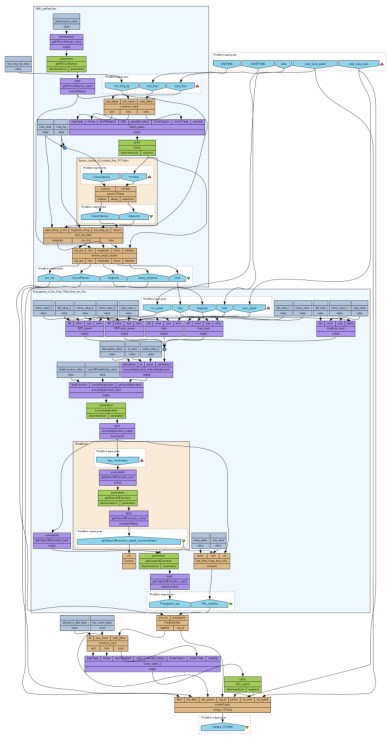
Taverna, myExperiment and HELIO services

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

What is a Taverna Workflow?

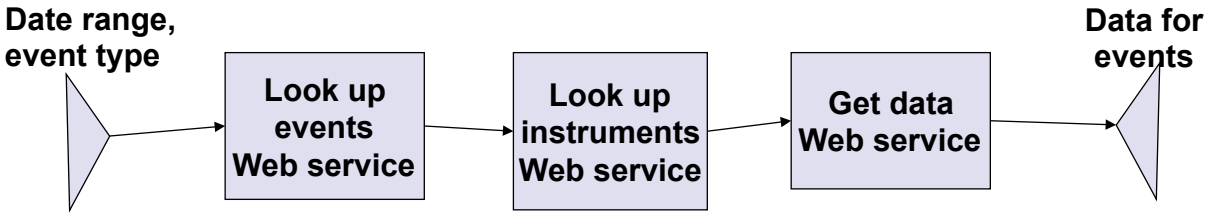
- Workflow management system
- Sophisticated analysis pipelines
- A set of **services** to analyze or manage data (either local or remote)
- Data flow through services
- Control of service invocation



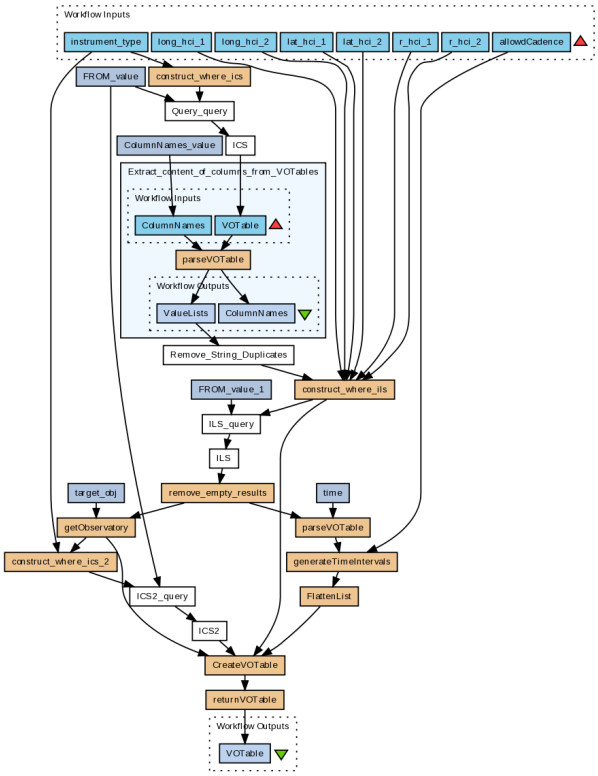
What is a Workflow?

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

BioCatalogue

Discover and reuse services

Share, discover and reuse workflows

myExperiment

Create and run workflows

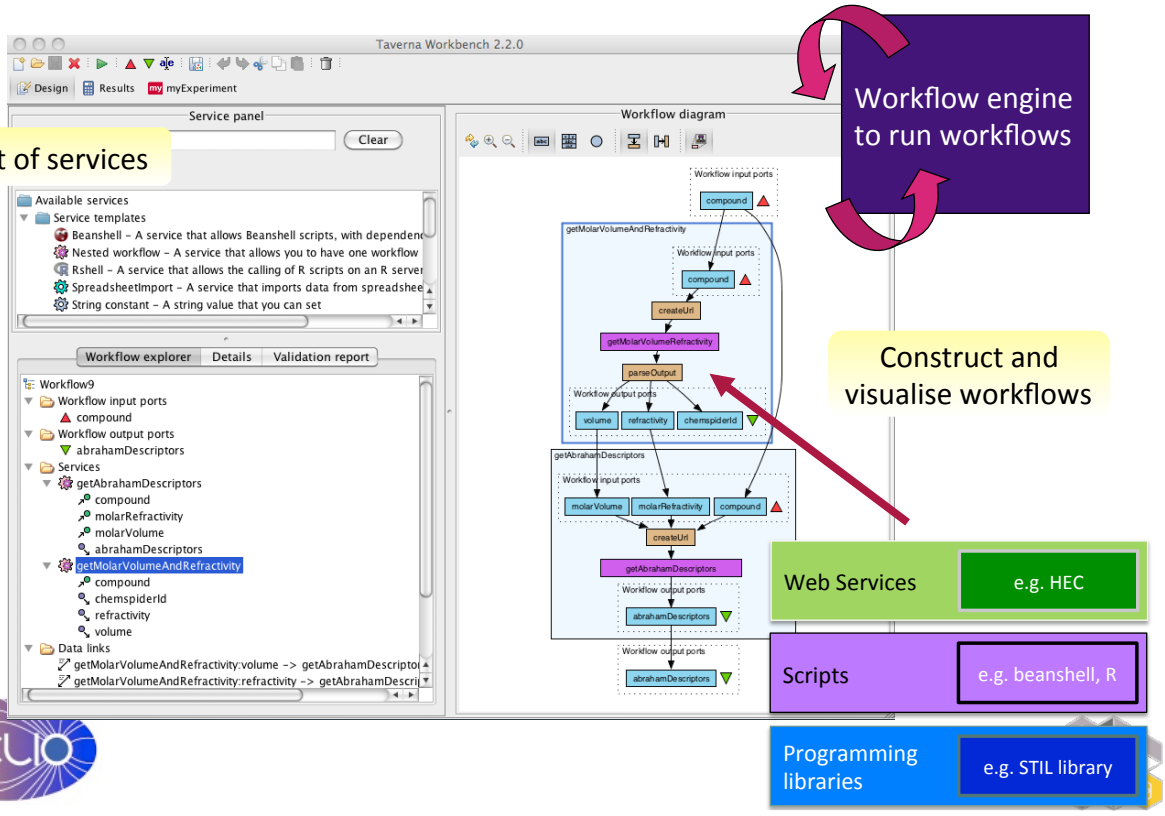
Taverna

Manage the metadata needed and generated

RDF, OWL

myGrid

Taverna Workbench



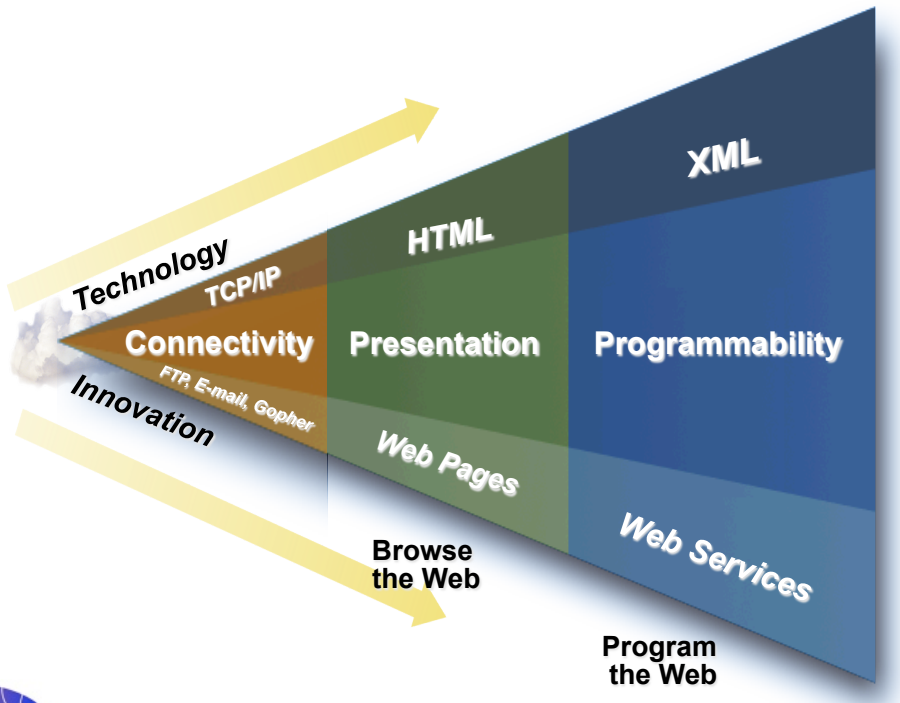
What are Web Services?

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



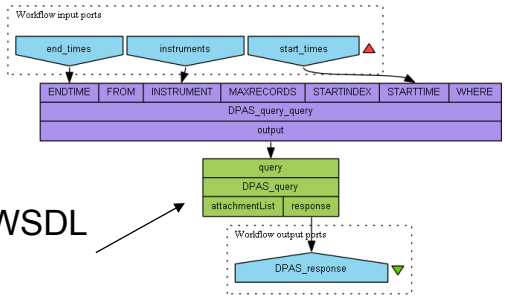
Taken from :http://www.softstar-inc.com/

What types of service?

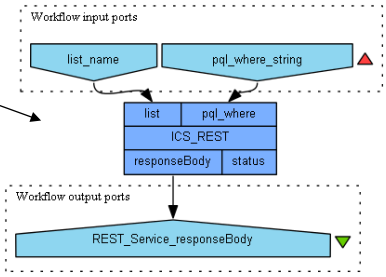
- WSDL Web Services
- REST
- R scripts
- Command line tools (local, ssh, grid)
- Java libraries
- Beanshell scripts (common scripts built-in)
- Nested workflows
- *… your plugin here!*

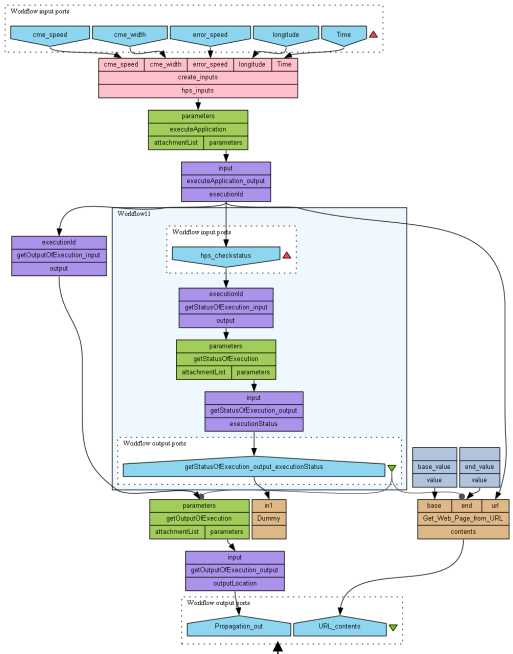
How do you use the Services?

Simple WSDL services



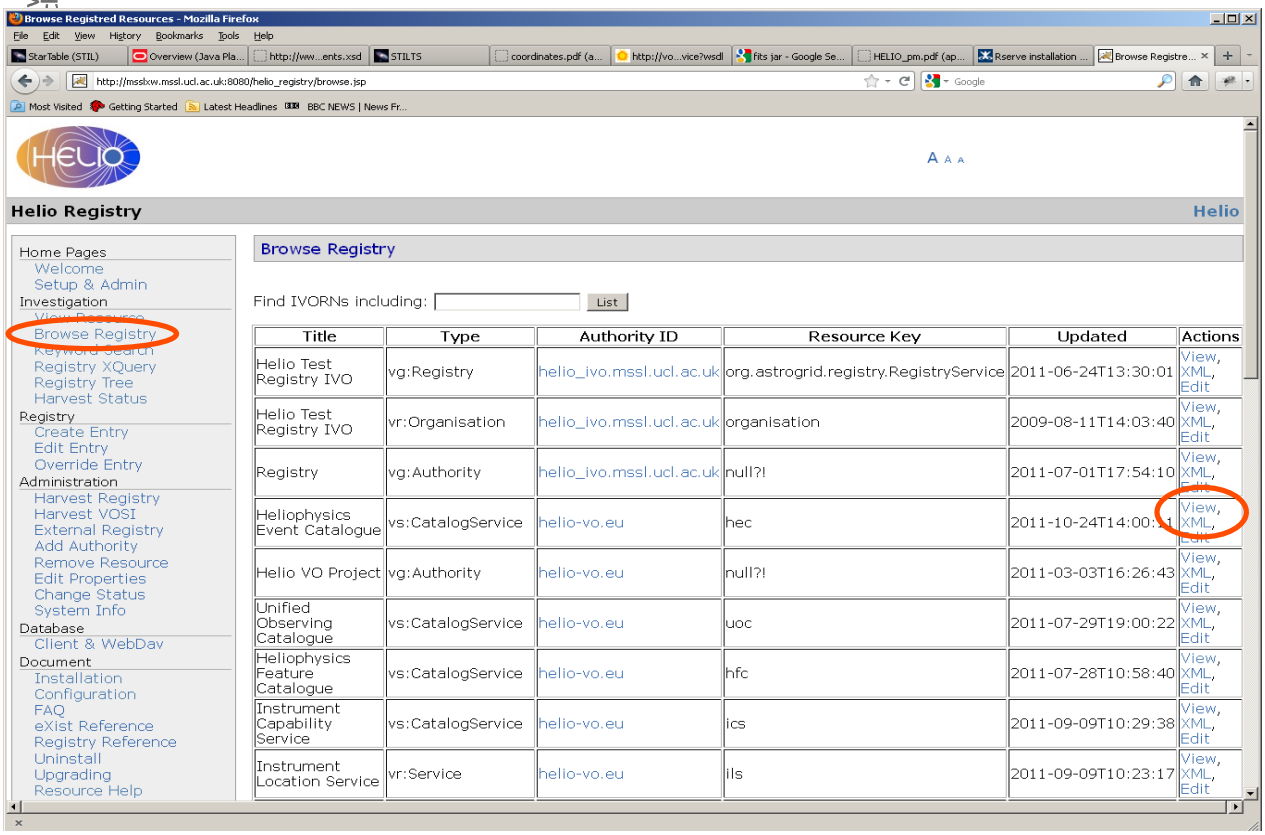
Simple REST service



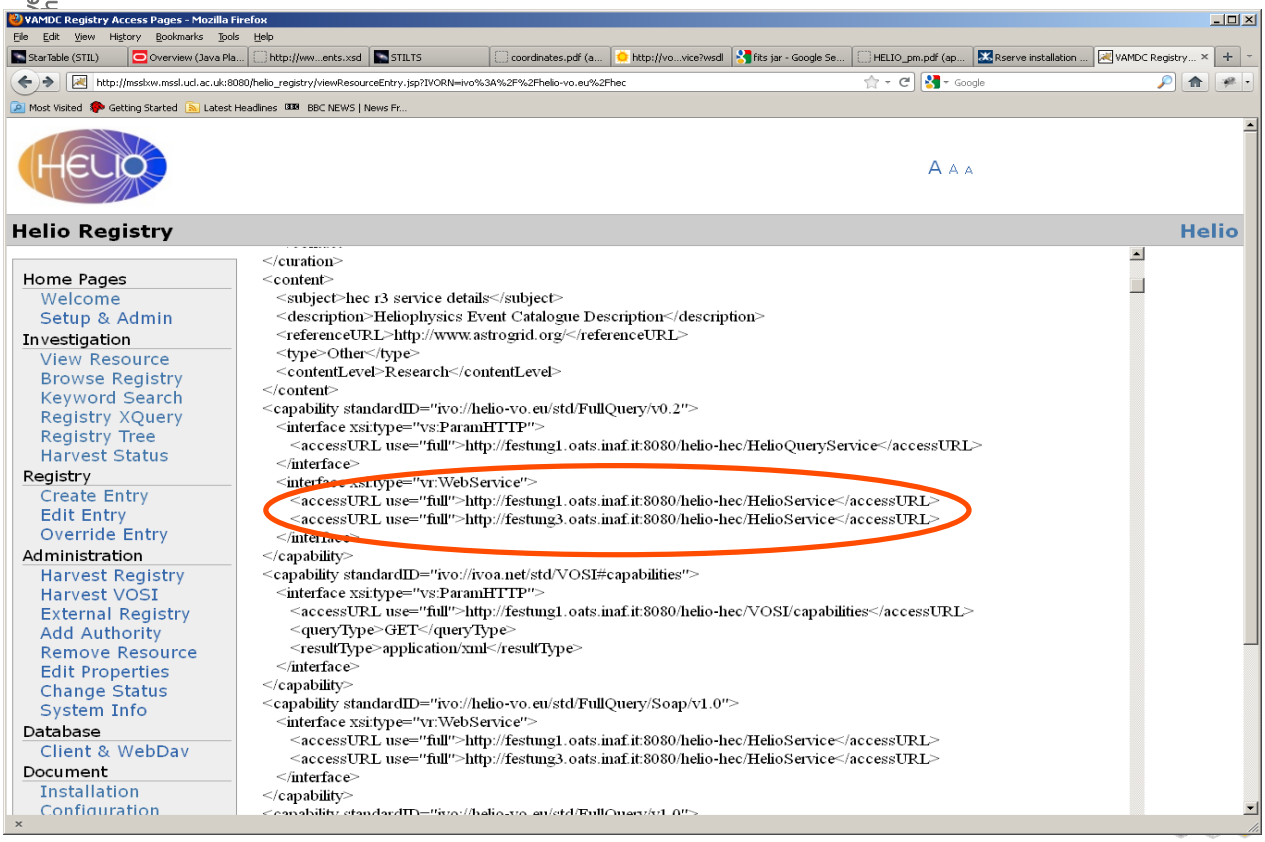


Asynchronous services

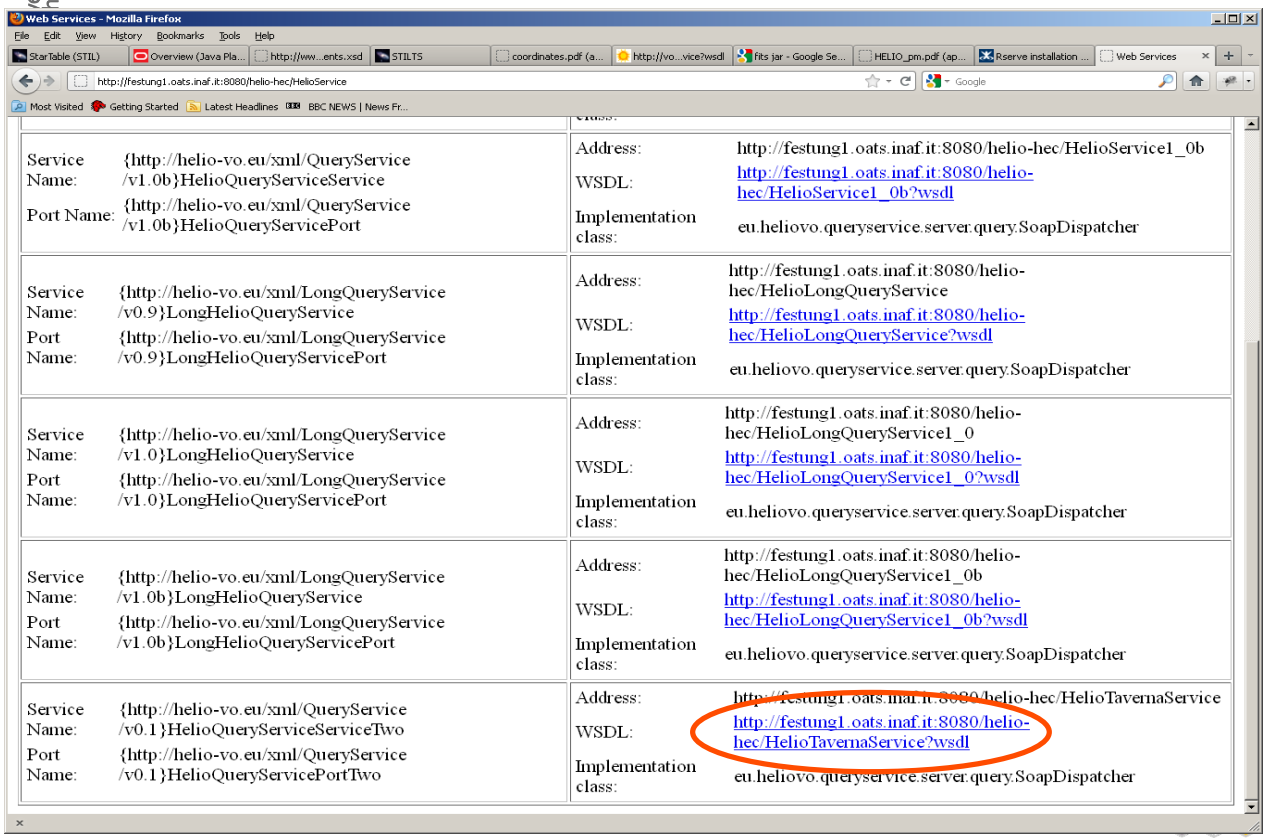
How to find HELIO services



How to find HELIO services cont …



How to find HELIO services



HELIO Services

Heliophysics Event Catalogue (HEC)	Maintain and provide access to *existing* 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 *newly-derived* parameter, etc.
Context Service (CXS)	Provide context information to help the user make a selection

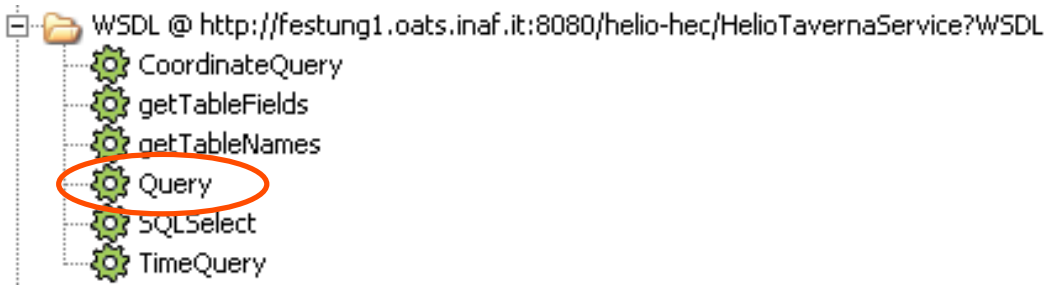
HELIO Services

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

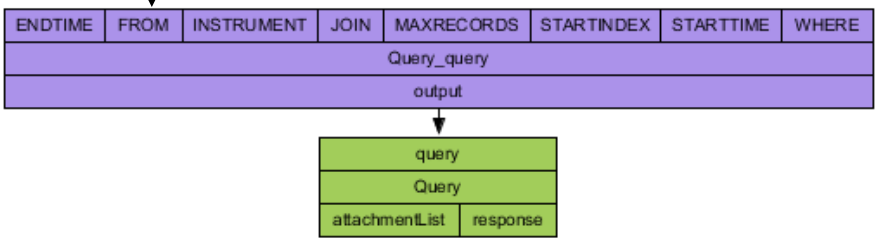
HELIO Services

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

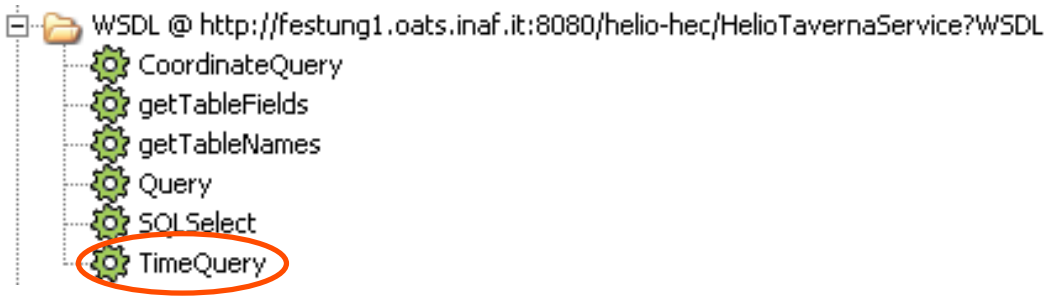


Only required field

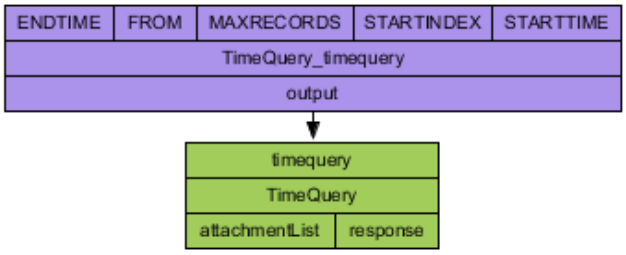


Details to Service Inputs

HELIO Query Interface

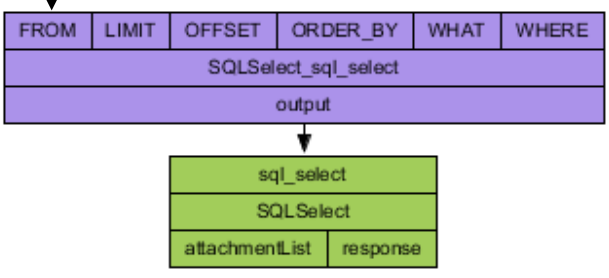


Only required field



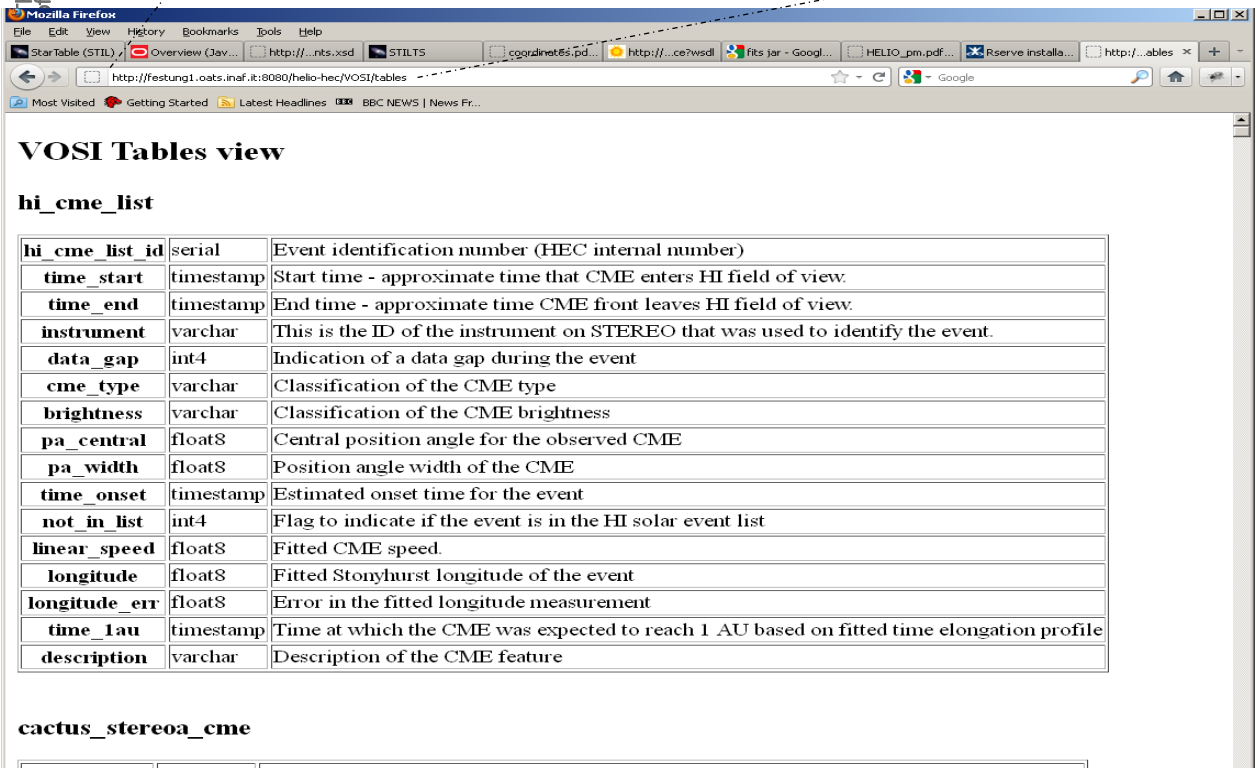
HELIO Query Interface

Only required field



How to find which ‘FROM’ to use

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



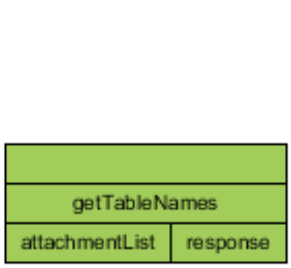
VOSI Tables view

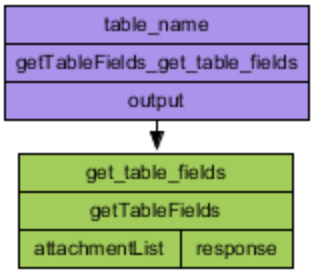
hi_cme_list

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

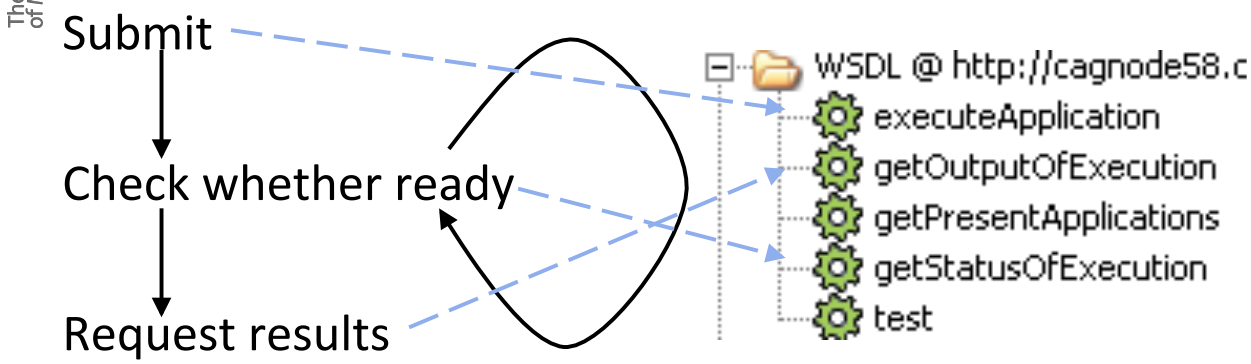
cactus_stereoa_cme

How to find which ‘FROM’ to use

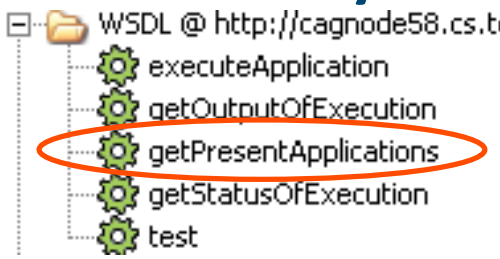


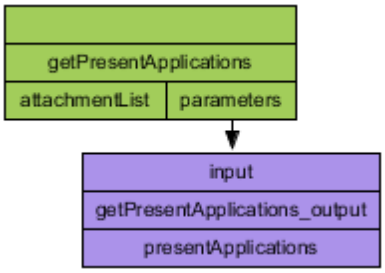


Asynchronous SOAP call - HPS

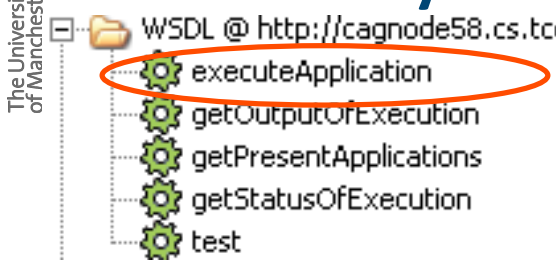


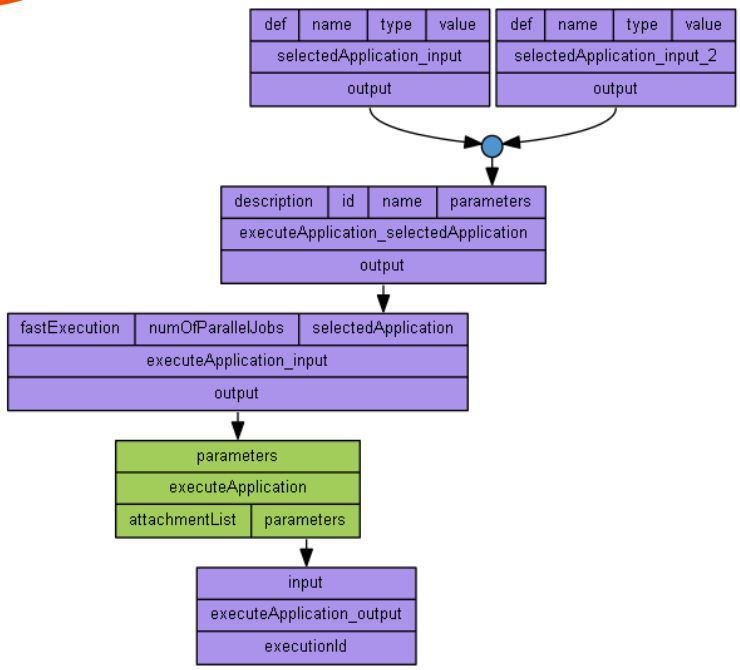
Asynchronous SOAP call - HPS



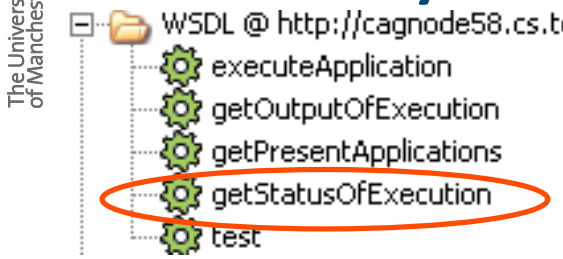


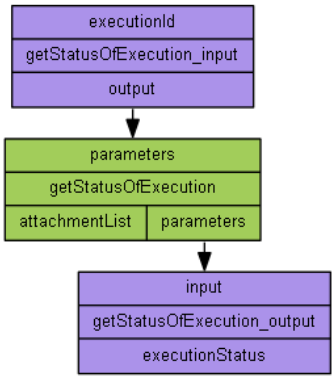
Asynchronous SOAP call - HPS



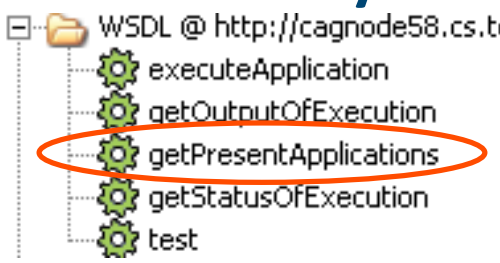


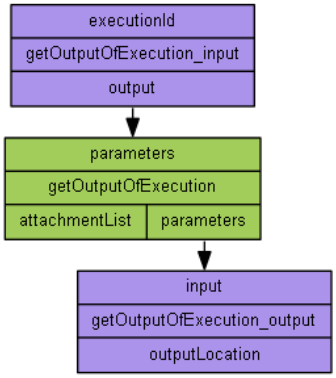
Asynchronous SOAP call - HPS





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

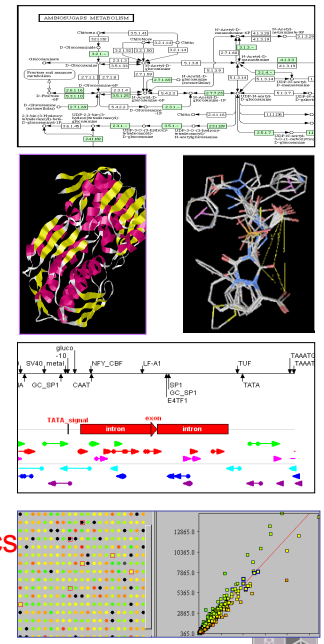
High throughput screening Proteomics Music

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

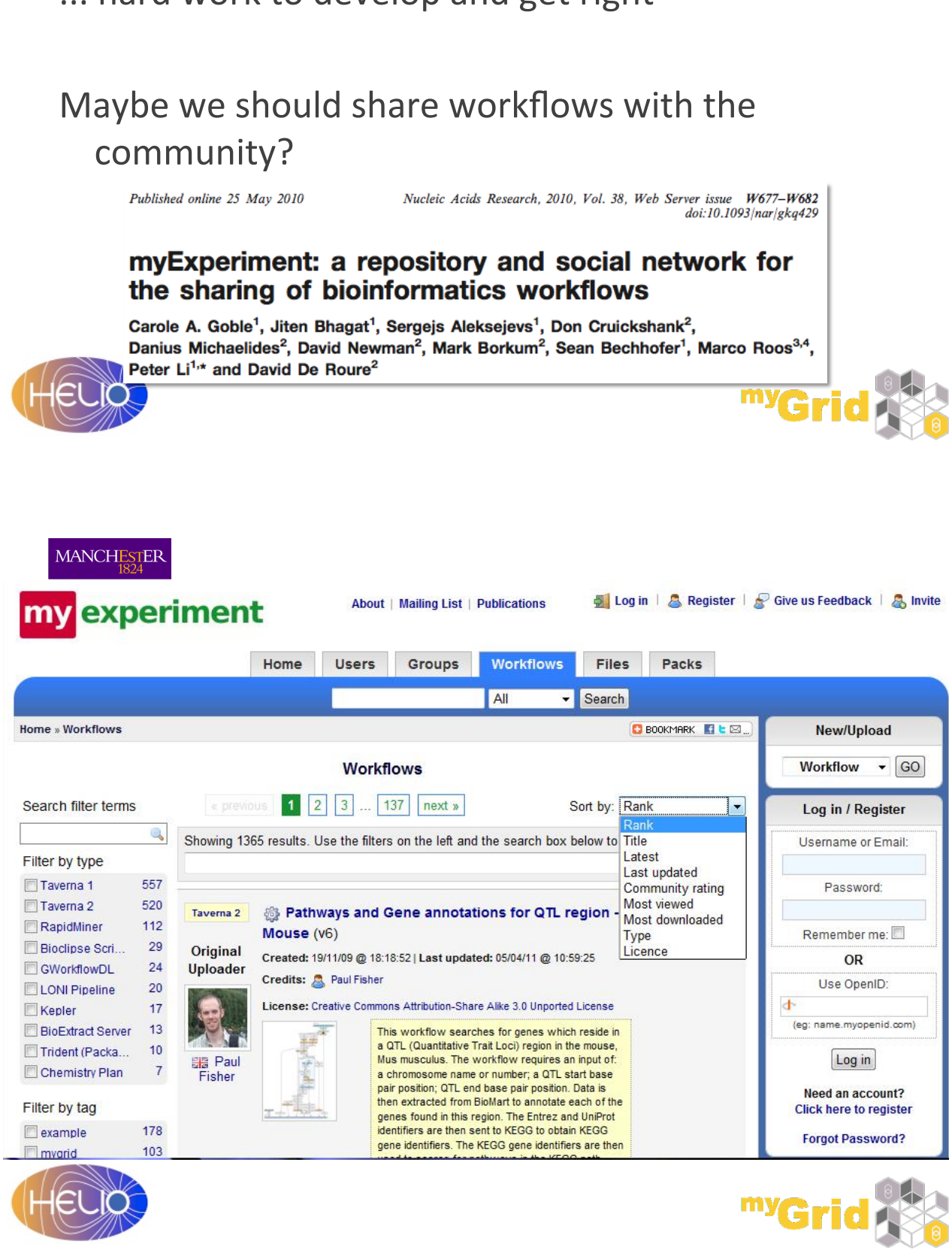
… hard work to develop and get right

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

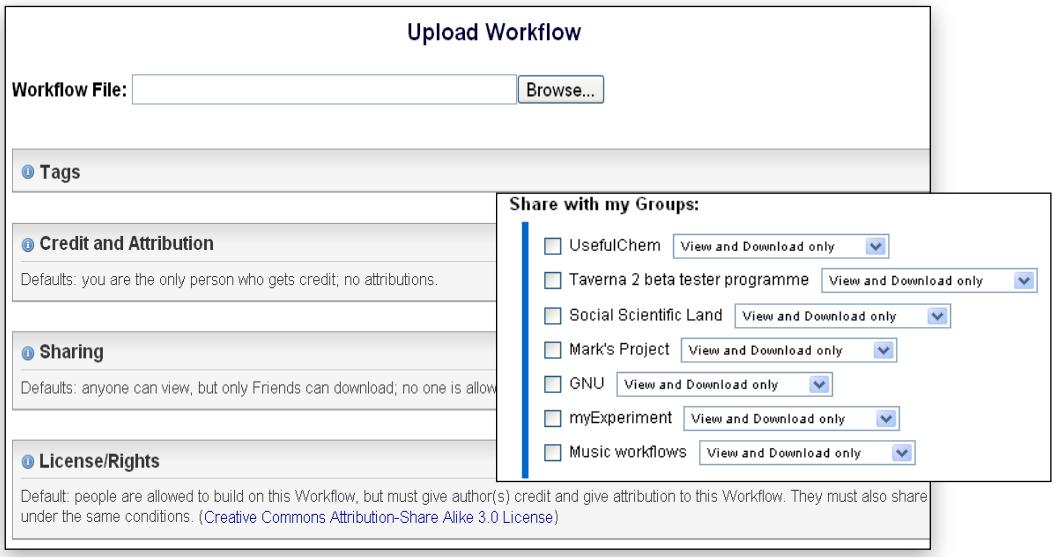
Carole A. Goble[1], Jiten Bhagat[1], Sergejs Aleksejevs[1], Don Cruickshank[2], Danius Michaelides[2], David Newman[2], Mark Borkum[2], Sean Bechhofer[1], Marco Roos[3,4], Peter Li[1,*] and David De Roure[2]



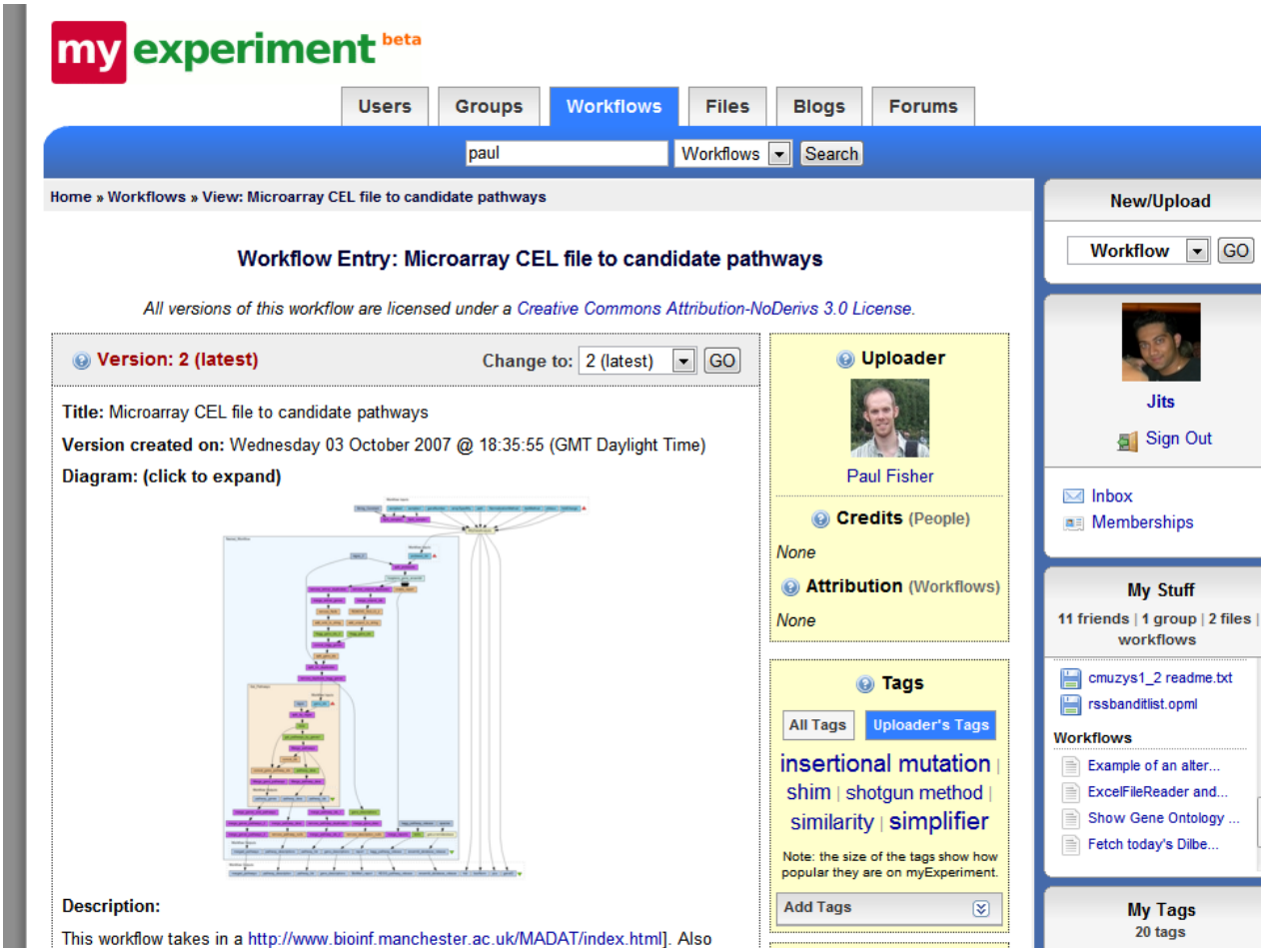
"Just Enough" Sharing….

- 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



The most important aspect of myExperiment - Designed by scientists



MANCHESTER 1824
The University of Manchester

Taverna on the cloud

amazon web services™

- 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

HELIO

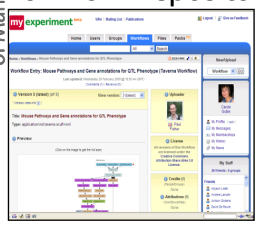
myGrid

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

myGrid Open Suite of Tools

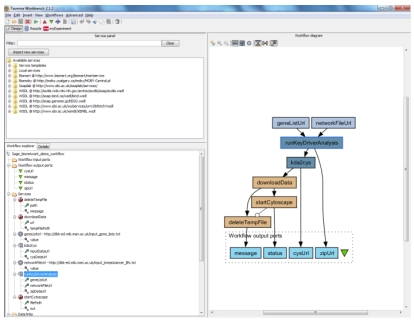
Workflow Repository



Service Catalogue

Activity and Service Plug-in Manager

Workflow GUI Workbench

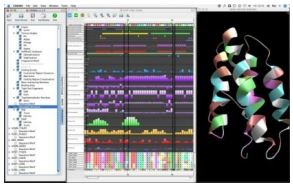


Provenance Store

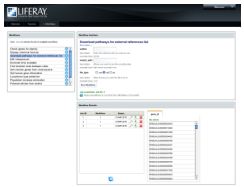
Workflow Server

Secure Service Access

Client User Interfaces



Third Party Tools

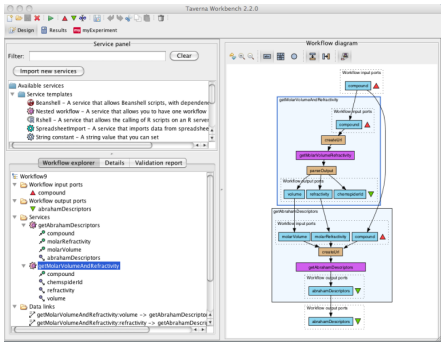


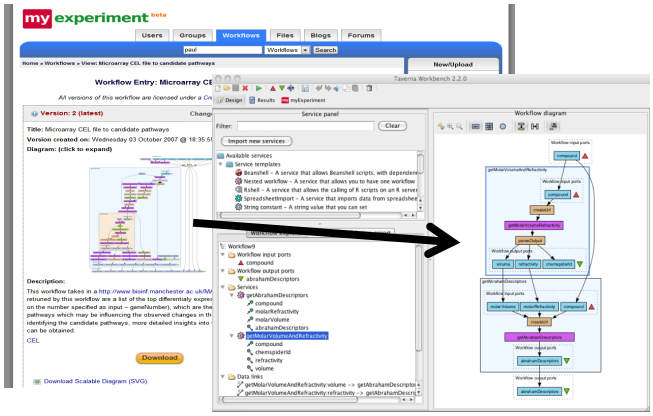
Web Portal

Programming and APIs

Spectrum of Users

Advanced users design and build workflows (informaticians)





Intermediate users reuse and modify existing workflows or components

Others "replay" workflows through web page

http://www.myexperiment.org

Load Data:

Run Workflow

Workflows and HELIO

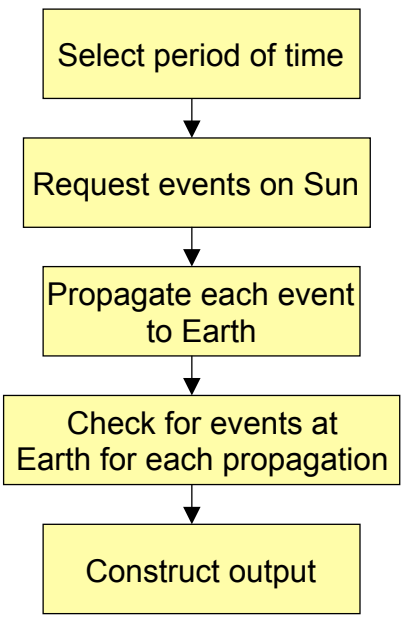
- Integral part of project
- Diverse use
  - Creation of pseudo services which can be integrated in Front End
  - Implementing complex use-case which can't be done in the Front End
  - Execute same use-case with changing parameters
  - Testing of Services
- All workflows developed shared with 'helio'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

More Information

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