

</curation>
<content>
  <subject>hec r3 service details</subject>
  <description>Heliophysics Event Catalogue Description</description>
  <referenceURL>http://www.astrogrid.org</referenceURL>
  <type>Other</type>
  <contentLevel>Research</contentLevel>
</content>
<capability standardID="ivo://helio-vo.eu/std/FullQuery/v0.2">
  <interface xsi:type="vs:ParamHTTP">
    <accessURL use="full">http://festung1.oats.inaf.it:8080/helio-hec/HelioQueryService</accessURL>
  </interface>
  <interface xsi:type="vr:WebService">
    <accessURL use="full">http://festung1.oats.inaf.it:8080/helio-hec/HelioService</accessURL>
    <accessURL use="full">http://festung3.oats.inaf.it:8080/helio-hec/HelioService</accessURL>
  </interface>
</capability>
<capability standardID="ivo://ivoa.net/std/VOSI#capabilities">
  <interface xsi:type="vs:ParamHTTP">
    <accessURL use="full">http://festung1.oats.inaf.it:8080/helio-hec/VOSI/capabilities</accessURL>
    <queryType>GET</queryType>
    <resultType>application/xml</resultType>
  </interface>
</capability>
<capability standardID="ivo://helio-vo.eu/std/FullQuery/Soap/v1.0">
  <interface xsi:type="vr:WebService">
    <accessURL use="full">http://festung1.oats.inaf.it:8080/helio-hec/HelioService</accessURL>
    <accessURL use="full">http://festung3.oats.inaf.it:8080/helio-hec/HelioService</accessURL>
  </interface>
</capability>
</capability standardID="ivo://helio-vo.eu/std/FullQuery/v1.0">

```

How to find HELIO services

Service Name: {http://helio-vo.eu/xml/QueryService/v1.0b}HelioQueryServiceService	Port Name: {http://helio-vo.eu/xml/QueryService/v1.0b}HelioQueryServicePort	Address: http://festung1.oats.inaf.it:8080/helio-hec/HelioService1_0b	WSDL: http://festung1.oats.inaf.it:8080/helio-hec/HelioService1_0b?wsdl	Implementation class: eu.heliovo.queryservice.server.query.SoopDispatcher
Service Name: {http://helio-vo.eu/xml/LongQueryService/v0.9}LongHelioQueryService	Port Name: {http://helio-vo.eu/xml/LongQueryService/v0.9}LongHelioQueryServicePort	Address: http://festung1.oats.inaf.it:8080/helio-hec/HelioLongQueryService	WSDL: http://festung1.oats.inaf.it:8080/helio-hec/HelioLongQueryService?wsdl	Implementation class: eu.heliovo.queryservice.server.query.SoopDispatcher
Service Name: {http://helio-vo.eu/xml/LongQueryService/v1.0}LongHelioQueryService	Port Name: {http://helio-vo.eu/xml/LongQueryService/v1.0}LongHelioQueryServicePort	Address: http://festung1.oats.inaf.it:8080/helio-hec/HelioLongQueryService1_0	WSDL: http://festung1.oats.inaf.it:8080/helio-hec/HelioLongQueryService1_0?wsdl	Implementation class: eu.heliovo.queryservice.server.query.SoopDispatcher
Service Name: {http://helio-vo.eu/xml/LongQueryService/v1.0b}LongHelioQueryService	Port Name: {http://helio-vo.eu/xml/LongQueryService/v1.0b}LongHelioQueryServicePort	Address: http://festung1.oats.inaf.it:8080/helio-hec/HelioLongQueryService1_0b	WSDL: http://festung1.oats.inaf.it:8080/helio-hec/HelioLongQueryService1_0b?wsdl	Implementation class: eu.heliovo.queryservice.server.query.SoopDispatcher
Service Name: {http://helio-vo.eu/xml/QueryService/v0.1}HelioQueryServiceServiceTwo	Port Name: {http://helio-vo.eu/xml/QueryService/v0.1}HelioQueryServicePortTwo	Address: http://festung1.oats.inaf.it:8080/helio-hec/HelioTavernaService	WSDL: http://festung1.oats.inaf.it:8080/helio-hec/HelioTavernaService?wsdl	Implementation class: eu.heliovo.queryservice.server.query.SoopDispatcher

HELIO Services

Heliophysics Event Catalogue (HEC)	Maintain and provide access to <i>existing</i> event data from all domains
Heliophysics Feature Catalogue (HFC)	Maintain and provide access to existing feature data from all domains
Data Evaluation Service (DES)	Allows the user to create an auxiliary event list based on a <i>newly-derived</i> parameter, etc.
Context Service (CXS)	Provide context information to help the user make a selection

HELIO Services

Instrument Capabilities Service (ICS)	Match required observation type to one or more instruments (each part of an observatory)
Instrument Location Service (ILS)	Determine the location of an instrument (part of an observatory) at a specified time
Universal Observing Catalogue (UOC)	Provide information on whether an instrument was making suitable observations at a specified time
Data Provider Access Service (DPAS)	Provide integrated access to data archives in all domains no matter how the data are stored or accessed

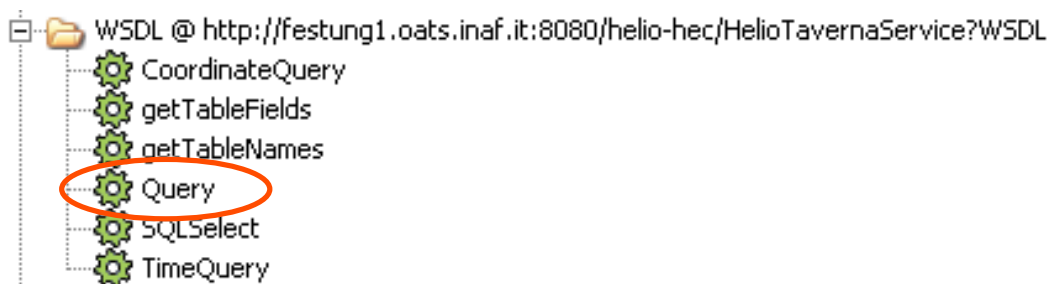


HELIO Services

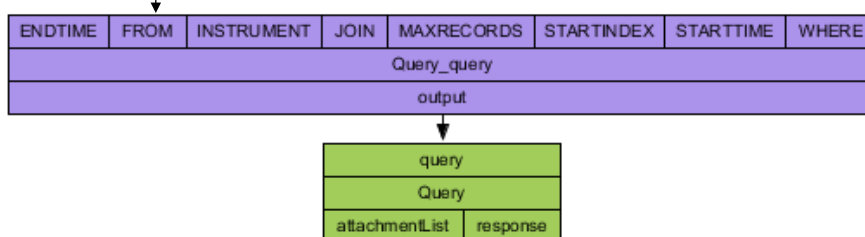
Processing Service (HPS)	Support processing on demand
Storage Service	Provide storage for user information (requires authentication)
Coordinate Transformation Service (CTS)	Translated between the different coordinate systems used by the communities
Semantic Mapping Service (SMS)	Maps terms used in the metadata from the different communities



HELIO Query Interface



Only required field



Workflow explorer Details Validation report

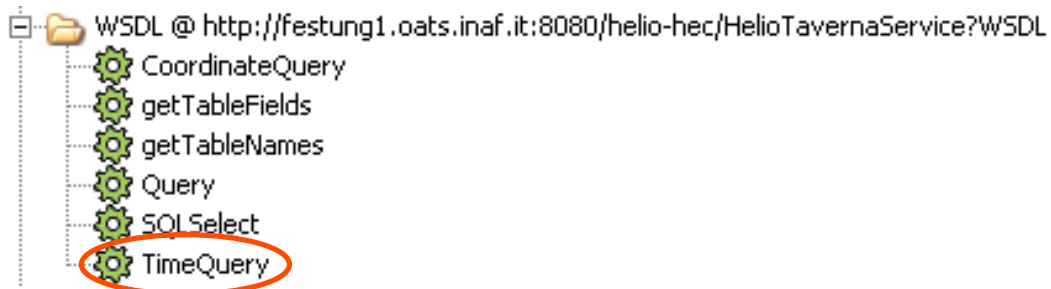
XML splitter

XML splitter	
Inputs	
JOIN	optional Depth:0 Query>JOIN
STARTINDEX	optional Depth:0 Query>STARTINDEX
FROM	Depth:1 Query>FROM
MAXRECORDS	optional Depth:0 Query>MAXRECORDS
WHERE	optional Depth:0 Query>WHERE
STARTTIME	optional Depth:1 Query>STARTTIME
	optional

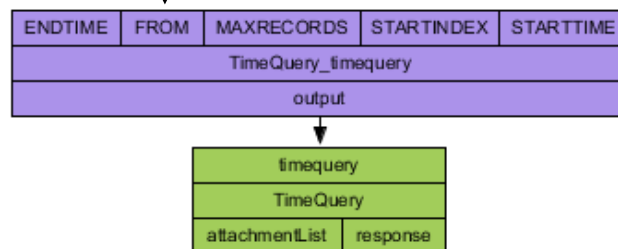
**Details to
Service
Inputs**



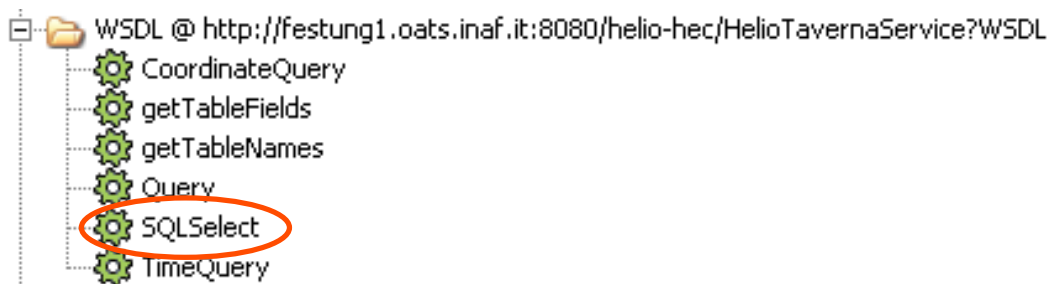
HELIO Query Interface



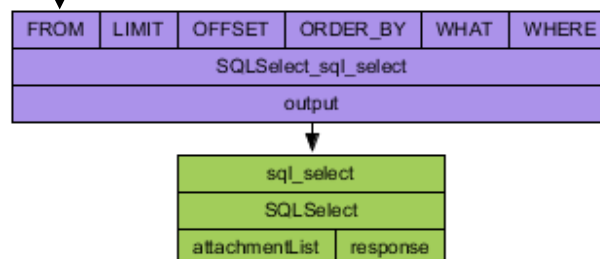
Only required field



HELIO Query Interface



Only required field



How to find which 'FROM' to use

<http://festung1.oats.inaf.it:8080/helio-hec/VOSI/tables>

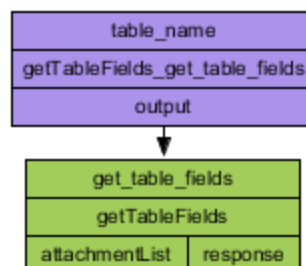
VOSI Tables view

hi_cme_list

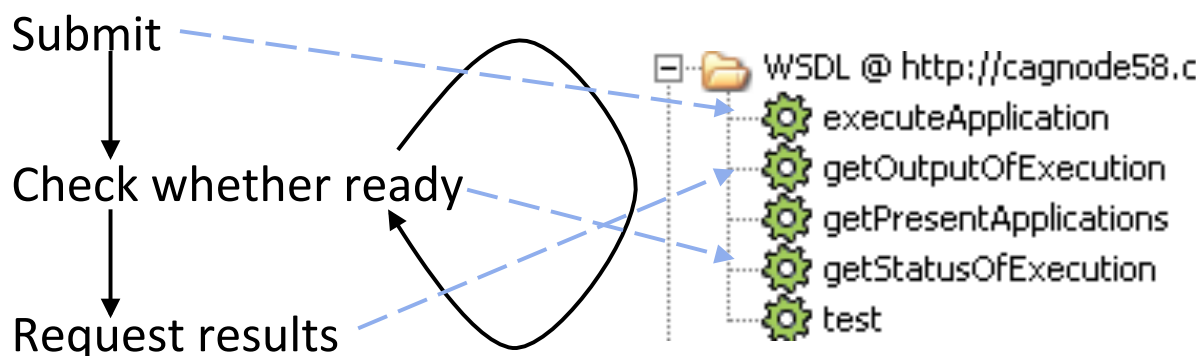
hi_cme_list_id	serial	Event identification number (HEC internal number)
time_start	timestamp	Start time - approximate time that CME enters HI field of view.
time_end	timestamp	End time - approximate time CME front leaves HI field of view.
instrument	varchar	This is the ID of the instrument on STEREO that was used to identify the event.
data_gap	int4	Indication of a data gap during the event
cme_type	varchar	Classification of the CME type
brightness	varchar	Classification of the CME brightness
pa_central	float8	Central position angle for the observed CME
pa_width	float8	Position angle width of the CME
time_onset	timestamp	Estimated onset time for the event
not_in_list	int4	Flag to indicate if the event is in the HI solar event list
linear_speed	float8	Fitted CME speed.
longitude	float8	Fitted Stonyhurst longitude of the event
longitude_err	float8	Error in the fitted longitude measurement
time_1au	timestamp	Time at which the CME was expected to reach 1 AU based on fitted time elongation profile
description	varchar	Description of the CME feature

cactus_stereo_cme

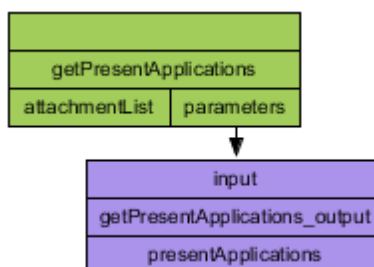
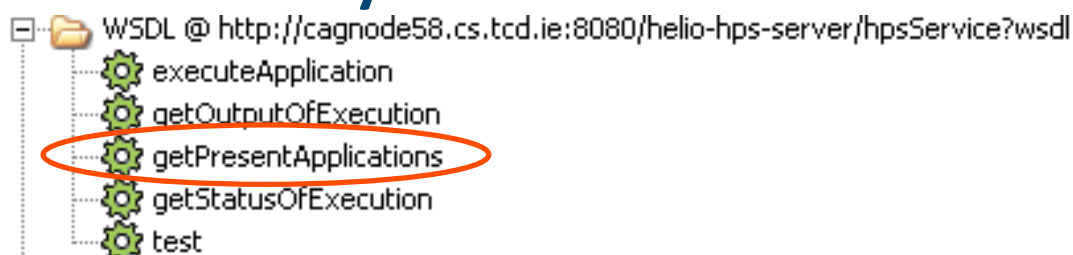
How to find which 'FROM' to use



Asynchronous SOAP call - HPS



Asynchronous SOAP call - HPS



Asynchronous SOAP call - HPS

The University
of Manchester

WSDL @ <http://cagnode58.cs.tcd.ie:8080/helio-hps-server/hpsService?wsdl>

-  executeApplication
-  getOutputOfExecution
-  getPresentApplications
-  getStatusOfExecution
-  test

def	name	type	value	def	name	type	value
	selectedApplication_input				selectedApplication_input_2		
	output				output		

description	id	name	parameters
		executeApplication_selectedApplication	
		output	

fastExecution	numOfParallelJobs	selectedApplication
		executeApplication_input
		output

parameters
executeApplication
attachmentList
parameters

input
executeApplication_output
executionId



Asynchronous SOAP call - HPS

The University
of Manchester

WSDL @ <http://cagnode58.cs.tcd.ie:8080/helio-hps-server/hpsService?wsdl>

-  executeApplication
-  getOutputOfExecution
-  getPresentApplications
-  getStatusOfExecution
-  test

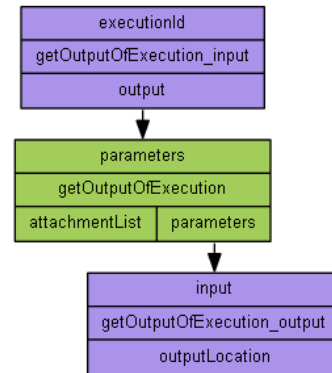
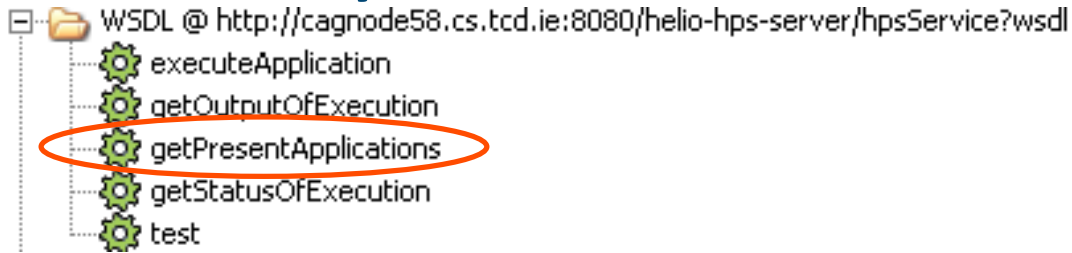
executionId
getStatusOfExecution_input
output

parameters
getStatusOfExecution
attachmentList
parameters

input
getStatusOfExecution_output
executionStatus



Asynchronous SOAP call - HPS



What do Scientists use Taverna for?

Systems biology model building

Sequence analysis Protein structure prediction

Gene/protein annotation Microarray data analysis

Phylogeny Model simulations sweeps **Astronomy**

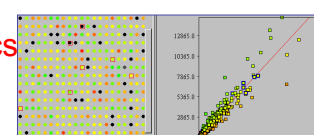
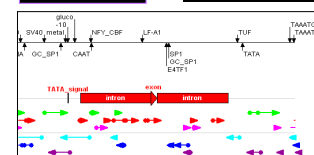
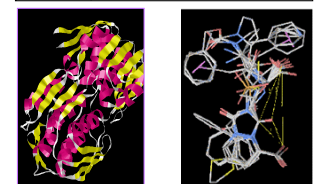
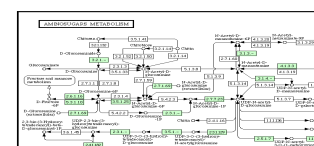
High throughput screening Proteomics **Music**

Phenotypical studies Text mining **Meteorology**

Public Health care epidemiology **Social Science**

Medical image analysis QTL studies **Cheminformatics**

QSAR studies Genome Wide Association Studies



Workflows are ...

... records and protocols

... know-how and intellectual property

... hard work to develop and get right

Maybe we should share workflows with the community?

Published online 25 May 2010

Nucleic Acids Research, 2010, Vol. 38, Web Server issue W677-W682
doi:10.1093/nar/gkq429

myExperiment: a repository and social network for the sharing of bioinformatics workflows

Carole A. Goble¹, Jiten Bhagat¹, Sergejs Aleksejevs¹, Don Cruickshank²,
Danius Michaelides², David Newman², Mark Borkum², Sean Bechhofer¹, Marco Roos^{3,4},
Peter Li^{1,*} and David De Roure²



MANCHESTER 1824

my experiment About | Mailing List | Publications Log in | Register | Give us Feedback | Invite

Home Users Groups **Workflows** Files Packs

Home » Workflows

Search filter terms

Filter by type

- Taverna 1 557
- Taverna 2 520
- RapidMiner 112
- Bioclipse Scri... 29
- GWorkflowDL 24
- LONI Pipeline 20
- Kepler 17
- BioExtract Server 13
- Trident (Packa... 10
- Chemistry Plan 7

Filter by tag

- example 178
- mvorid 103

Showing 1365 results. Use the filters on the left and the search box below to

Sort by: Rank

Rank

Title

Latest

Last updated

Community rating

Most viewed

Most downloaded

Type

Licence

Taverna 2 Pathways and Gene annotations for QTL region - Mouse (v6)

Original Uploader

Created: 19/11/09 @ 18:18:52 | Last updated: 05/04/11 @ 10:59:25

Credits: Paul Fisher

License: Creative Commons Attribution-Share Alike 3.0 Unported License

This workflow searches for genes which reside in a QTL (Quantitative Trait Loci) region in the mouse, *Mus musculus*. The workflow requires an input of: a chromosome name or number; a QTL start base pair position; QTL end base pair position. Data is then extracted from BioMart to annotate each of the genes found in this region. The Entrez and UniProt identifiers are then sent to KEGG to obtain KEGG gene identifiers. The KEGG gene identifiers are then used to search for pathways in the KEGG path...

New/Upload

Workflow GO

Log in / Register

Username or Email:

Password:

Remember me: ☐

OR

Use OpenID:

(eg: name.myopenid.com)

Log in

Need an account? Click here to register

Forgot Password?



“Just Enough” Sharing....

- myExperiment can provide a central location for workflows from one community/group
- myExperiment allows you to say
 - Who can look at your workflow
 - Who can download your workflow
 - Who can modify your workflow
 - Who can run your workflow
- Ownership and attribution



Ownership and Attribution

Upload Workflow

Workflow File:

Tags

Credit and Attribution

Defaults: you are the only person who gets credit, no attributions.

Sharing

Defaults: anyone can view, but only Friends can download; no one is allowed to modify.

License/Rights

Default: people are allowed to build on this Workflow, but must give author(s) credit and give attribution to this Workflow. They must also share under the same conditions. ([Creative Commons Attribution-Share Alike 3.0 License](#))

Share with my Groups:

☐ UsefulChem	View and Download only
☐ Taverna 2 beta tester programme	View and Download only
☐ Social Scientific Land	View and Download only
☐ Mark's Project	View and Download only
☐ GNU	View and Download only
☐ myExperiment	View and Download only
☐ Music workflows	View and Download only



The most important aspect of
myExperiment - Designed by scientists



my experiment beta

Users

Groups

Workflows

Files

Blogs

Forums

paul

Workflows

Search

Home » Workflows » View: Microarray CEL file to candidate pathways

Workflow Entry: Microarray CEL file to candidate pathways

All versions of this workflow are licensed under a [Creative Commons Attribution-NoDerivs 3.0 License](#).

Version: 2 (latest)

Change to: 2 (latest)

GO

Title: Microarray CEL file to candidate pathways

Version created on: Wednesday 03 October 2007 @ 18:35:55 (GMT Daylight Time)

Diagram: (click to expand)

Description:

This workflow takes in a <http://www.bioinf.manchester.ac.uk/MADAT/index.html>. Also

Uploader

Paul Fisher

Credits (People)

None

Attribution (Workflows)

None

Tags

All Tags

Uploader's Tags

insertional mutation

shim

shotgun method

similarity

simplifier

Note: the size of the tags show how popular they are on myExperiment.

Add Tags

New/Upload

Workflow

GO

Jits

Sign Out

Inbox

Memberships

My Stuff

11 friends | 1 group | 2 files | workflows

cmuzys1_2 readme.txt

rssbanditlist.opml

Workflows

Example of an alter...

ExcelFileReader and...

Show Gene Ontology ...

Fetch today's Dilbe...

My Tags

20 tags

Taverna on the cloud

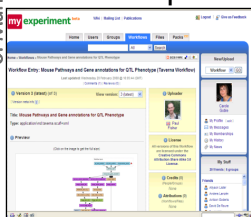
- Use-case:
 - SNP analysis and annotation of African cattle
 - ~11 million SNPs per cow
 - Genetic differences between resistant and susceptible cattle infected with Trypanosomiasis
 - Amazon EC2 with Taverna Server and local services
 - Custom (built-in-a-week) Ruby on Rails web interface
 - Runs through 30 chromosomes in ~2 hours using 10 instances ~ \$5-10

Data and Provenance

- Workflows can generate vast amount of data - how can we manage and track it?
- We need to manage data **AND** metadata **AND** experimental provenance
- Scientists need to check back over past results, compare workflow runs and share workflow runs with colleagues
- Scientists need to look at intermediate results when designing and debugging



Workflow Repository



Service Catalogue

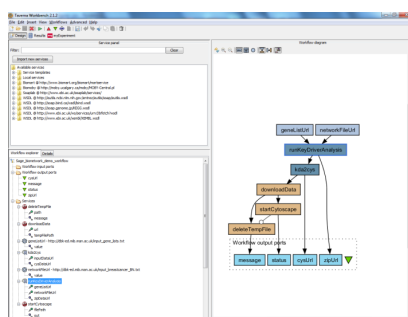


Activity and Service Plug-in Manager

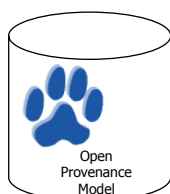


myGrid Open Suite of Tools

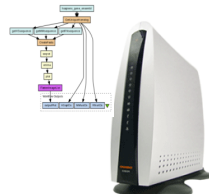
Workflow GUI Workbench



Provenance Store

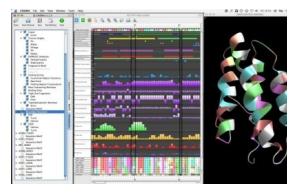


Workflow Server

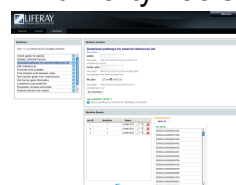


Secure Service Access

Client User Interfaces



Third Party Tools



Web Portal

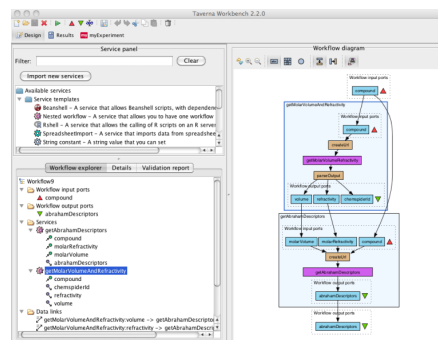
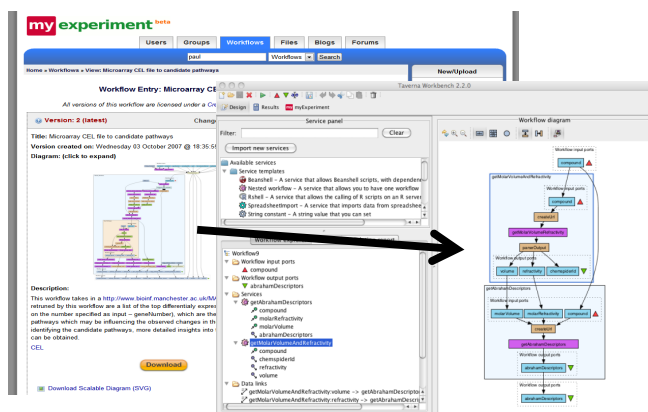


Programming and APIs



Spectrum of Users

Advanced users design and build workflows (informaticians)



Intermediate users reuse and modify existing workflows or components

<http://www.myexperiment.org>

Load Data:

Run Workflow



Others “replay” workflows through web page

Workflows and HELIO

- Integral part of project
- Diverse use
 - Creation of pseudo services which can be integrated in Front End
 - Implementing complex use-case which can't be done in the Front End
 - Execute same use-case with changing parameters
 - Testing of Services
- All workflows developed shared with ‘helio’ MyExprimment group



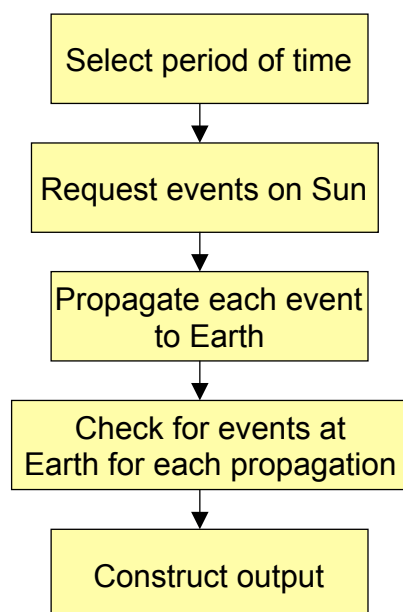
Case Study: Propagate Solar Events to Earth and find Related Events

Practical Session

- Build a workflow which requests flare events from a catalogue, propagates the position to Earth and checks for proton events around the predicted time



Propagate Solar Events to Earth and find Related Events



How long is a sensible period?

Restrictions on request?

Which type of propagation?

Which list to use?

What information do I need and from where?



Case Study: Propagate Solar Events to Earth and find Related Events

Practical Session

- Finding and invoking services
 - REST, WSDL, local services
 - Shims (helper services)
- Understanding the Taverna engine
 - Iteration, looping, retries, control links v' s data links
- Sharing and reusing workflows
- Plans for future workflow development



Summary

- Informatics often relies on data integration and large-scale data analysis
- Taverna workflows are a mechanism for linking together resources and analyses
- myExperiment allows you to reuse workflows and benefit from others work
- Workflows provide reusable analysis methods that you can easily make available to research communities



The Rest of Today

Tutorial

- Work through the exercises
- If something is unclear or you have a question call one of us
- Follow your curiosity
- At the end give us some feedback



myGrid

- <http://www.mygrid.org.uk>

- Taverna

- <http://www.taverna.org.uk>

- Taverna User Manual

- <http://www.mygrid.org.uk/dev/wiki/display/taverna/User+Manual>

- myExperiment

- <http://www.myexperiment.org>

- HELIO

- <http://www.helio-vo.eu>



More Information

