



The University of Manchester

MANCHESTER 1824

## Dictionary Terms for Taverna

Dr. Georgina Moulton  
The University of Manchester  
([georgina.moulton@manchester.ac.uk](mailto:georgina.moulton@manchester.ac.uk))  
(on behalf of the myGRID team)



myGrid  

The University of Manchester

MANCHESTER 1824

## What you will learn

- No prior knowledge of workflow technology
- By the end you will know how to
  - install the workbench software, import and run existing workflows and build their own from components available on the public internet.
  - use the semantic search technologies in myGrid assist this process by enabling service discovery
  - do basic troubleshooting of workflows using Taverna's fault tolerance and debug mechanisms
  - manage the import and export of data to and from the workflow system.

myGrid  



The University of Manchester

MANCHESTER 1824

## T is for Taverna

Taverna enables the interoperation between databases and tools by providing a toolkit for composing, executing and managing workflow experiments

- Access to local and remote resources and analysis tools
- Automation of data flow
- Iteration over large data sets

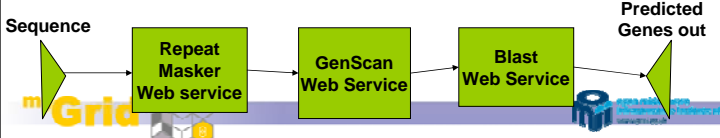
myGrid  



The University of Manchester

MANCHESTER 1824

## Workflows

- Workflow language specifies how processes (web services) fit together
- Describes *what* you want to do, not *how* you want to do it
- High level workflow diagram separated from any lower level coding – *you don't have to be a coder to build workflows*
- *Workflow is a kind of script or protocol that you configure when you run it.*
- Easier to explain, share, relocate, reuse and repurpose.
- Workflow  $\Leftrightarrow$  Model
- Workflow is the integrator of knowledge

Sequence 

myGrid  

The University of Manchester

MANCHESTER

## Two Types of Workflows

- Data workflows
  - A task is invoked once its **expected data** has been received, and when complete passes any resulting data downstream
- Control workflows
  - A task is invoked once its **dependant tasks** have completed

```

graph TD
  A --> B
  B --> C
  B --> D
  C --> E
  D --> E
  E --> F
  
```

myGrid

The University of Manchester

MANCHESTER

## Workflow Advantages

- Automation
  - Capturing processes in an explicit manner
  - Tedium! Computers don't get bored/distracted/hungry/impatient!
  - Saves repeated time and effort
- Modification, maintenance, substitution and personalisation
- Easy to share, explain, relocate, reuse and build
- Releases Scientists/Bioinformaticians to do other work
- Record
  - Provenance: what the data is like, where it came from, its quality
  - Management of data (LSID - Life Science Identifiers)

myGrid

The University of Manchester

myGrid

## It's more than plumbing....

- **Workflows are protocols and records.**
  - Explicit and precise descriptions of a scientific protocol
  - Scientific transparency. Easier to explain, share, relocate, reuse and repurpose and remember.
  - Provenance of results for credibility.
- **Workflows are know-how.**
  - Specialists create applications; experts design and set parameters; inexperienced punch above their weight with sophisticated protocols
- **Workflows are collaborations.**
  - Multi-disciplinary workflows promote even broader collaborations.

myGrid

The University of Manchester

MANCHESTER

## I is for the *in silico* experiment lifecycle

```

graph TD
  1((1)) --> 2((2))
  2 --> 3((3))
  3 --> 4((4))
  4 --> 5((5))
  5 --> 1
  
```

-Service-oriented middleware and tools that formalize and support the lifecycle:

- Service/Experiment Discovery } Feta
- Service Selection
- Service Composition } Taverna & Freeflow
- Service Execution & Execution Reporting
- Result Display
- Result Storage and Management } Provenance

Using Workflows is one way to make these experiments structured, shareable, repeatable and verifiable.

The University of Manchester

## Taverna Workflow Workbench

Workflow Object	Address	Delay	Back	Th
Compare Functions of genes on 1				
Workflow Input				
Workflow Output				
Graph				
Processes				
GetLineage0				
GetLineage1				
GetLineage2				
GetLineage3				
GetLineage4				
GetLineage5				
GetLineage6				
GetLineage7				
GetLineage8				
GetLineage9				
GetLineage10				
GetLineage11				
GetLineage12				
GetLineage13				
GetLineage14				
GetLineage15				
GetLineage16				
GetLineage17				
GetLineage18				
GetLineage19				
GetLineage20				
GetLineage21				
GetLineage22				
GetLineage23				
GetLineage24				
GetLineage25				
GetLineage26				
GetLineage27				
GetLineage28				
GetLineage29				
GetLineage30				
GetLineage31				
GetLineage32				
GetLineage33				
GetLineage34				
GetLineage35				
GetLineage36				
GetLineage37				
GetLineage38				
GetLineage39				
GetLineage40				
GetLineage41				
GetLineage42				
GetLineage43				
GetLineage44				
GetLineage45				
GetLineage46				
GetLineage47				
GetLineage48				
GetLineage49				
GetLineage50				
GetLineage51				
GetLineage52				
GetLineage53				
GetLineage54				
GetLineage55				
GetLineage56				
GetLineage57				
GetLineage58				
GetLineage59				
GetLineage60				
GetLineage61				
GetLineage62				
GetLineage63				
GetLineage64				
GetLineage65				
GetLineage66				
GetLineage67				
GetLineage68				
GetLineage69				
GetLineage70				
GetLineage71				
GetLineage72				
GetLineage73				
GetLineage74				
GetLineage75				
GetLineage76				
GetLineage77				
GetLineage78				
GetLineage79				
GetLineage80				
GetLineage81				
GetLineage82				
GetLineage83				
GetLineage84				
GetLineage85				
GetLineage86				
GetLineage87				
GetLineage88				
GetLineage89				
GetLineage90				
GetLineage91				
GetLineage92				
GetLineage93				
GetLineage94				
GetLineage95				
GetLineage96				
GetLineage97				
GetLineage98				
GetLineage99				

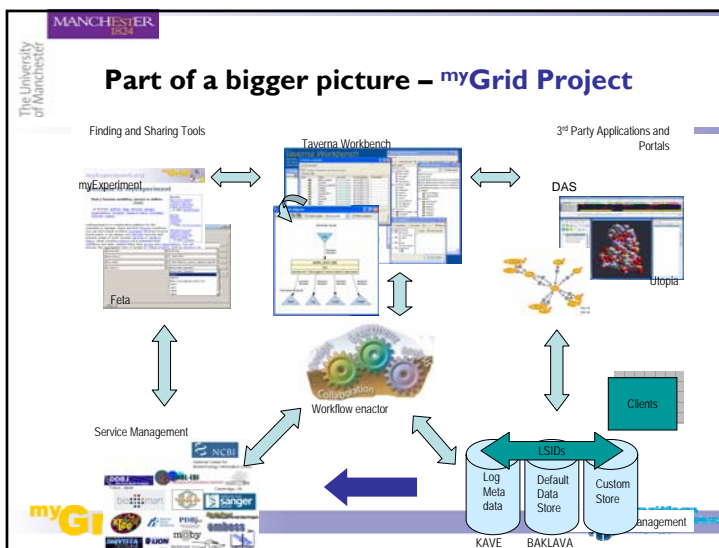
 The bottom of the window shows a 'Rendering done' status. Logos for myGrid and the University of Manchester are visible at the bottom of the slide."/>

The University of Manchester

## Taverna

- Taverna is :
  - A workflow language based on a dataflow model.
  - A graphical editing environment for that language.
  - An invocation system to run instances of that language on data supplied by a user of the system.
- When you download it you get all this rolled into a single piece of desktop software
- The enactor can be run independently of the GUI
- Java based, runs on Windows, Mac OS, Linux, Solaris ....
- It doesn't necessarily run "on a grid".
- Can be used to access resources, either on a grid, or anywhere else.

Logos for myGrid and the University of Manchester are visible at the bottom of the slide.

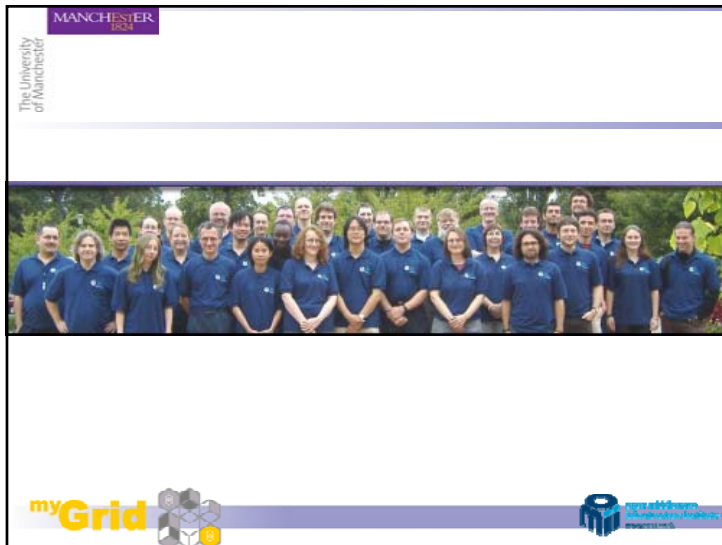


The University of Manchester

## OMII-UK

- Funded through the Open Middleware Infrastructure Institute (OMII-UK) as part of the myGrid project run by Carole Goble
- Four years old, funding secured through 2008 and beyond.
- Development team at Manchester & Hinxton, UK
- Wide group of 'friends and allies' across the world particularly within UK eScience
- Implemented in Java, released under LGPL licence.

Logos for myGrid and the University of Manchester are visible at the bottom of the slide.



## S is for Services and Service Providers

- Independent third party world-wide service providers of applications, tools and data sets. In the Cloud.
  - 850 databases, 166 web servers Nucleic Acids Research Jan 2006
- My local applications, tools and datasets. In the Enterprise. In the laboratory.
- Easily incorporate new service without coding. So even more services from the cloud and enterprise.

- 3500+ service operations
- All major providers
- Integration application for service providers like BioMOBY and BioMART

## Services

- Taverna can interoperate the following by default :
  - SOAP based web services
  - Biomart data warehouses
  - Soaplab wrapped command line tools
  - BioMoby services and object constructors (talk tomorrow)
    - Inline interpreted scripting (Java based)
- Other service classes can be added through an extension point (but you probably don't need to)

Type	Example	Description
String Constant	namespace	Local operation to emit a single constant string.
BioMoby	BlastFasta...	Service based on BioMoby, see <a href="http://www.biomoby.org">http://www.biomoby.org</a> .
Local Java	Create_moby_data	Local operation coded as a Java class, used for common or particularly generic functionality.
Soaplab	FormatSequence	Service wrapper around a legacy command line tool such as, in this case, seqret from the EMBOSS tool set.
Workflow	FixGeneOntologyID	A nested workflow exposed as a single operation.
Web Service	createSession	A processor accessing a standard SOAP service.
Beanshell	FetchOntoGlyph	User editable scripting operation, in this instance using interpreted Java.
Biomart	<i>Not present in diagram</i>	Configurable parameterized query over a Biomart data warehouse dataset. More information is available at <a href="http://www.ebi.ac.uk/biomart">http://www.ebi.ac.uk/biomart</a> .
Styx Grid Service	<i>Not present in diagram</i>	Processor based on the Styx Grid Service framework developed at Reading eScience Centre, link at <a href="http://www.resc.rdg.ac.uk/projects.php">http://www.resc.rdg.ac.uk/projects.php</a>
API Consumer	<i>Not present in diagram</i>	Allows import of Java APIs as workflow components, processors correspond to constructors and static or instance methods.
Interaction Service	<i>Not present in diagram</i>	Allows asynchronous interaction with remote users in a configurable fashion

The University of Manchester

MANCHESTER

Available services

Workflow diagram

Biomart query

Soaplab operation wrapping an EMBOSS tool

Advanced Model Explorer:  
Tree view of workflow structure

Version 1.5.1 Shown running on a Mac but written in Java,  
Runs & developed on Windows, OS X and Linux.

The University of Manchester

MANCHESTER

## Introduced Terms – A Test??

- Advanced Model Explorer
- Available services
- Biomoby
- Biomart
- myGrid Project
- *in silico* experimental lifecycle
- OMII-UK
- Soap
- Workbench
- Workflow
- Web services

myGrid