MANCHESTER 1824

The University of Manchester

The Taverna Workbench: Integrating and analysing biological and clinical data with computerised workflows

Dr Katy Wolstencroft

myGrid

University of Manchester

Vrije Universiteit, Amsterdam

myGrid

MANCHESTER 1824

The University of Manchester

Outline

- Why workflows are important
- WSDL, REST and other Workflow Services
- Getting started with Taverna
- Taverna in Use
- Sharing and reusing workflows
- Workflows on servers, grids and clouds
- Taverna Future Plans

myGrid

MANCHESTER 1824

The University of Manchester

www.taverna.org.uk

Download, unpack and run

myGrid

Automation

- 21st century is the century of information
- eGovernment
- World bank data
- Climate change data
- Large scale physics
  - Large Hadron collider
  - Astronomy
- ‘Omics data
- Next Gen Sequencing

Where is the data?

- In repositories run by major service providers (e.g. NCBI, EBI)
- Group/Institute web sites
- On ftp servers
- In local project stores

- Few defined formats
- Inconsistent metadata

NCBI
National Center for Biotechnology Information (USA)

DDBJ DNA Data Bank of Japan
Tokyo, Japan

EMBL-EBI European Bioinformatics Institute
Cambridge, UK

biomart

SoapLab

wellcome trust sanger institute

KEGG Kyoto Encyclopedia of Genes and Genomes

MOUNT SINAI HOSPITAL

PDBjXP Protein Data Bank Japan Experimetal Version

PathPort The Pathogen Portal Web Project

emboss

moby

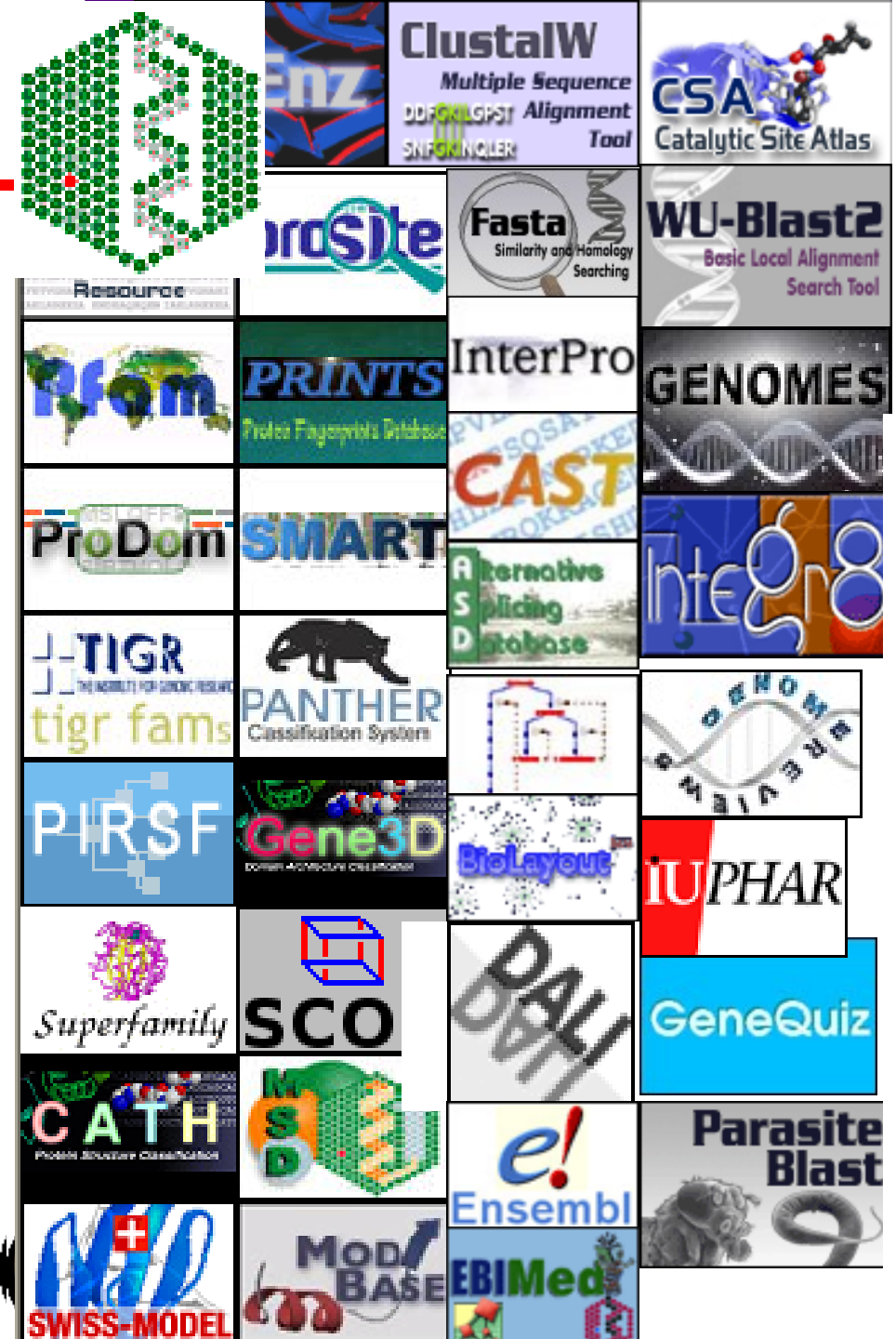
SeqVISTA

LION
SRS

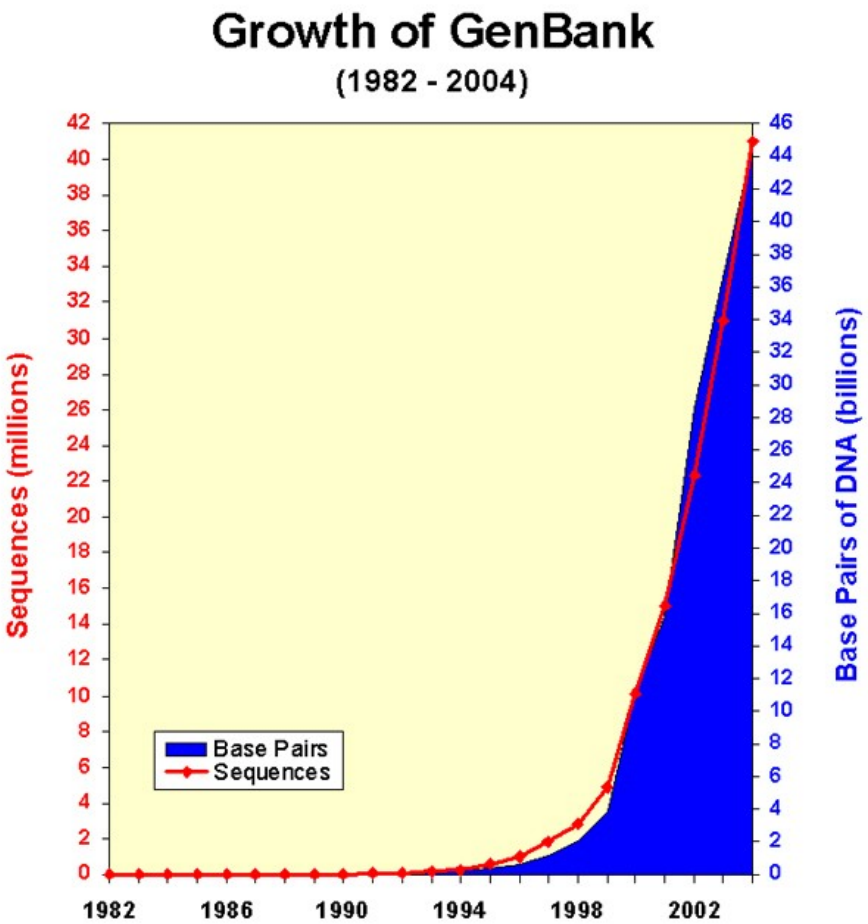
SeqHound

caBIO

Lots of Resources



NAR 2012 – 1500 databases

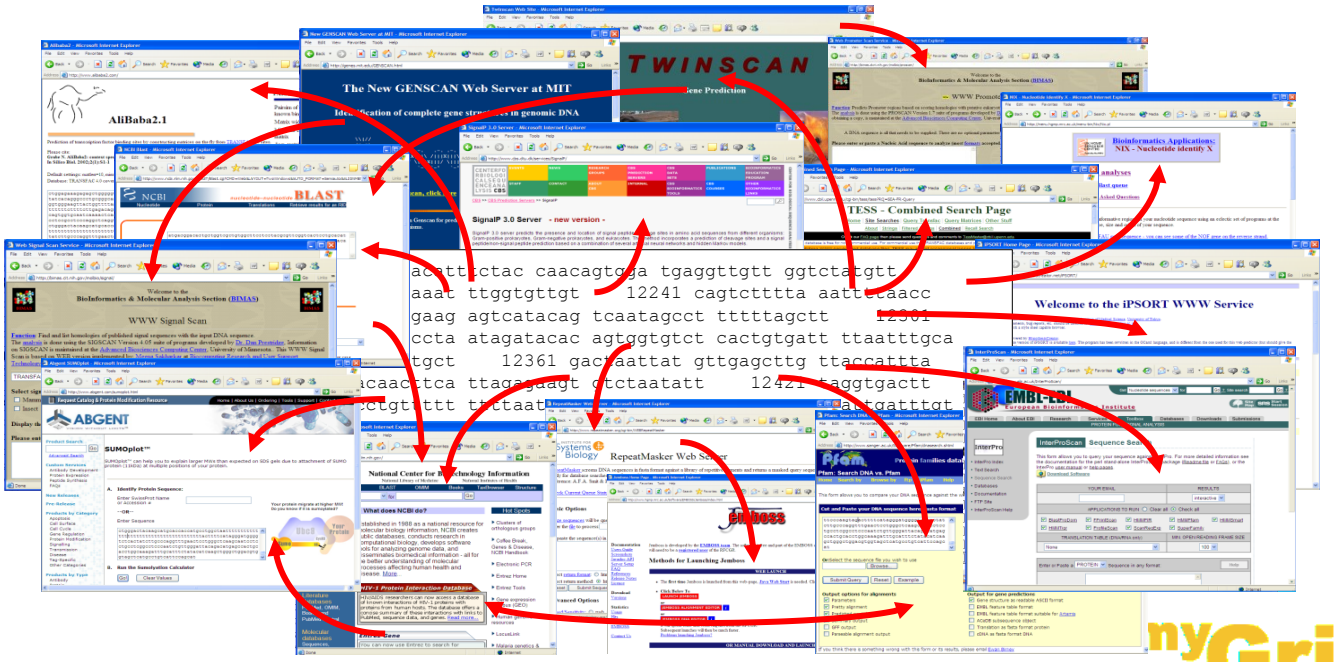


Distribution

- Data resources – databases, analysis tools
- Computational power – servers, clusters, cloud/grid
- Researchers and collaborators – skills and expertise need to be shared and exchanged
  Analysis scripts need to be shared and exchanged

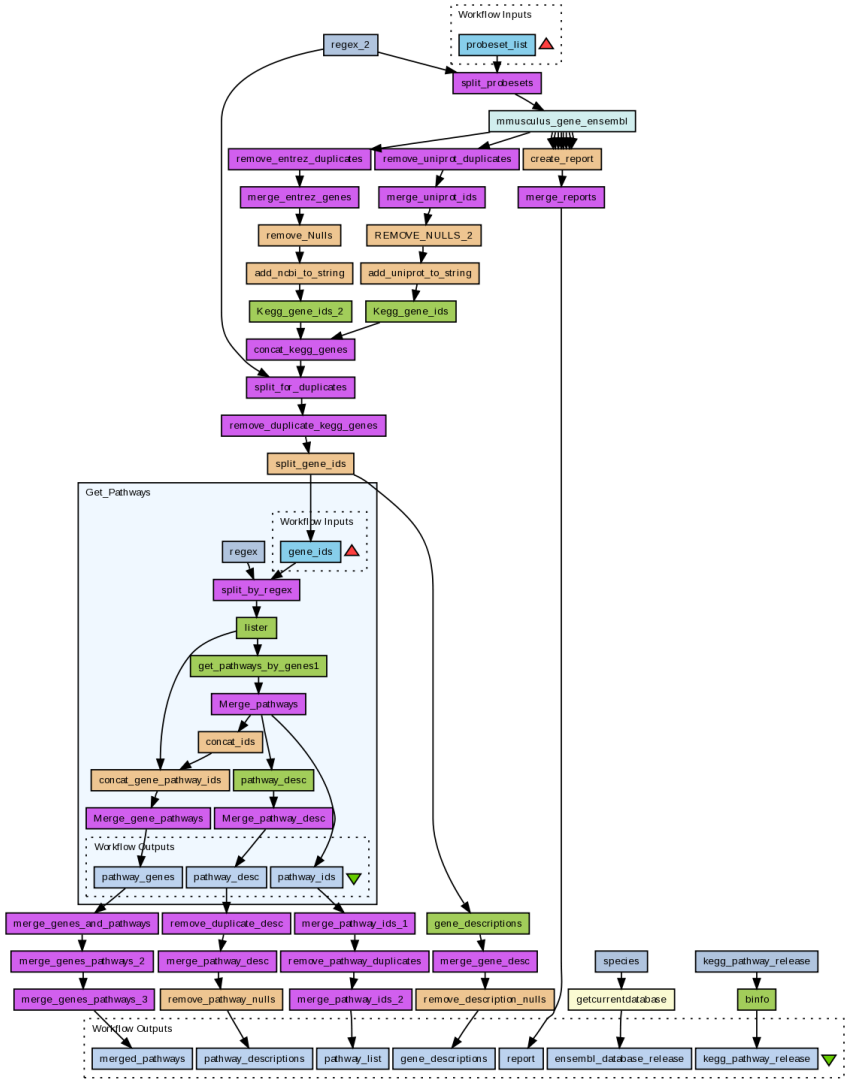
What that means for Bioinformatics

- Sequential use of distributed tools
- Incompatible input and output formats
- Analysis of large data sets by multiple researchers
- Difficult to record parameter selections
- Difficult to reproduce analyses



Workflow as a Solution

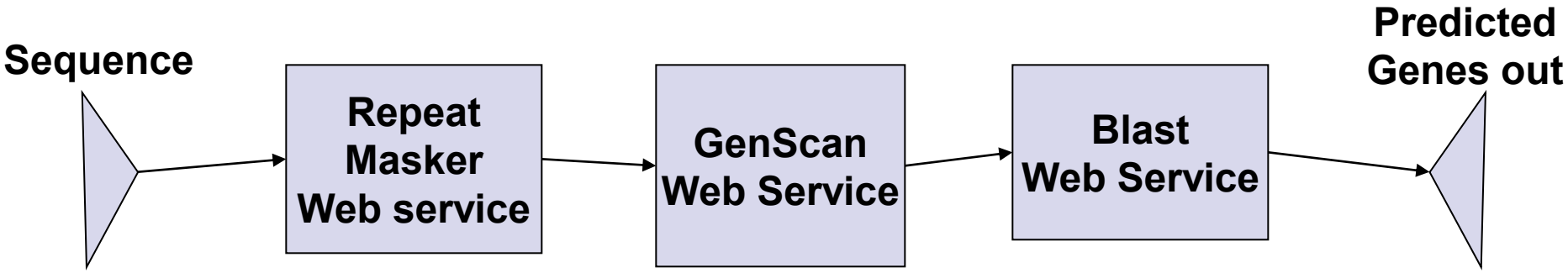
- Automating the process
- Sophisticated analysis pipelines
- A set of **services** to analyse or manage data (either local or remote)
- Data flow through services
- Control of service invocation
- Iteration



What is a Workflow?

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

Simple language specifies how processes fit together



Workflows are ideal for…

- High throughput analysis
  - Transcriptomics, proteomics, next gen sequencing
- Data integration, data interoperation
- Data management
  - Model construction
  - Data format manipulation
  - Database population
  - Semantic integration
  - Visualisation

Promoting Reproducible Research

Informatics involves

- Complex, multi-step analyses
- Lots of data as inputs
- Lots of data generated
- Workflows encapsulate the methods and parameters
- Workflows allow you to visualise the methods

MANCHESTER 1824

The University of Manchester

Preventing Irreproducible Research

- An array of errors

http://www.economist.com/node/21528593

- Duke University, 2006 -Prediction of the course of a patient's lung cancer using expression arrays and recommendations on different chemotherapies from cell cultures
- 3 different groups could not reproduce the results and uncovered mistakes in the original work

myGrid

If the Analyses were done using Workflows.....

- Reviewers could re-run experiments and see results for themselves
- Methods could be properly examined and criticised
- Mistakes could be pinpointed

MANCHESTER 1824

The University of Manchester

Workflows are …

… records and protocols (i.e. your *in silico* experimental method)

… know-how and intellectual property

… hard work to develop and get right

…..re-usable methods (i.e. you can build on the work of others)

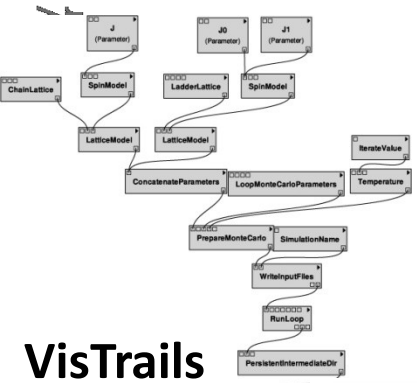
So why not share and re-use them

myexperiment

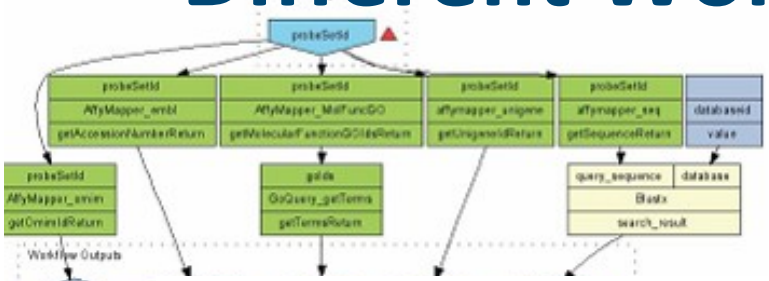
myGrid

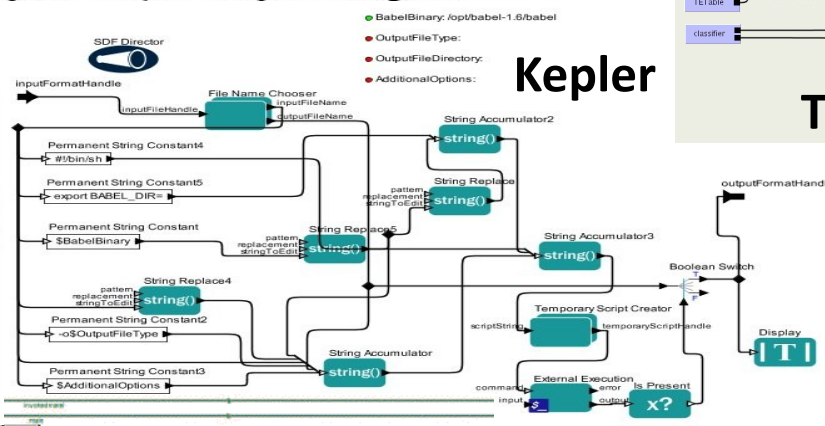
WORKFLOW SYSTEMS

Different Workflow Systems

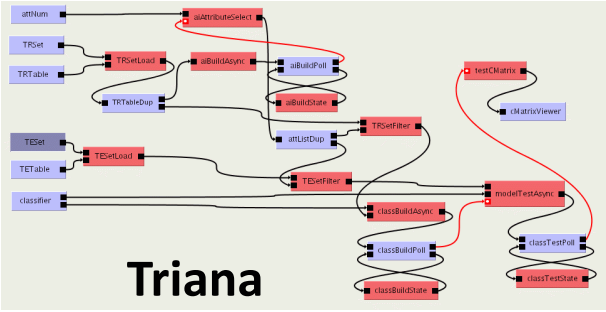


VisTrails

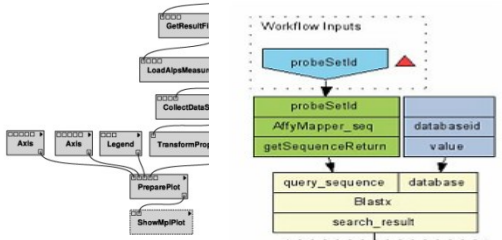


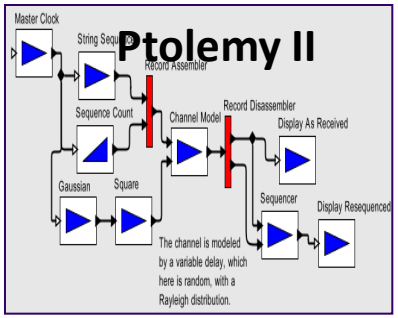


Kepler

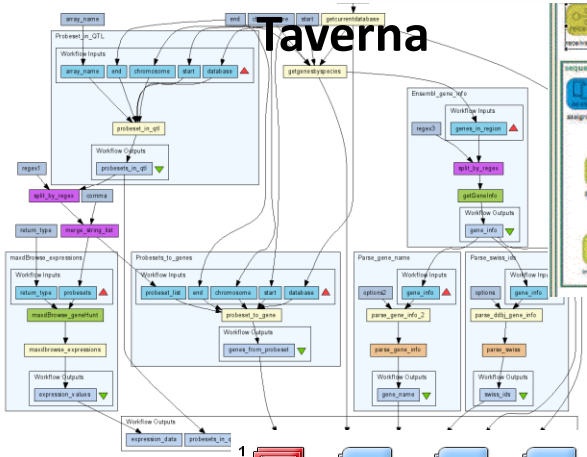


Triana

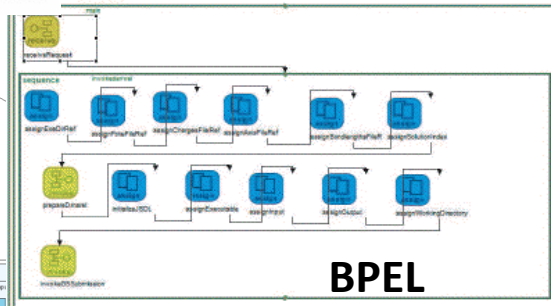




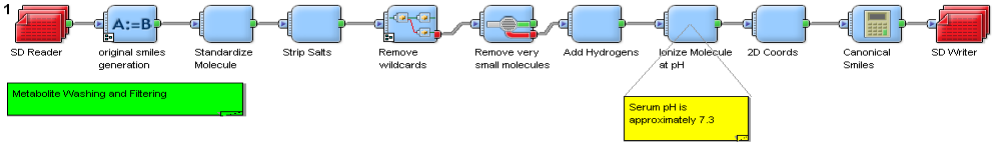
Ptolemy II



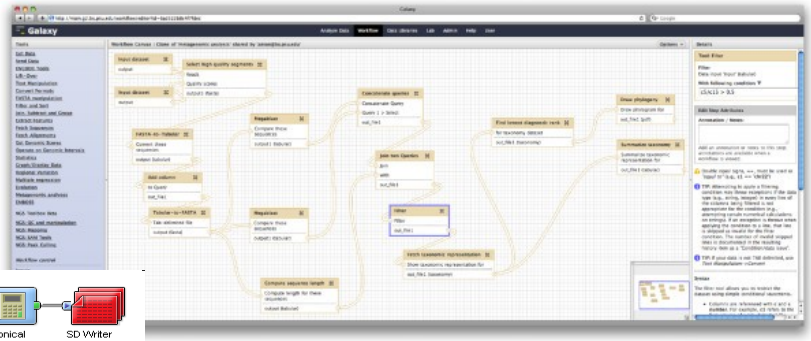
Taverna



BPEL

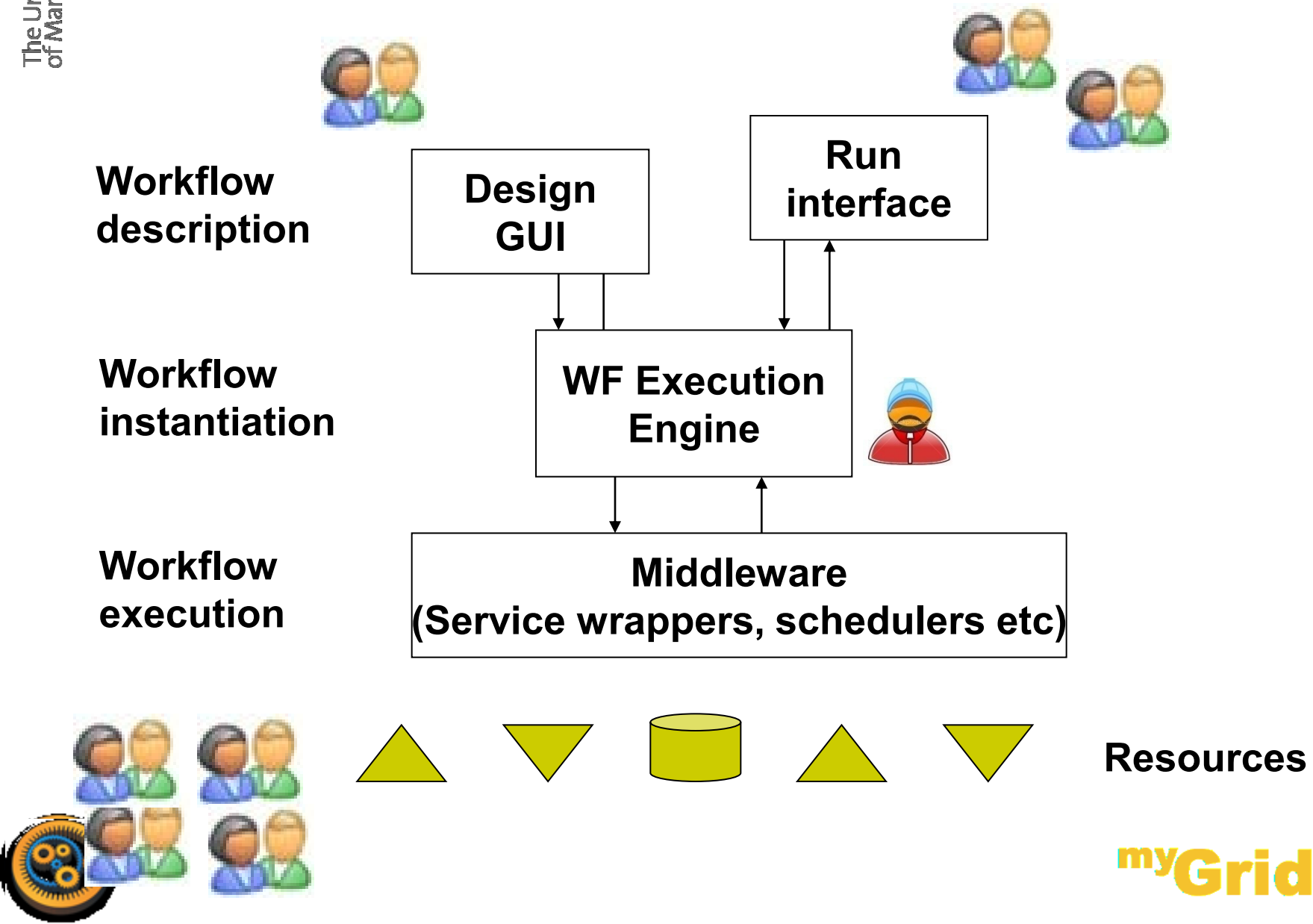


Pipeline Pilot



Galaxy

All Workflow Systems at 50,000 feet



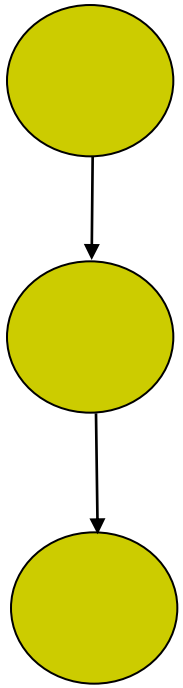
Different Types of Workflows

- Sequences of concatenated steps
- Two types of workflows:
  - Data workflows

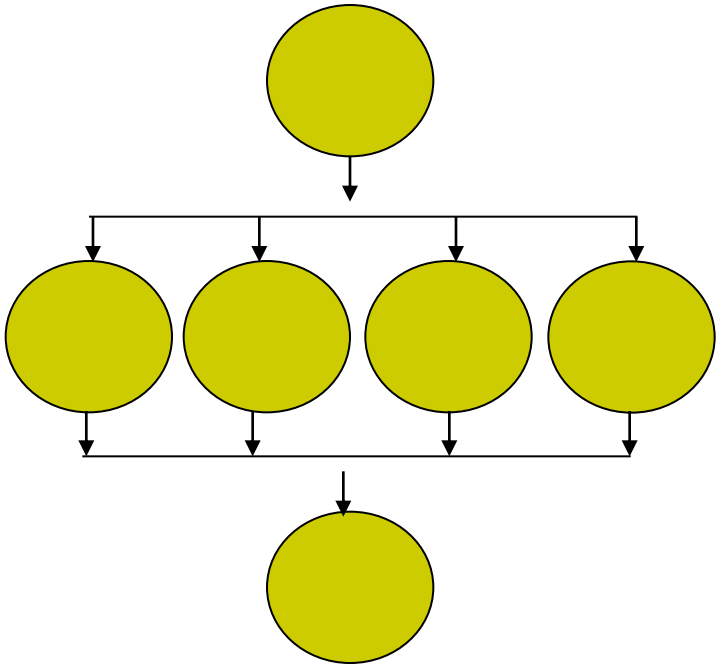
    A task is invoked once its **expected data** has been received. When complete, it passes any resulting data downstream
  - Control workflows

    A task is invoked once its **dependant tasks** have been completed

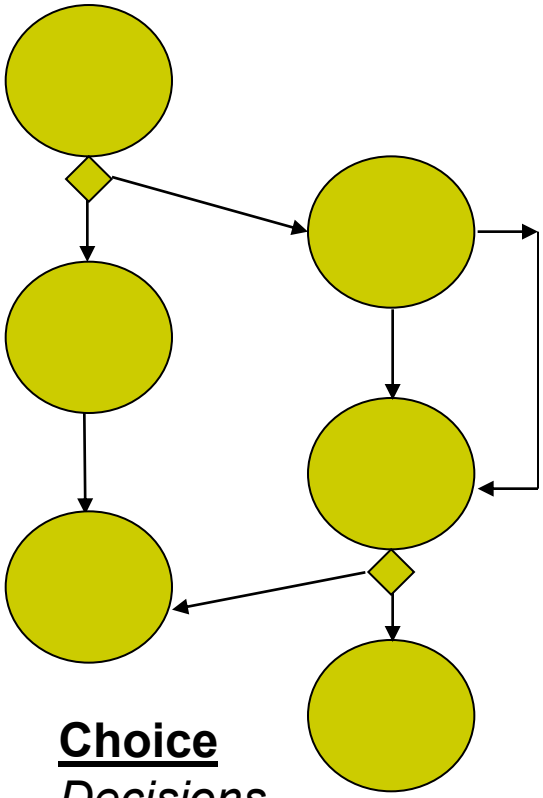
Possible Workflow Structures



Sequence
Store intermediate results



Parallel
Apply multiple components to a set of data



Choice
Decisions at runtime

Iteration
Loop through datasets

MANCHESTER 1824

The University of Manchester

Taverna Workbench

http://www.taverna.org.uk/

Freely available
open source
Current Version 2.4

80,000+ downloads
across version

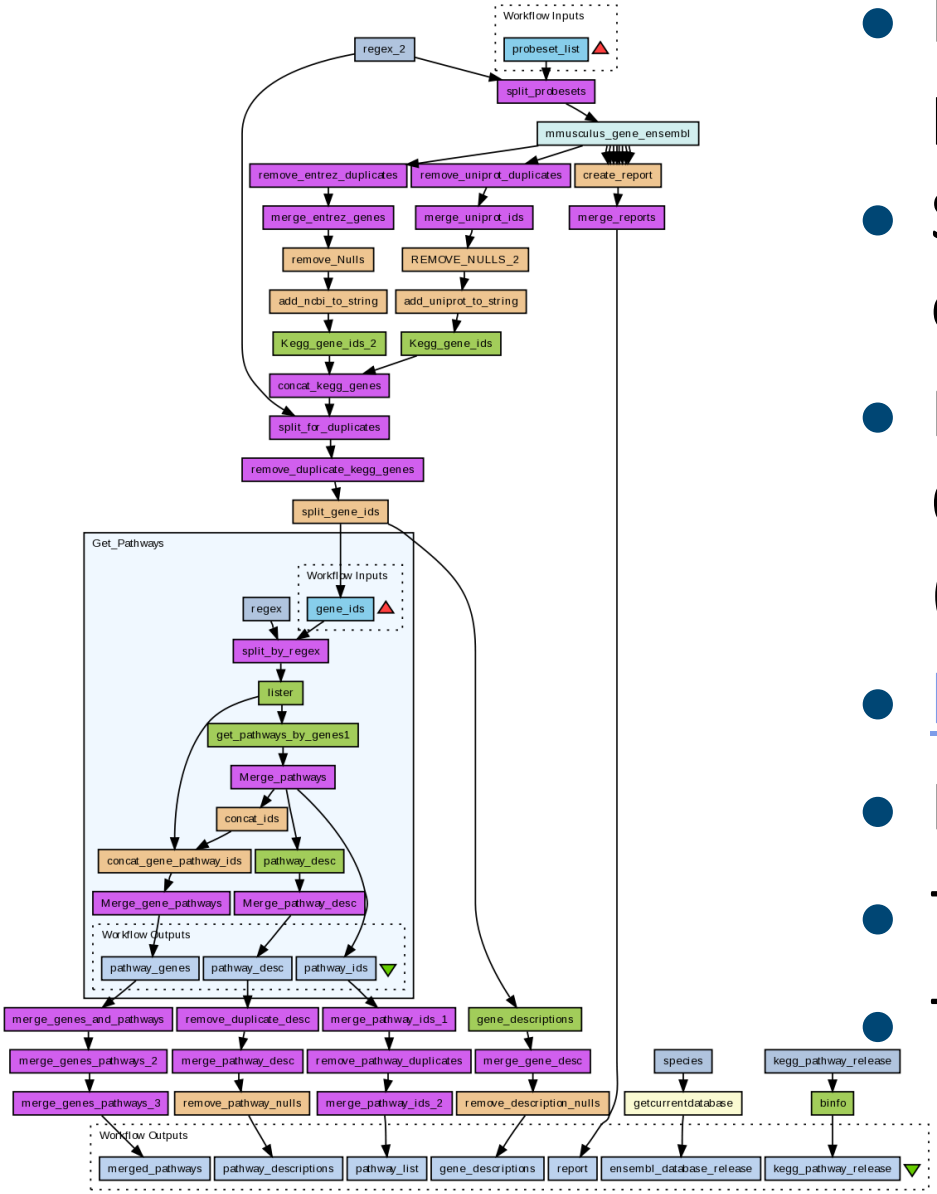
Part of the myGrid Toolkit

Windows/Mac OS X/
Linux/unix

Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W729-32.
Taverna: a tool for building and running workflows of services.
Hull D, Wolstencroft K, Stevens R, Goble C, Pocock MR, Li P, Oinn T.

myGrid

Taverna Workflows



- Part of UK E-Science myGrid project
- Started in 2001, collaboration across UK
- Now: Manchester (Goble), Oxford/Southampton (DeRoure)
- http://www.taverna.org.uk
- Local Taverna desktop
- Taverna Server
- Taverna on the cloud

MANCHESTER 1824

The University of Manchester

Open source, open development

- Taverna suite of tools are all **open source,** free to use and **customise**
- Large user **community**, active mailing lists
- Lead developers: **myGrid** in Manchester UK
- **Contributors** from across the world
- **Plugins** developed and shared by contributors
  - XPath, REST, R, BioCatalogue, PBS, SADI, External Tools (UseCase), UNICORE, CDK, Opal, caGrid, XWS, gLite

myGrid

Taverna Workbench

Taverna Workbench 2.2.0

Design | Results | myExperiment

Service panel

Clear

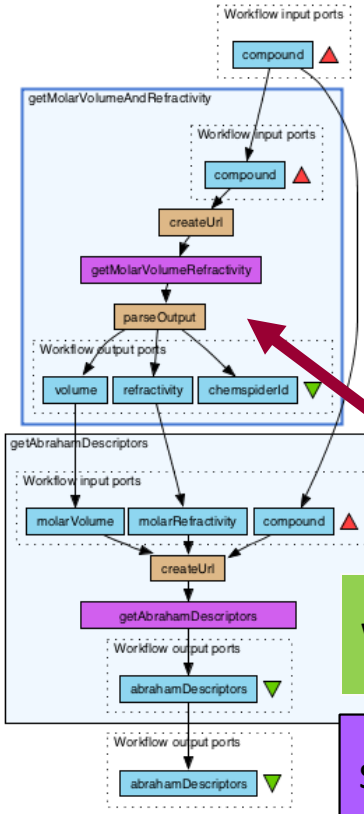
List of services

Available services
- Service templates
  - Beanshell – A service that allows Beanshell scripts, with dependenc
  - Nested workflow – A service that allows you to have one workflow
  - Rshell – A service that allows the calling of R scripts on an R server
  - SpreadsheetImport – A service that imports data from spreadshee
  - String constant – A string value that you can set

Workflow explorer | Details | Validation report

- Workflow9
  - Workflow input ports
    - compound
  - Workflow output ports
    - abrahamDescriptors
  - Services
    - getAbrahamDescriptors
      - compound
      - molarRefractivity
      - molarVolume
      - abrahamDescriptors
    - getMolarVolumeAndRefractivity
      - compound
      - chemspiderId
      - refractivity
      - volume
  - Data links
    - getMolarVolumeAndRefractivity:volume -> getAbrahamDescriptor
    - getMolarVolumeAndRefractivity:refractivity -> getAbrahamDescri

Workflow diagram



Workflow engine to run workflows

Construct and visualise workflows

Web Services — e.g. KEGG

Scripts — e.g. beanshell, R

Programming libraries — e.g. libSBML

Workflows and the in Silico Life Cycle

Create and run workflows

Taverna

Workflows and the in Silico Life Cycle

BioCatalogue

Discover, understand and assess services

Create and run workflows

Taverna

Workflows and the in Silico Life Cycle

BioCatalogue

Discover, understand and assess services

Discover, reuse and share workflows

myexperiment

Create and run workflows

Taverna

myGrid

MANCHESTER 1824

The University of Manchester

Workflows and the in Silico Life Cycle

BioCatalogue

Discover, understand and assess services

Discover, reuse and share workflows

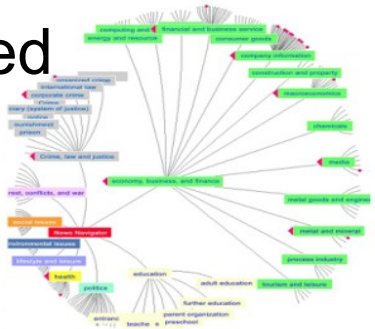
myexperiment

Create and run workflows

Taverna

Manage the metadata needed and generated

RDF, OWL



myGrid

SERVICES IN WORKFLOWS

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

Therefore, you can automatically connect to and use remote services from your computer in an automated way

Web Services – Brief Glossary

- WSDL (Web Service Definition Language)
  - A machine-readable description of the operations supported
- SOAP (Simple Object Access Protocol)
  - An xml protocol for passing messages
- REST (Representational State Transfer)
  - An alternative interface to SOAP

Using Remote Tools and Services with Taverna

- Web Services
  - WSDL
  - REST
- Grid Services
- Local services
- Beanshell (small, local scripts)
- Secure Services
- Workflows
- BioMart
- R-processor
- And more.....

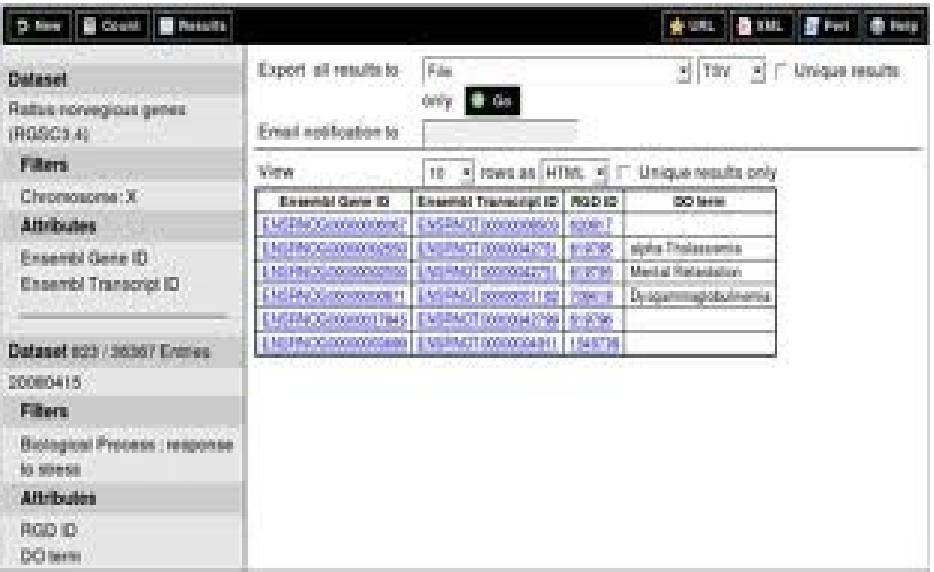
Specialist services

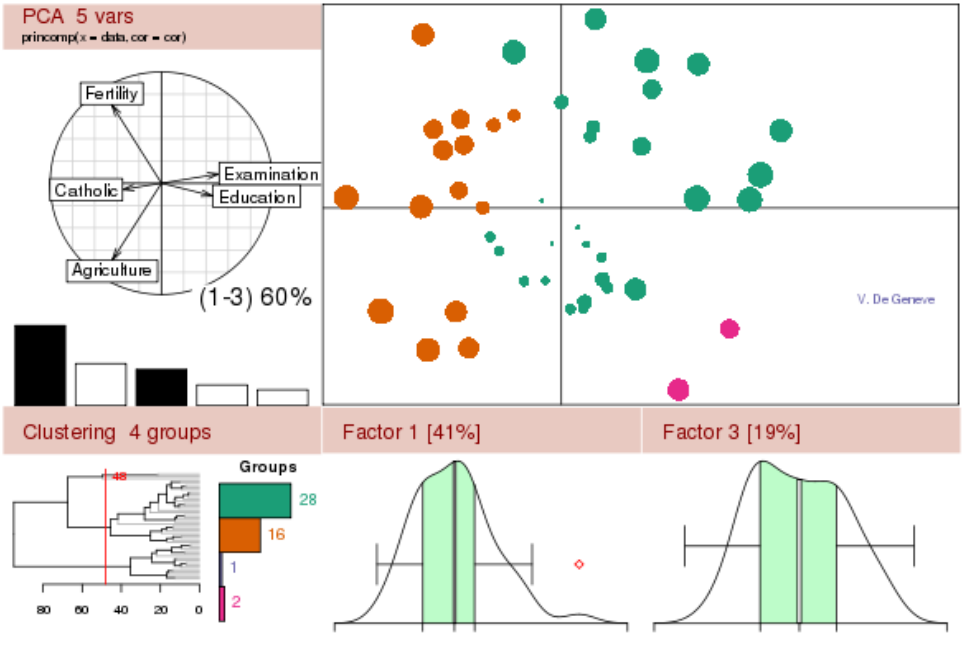
BioMart Queries

- Federated database system that provides unified access to distributed data sources
- Ensembl, Pride.....

R-scripts

- R is a free software environment for statistical computing and graphics





Different Approaches to Service Connections

- Open – connect to ANY service regardless of type and structure
  - More services, but more heterogeneity
  - Easy to add new services
  - Taverna, Kepler
- Closed – connect to services designed specifically to work together,
  - Less heterogeneity, but fewer services
  - Harder to add new services
  - Galaxy server, Knime

Who Provides the Services?

Open domain services and resources

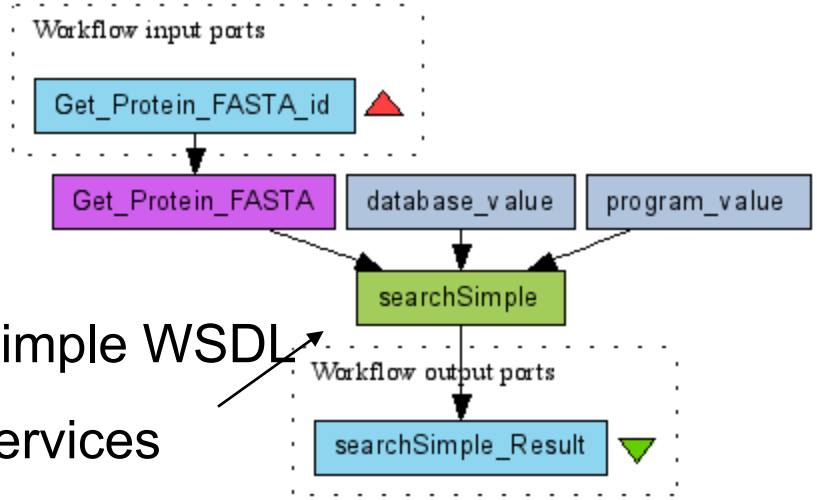
- Taverna accesses thousands of services
- Third party – we don't own them – we didn't build them
- All the major providers
  - NCBI, DDBJ, EBI …
- Enforce NO common data model.

National Center for Biotechnology Information (USA)

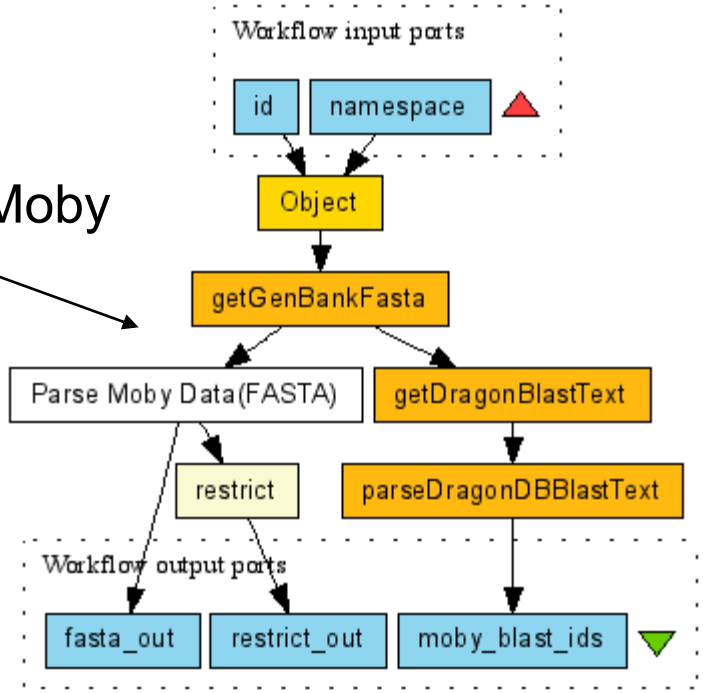
Tokyo, Japan

Cambridge, UK

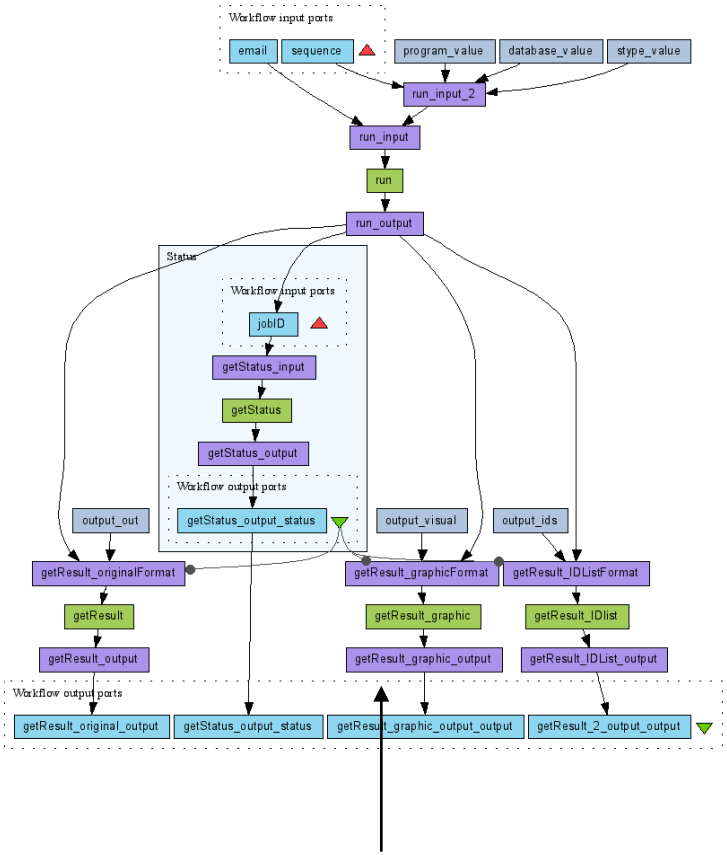
How do you use the services?



Simple WSDL services



SADI / BioMoby 'Semantic' Services



Asynchronous services

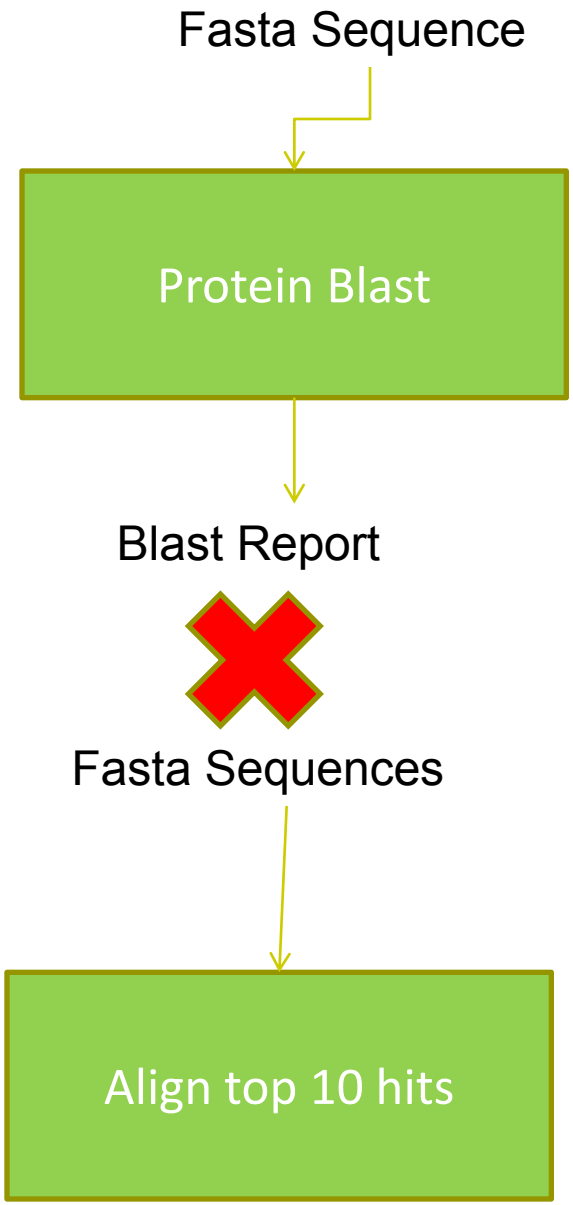
Managing Heterogeneities

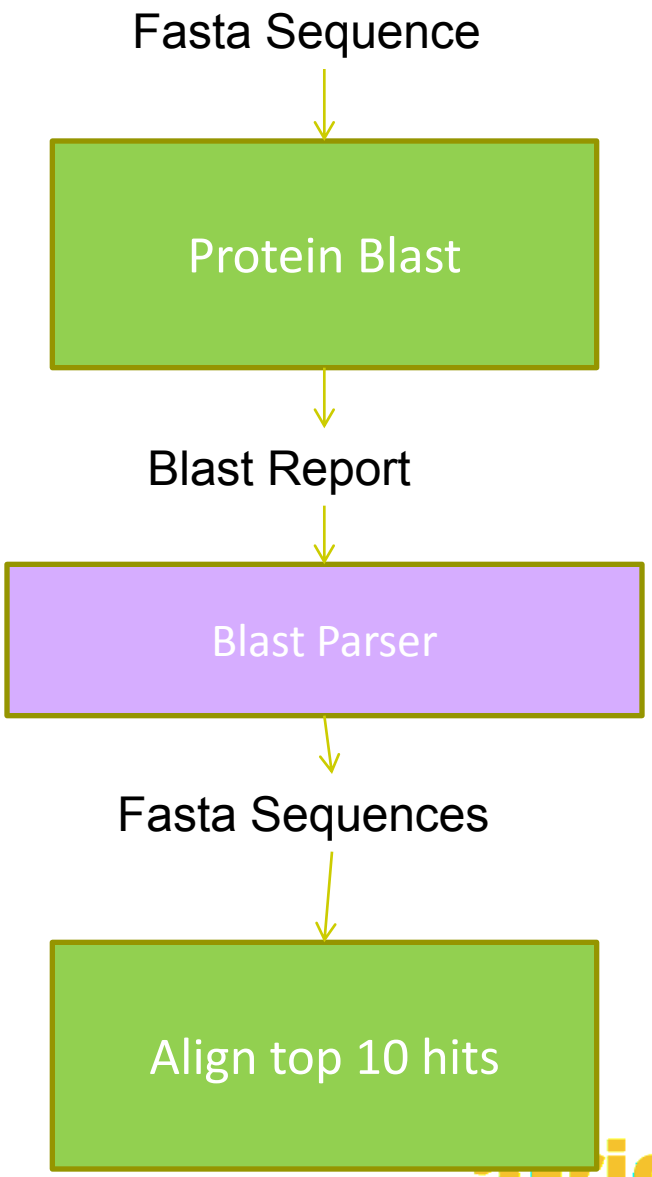
1. Understand how services work – inputs, outputs, dependencies → service descriptions and documentation
2. Find and use SHIM (or helper) services to combat incompatibilities

A Shim Service is a service that:

- doesn't perform an experimental function, but acts as a connector, or glue, when 2 experimental services have incompatible outputs and inputs

Shim Example





Understanding how services work

BioCatalogue beta
"The Life Science Web Service Registry"

Getting Started | About Us | Contact Us | API Docs

Sign up | Sign in

Search: Go! | Home | Services | Register a Service | Providers

Home »

SHARE

The BioCatalogue: providing a curated catalogue of Life Science Web Services

Latest Activity

Last 7 days

- vasunun **joined** the BioCatalogue
- Peter Taschner **added** a publication annotation to the Soap Service of Service: MutalyzerService
- Peter Taschner **added** a contact annotation to the Service Deployment of Service: MutalyzerService
- Peter Taschner **added** a tag annotation to Service: MutalyzerService
- Peter Taschner **added** a tag annotation to Service: MutalyzerService
- Peter Taschner **added** a tag annotation to Service: MutalyzerService
- Peter Taschner **added** a alternative name annotation to Service: MutalyzerService
- Peter Taschner **added** a documentation url annotation to the Soap Service of Service: MutalyzerService
- Peter Taschner **added** a description annotation to the Soap Service of Service:

The BioCatalogue currently has 1730 services, 130 service providers and 445 members

"Web Services are hard to find"
DISCOVER
- Find the right Web Service
- Powerful search and filtering
- Information from providers and community

More info

"My Web Services are not visible"
REGISTER
- Easily register Web Services
- Instantly available to everyone
- Providers can advertise, describe and monitor their Services

More info

"Web Services are poorly described"
ANNOTATE
- Anyone can describe and annotate
- Ongoing expert curation
- Social curation by the community

More info

"Web Services are volatile"
MONITOR
- Services change and get outdated
- BioCatalogue monitors Services
- Monitors availability and reliability

More info

Our Partners

EMBRACE Grid | myexperiment | myGrid | seekda! | moby

The EMBRACE Registry and the BioCatalogue have now been merged

Site Announcements

Have your say about BioCatalogue by taking part in the BioCatalogue users's survey
By Franck Tanoh (4 days ago

BioCatalogue Maintenance - 7 December 2010 @ 9:30 am (GMT)
By Eric E. Nzuobontane (6 days ago

BioCatalogue iPhone and iPad app now available to download for free
By Franck Tanoh (2 months ago

The BioCatalogue Functional Unit paper presented at IEEE 2010 Fourth International Workshop on Scientific Workflows
By Franck Tanoh (2 months ago

The National Cancer Research Institute (NCRI) joins forces with the BioCatalogue
By Franck Tanoh (4 months ago

More

Latest Services

MutalyzerService
dbfetch
graphtools
PRANK (REST)
EASTM (REST)

BioCatalogue.org - Services - InterProScan (REST) (REST) [Provided by: European Bioinformatics Institute (EBI)] - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.biocatalogue.org/services/2769

the univeristy of manchester

Most Visited | Getting Started | Latest Headlines | Post to CiteULike | Using Google Maps in ...

BioCatalogue.org - Services - InterPr...

BioCatalogue beta

"The Life Science Web Service Registry"

Getting Started | About Us | Contact Us | API Docs

Sign up | Sign in

Search: Go! | Home | Services | Register a Service | Providers

Home » Services » InterProScan (REST)

SHARE

InterProScan (REST) REST

99 0

aka JDispatcher aka InterProScan

Categories: Function Prediction

Monitoring

Annotations: 30 0 30 0

Overview | REST Endpoints (6) | Monitoring | News

Provider:
European Bioinformatics Institute (EBI)

Provider

Location:
UNITED KINGDOM

Submitter

Submitter / Source:
Hamish McWilliam (about 1 month ago)

Service Description

Base URL:
http://www.ebi.ac.uk/Tools/services/rest/iprscan

Tags

Tags (19)

bioinformatics | ebi | embl-ebi | Gene3D | HAMAP | interpro | interproscan | Panther | Pfam | PIRSF | PRINTS | ProDom | ProSite | protein domain | protein family | protein function | SMART | SUPERFAMILY | TIGRFAMs

Login to add tags

Favourited By (0)

Done

Managing Changes to Services

Monitoring detects changes, but the community site can notify users about changes → advanced warning

- EBI – Soaplab EMBOSS tools discontinued Feb 13
  - Redirect to alternative services (also from EBI)
- KEGG – SOAP services discontinued December 12
  - Replacing with equivalent REST services
- Help identify equivalent or similar services

MANCHESTER 1824

The University of Manchester

GETTING STARTED WITH TAVERNA: DEMO

myGrid

Enrichment Analysis

Many experiments result in a list of genes (e.g. microarray analysis, Chip-Seq, SNP identification etc)

- Today, we will use Taverna to perform enrichment analyses on a list of genes
- We will enrich our dataset by discovering:

1. Which pathways our genes are involved in and visualising those pathways
2. The functions of the genes using Gene Ontology annotations

MANCHESTER 1824

The University of Manchester

TAVERNA IN USE

myGrid

MANCHESTER 1824

The University Manchester

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

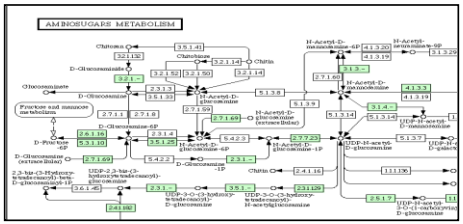
High throughput screening Proteomics

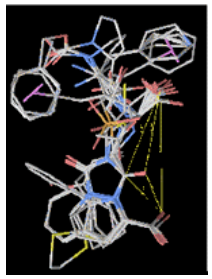
Phenotypical studies Text mining

Public Health care epidemiology

Medical image analysis QTL studies

QSAR studies Genome Wide Association Studies





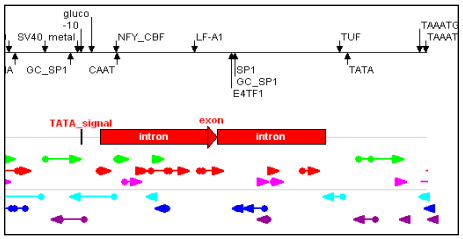
Astronomy

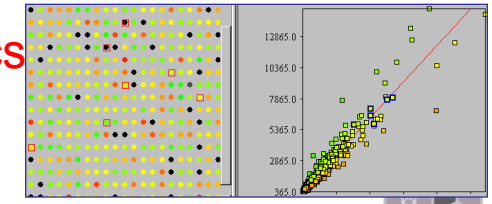
Music

Meteorology

Social Science

Cheminformatics

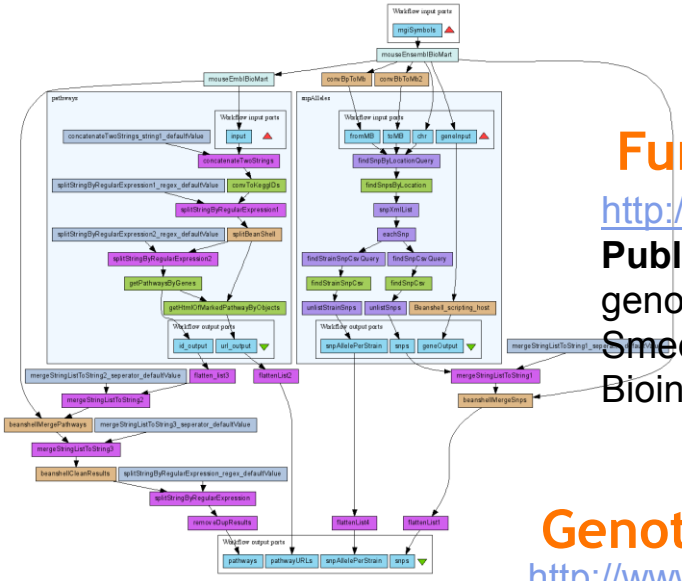




myGrid

MANCHESTER 1824

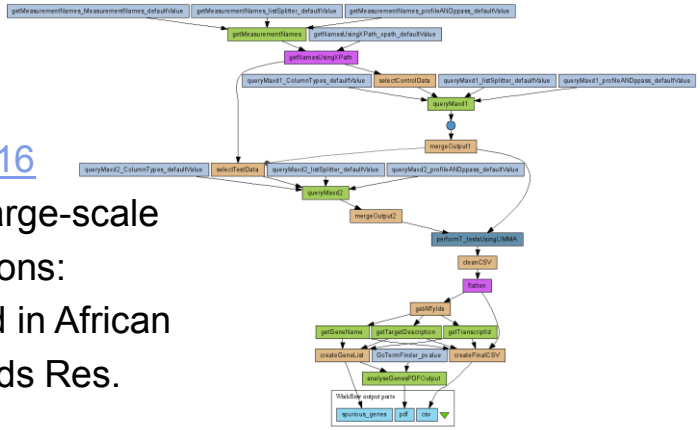
Taverna for Omics



Functional Genomics

http://www.myexperiment.org/workflows/126

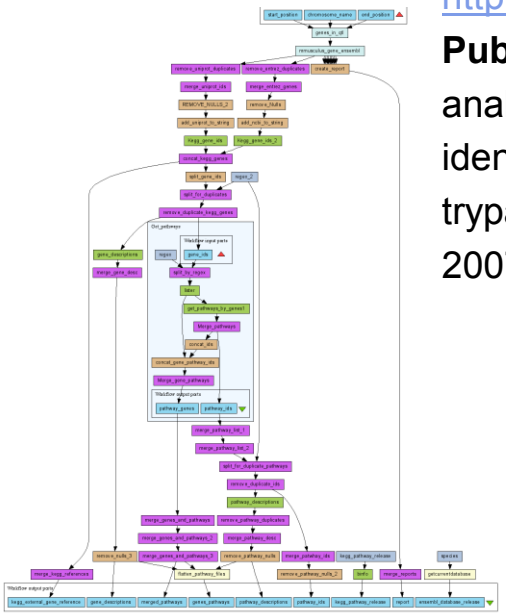
Publication: Solutions for data integration in functional genomics: a critical assessment and case study. Smedley, Swertz and Wolstencroft, et al Briefings in Bioinformatics. 2008 Nov;9(6):532-44.



Genotype to Phenotype

http://www.myexperiment.org/workflows/16

Publication: A systematic strategy for large-scale analysis of genotype phenotype correlations: identification of candidate genes involved in African trypanosomiasis. Fisher et al Nucleic Acids Res. 2007;35(16):5625-33



Next Generation Sequencing

- Whole Genome SNP analysis of different cattle species in response to trypanosomiasis infection (sleeping sickness)
- Large data processing strategies
- Taverna in the cloud – deploying and running large data processes using cloud computing services

myGrid

MANCHESTER 1824

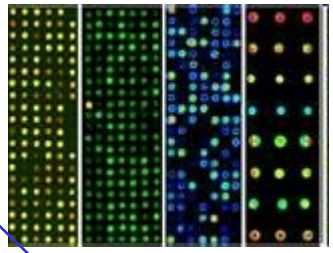
The University of Manchester

Research Example

Lymphoma Prediction Workflow

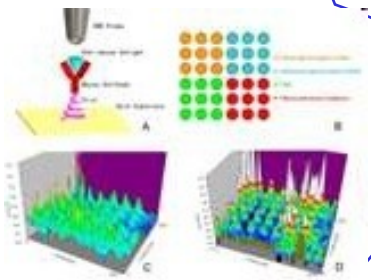
caBIG

MicroArray from tumor tissue

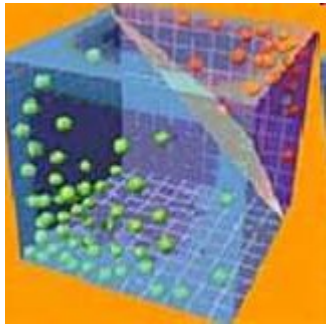


caArray

Use gene-expression patterns associated with two lymphoma types to predict the type of an unknown sample.

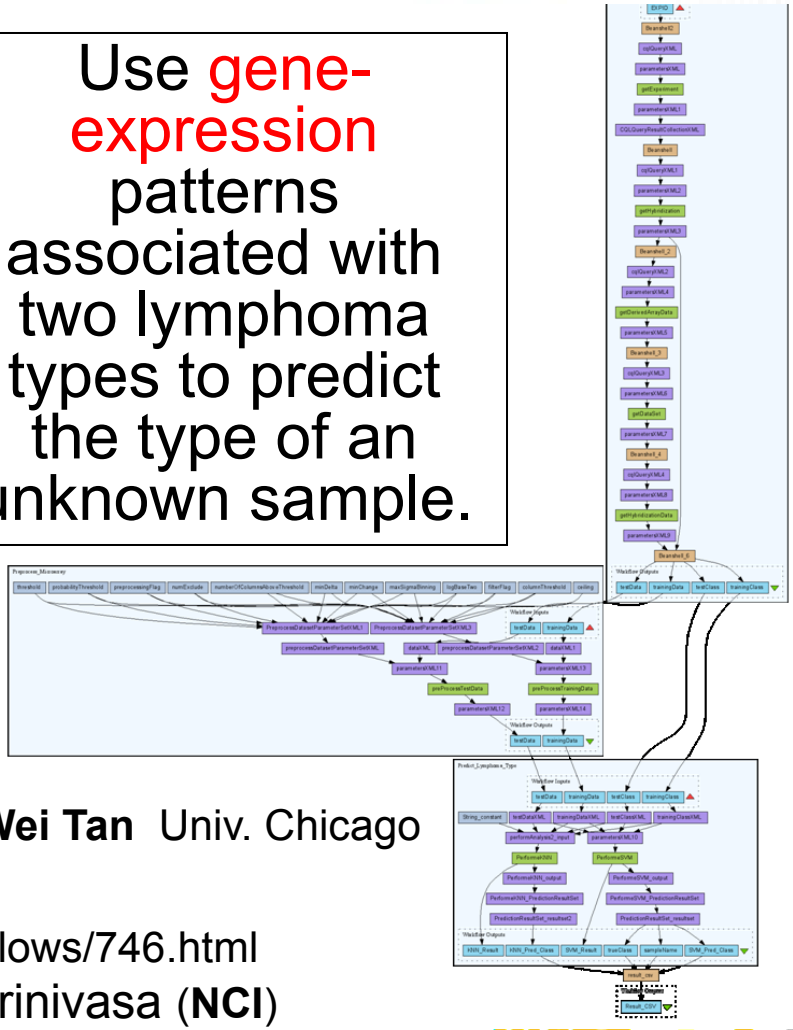


Microarray preprocessing



Lymphoma prediction

GenePattern



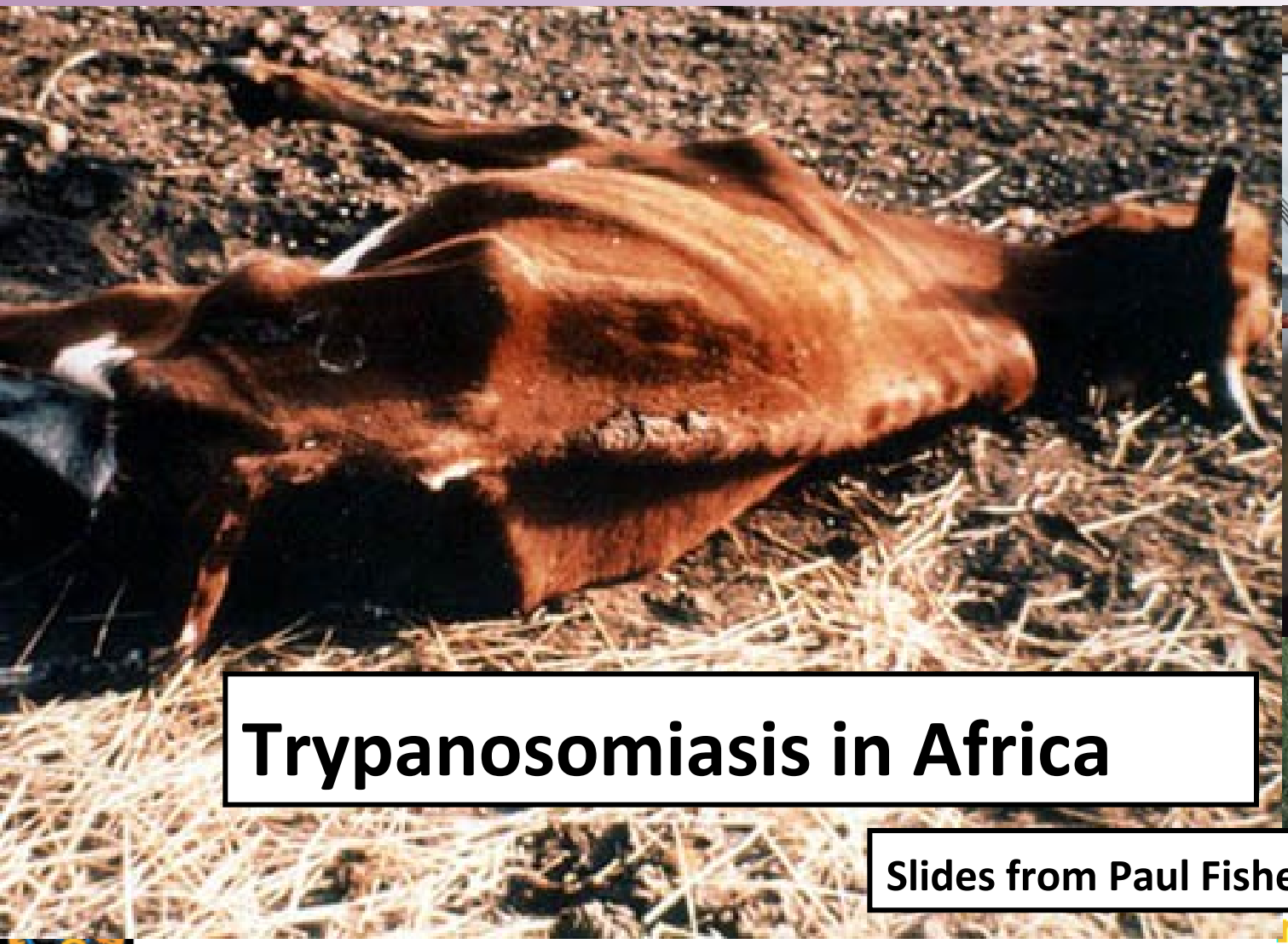
Wei Tan Univ. Chicago

Wei Tan: http://www.myexperiment.org/workflows/746.html

Ack. Juli Klemm, Xiaopeng Bian, Rashmi Srinivasa (**NCI**)
Jared Nedzel (**MIT**)

myGrid

The Wellcome Trust Funded Host-Pathogen Project



Steve Kemp

Andy Brass

Paul Fisher

Trypanosomiasis in Africa

Slides from Paul Fisher

http://www.genomics.liv.ac.uk/tryps/trypsindex.html

Cattle Disease Research

$4 billion US

Different breeds of African Cattle
- Some resistant
- Some susceptible

African Livestock adaptations:
- More productive
- Increases disease resistance
- Selection of traits

Potential outcomes:
- Food security
- Understanding resistance
- Understanding environmental
- Understanding diversity

http://www.bbc.co.uk/news/10403254

LIVE BBC NEWS CHANNEL

Page last updated at 18:03 GMT, Thursday, 24 June 2010 19:03 UK

E-mail this to a friend Printable version

African livestock offers 'untapped genetic resource'

By Mark Kinver
Science and environment reporter, BBC News

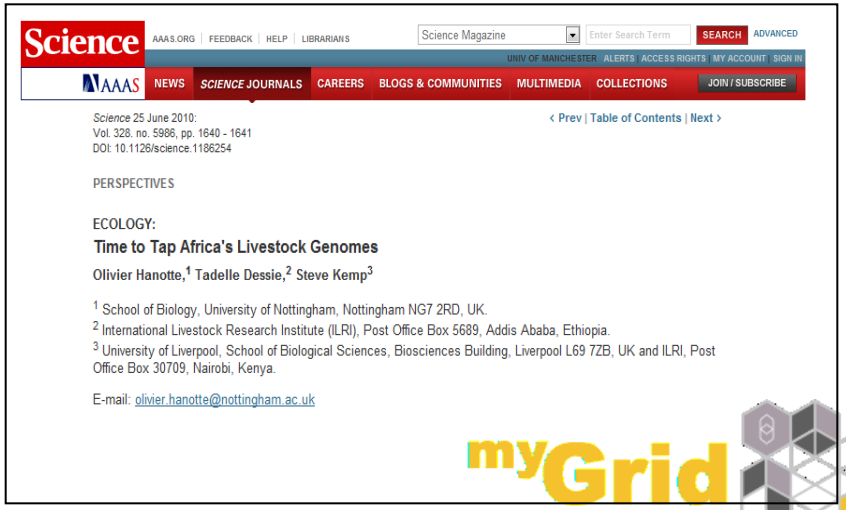
Over the centuries, indigenous breeds have adapted to conditions in Africa

The genetic diversity of Africa's indigenous livestock needs to be tapped before it is lost forever, researchers have warned.

SEE ALSO
- Livestock breeds face 'meltdown' 03 September 07 | Sci/Tech
- Developing nations need cloning 25 January 07 | Sci/Tech
- Warning signs on Kenya's drought road 14 November 06 | Sci/Tech
- 'Barren future' for Africa's soil 30 March 06 | Sci/Tech
- Africa bird flu disaster warning 22 February 06 | Africa

RELATED INTERNET LINKS
- University of Nottingham
- ILRI
- Science

The BBC is not responsible for the content of external internet sites



Science AAAS.ORG | FEEDBACK | HELP | LIBRARIANS Science Magazine Enter Search Term SEARCH ADVANCED

UNIV OF MANCHESTER | ALERTS | ACCESS RIGHTS | MY ACCOUNT | SIGN IN

AAAS NEWS SCIENCE JOURNALS CAREERS BLOGS & COMMUNITIES MULTIMEDIA COLLECTIONS JOIN / SUBSCRIBE

Science 25 June 2010:
Vol. 328, no. 5986, pp. 1640 - 1641
DOI: 10.1126/science.1186254

‹ Prev | Table of Contents | Next ›

PERSPECTIVES

ECOLOGY:

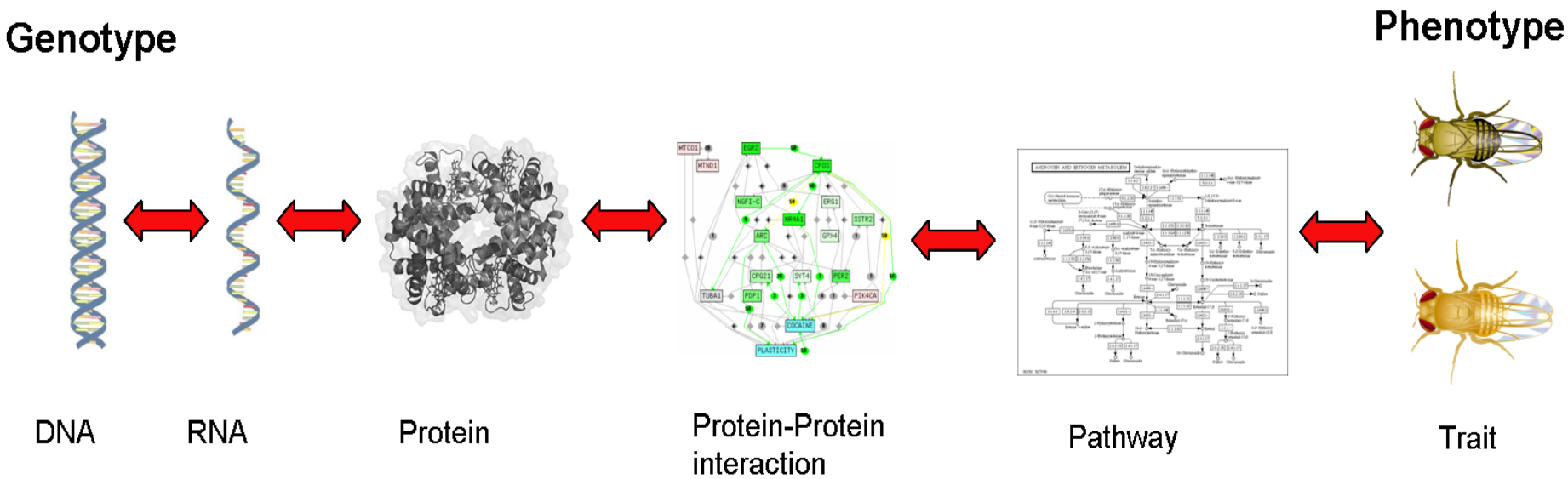
Time to Tap Africa's Livestock Genomes

Olivier Hanotte,[1] Tadelle Dessie,[2] Steve Kemp[3]

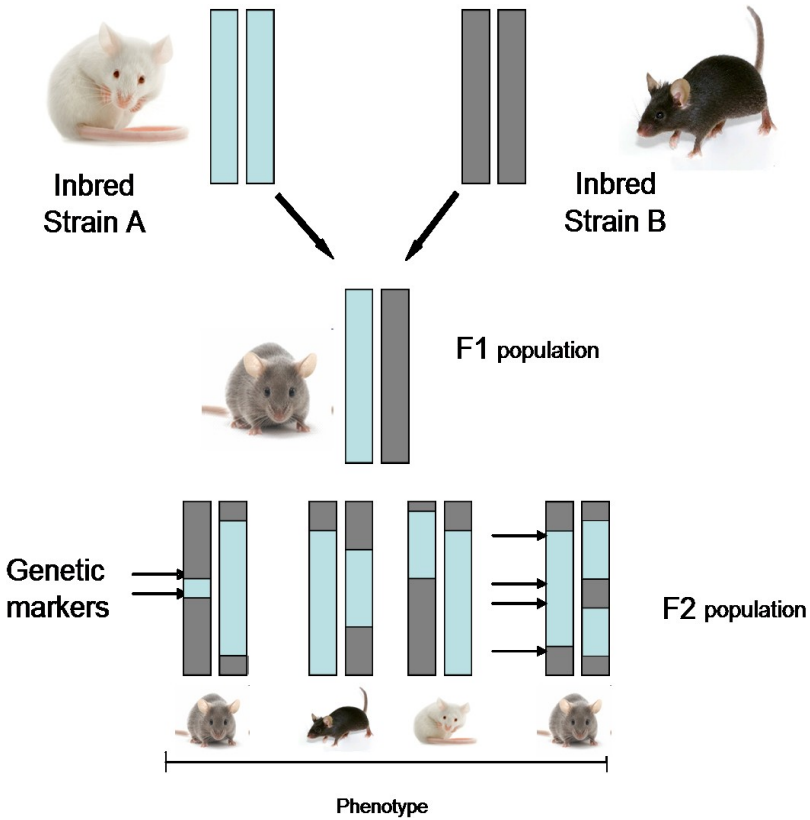
[1] School of Biology, University of Nottingham, Nottingham NG7 2RD, UK.
[2] International Livestock Research Institute (ILRI), Post Office Box 5689, Addis Ababa, Ethiopia.
[3] University of Liverpool, School of Biological Sciences, Biosciences Building, Liverpool L69 7ZB, UK and ILRI, Post Office Box 30709, Nairobi, Kenya.

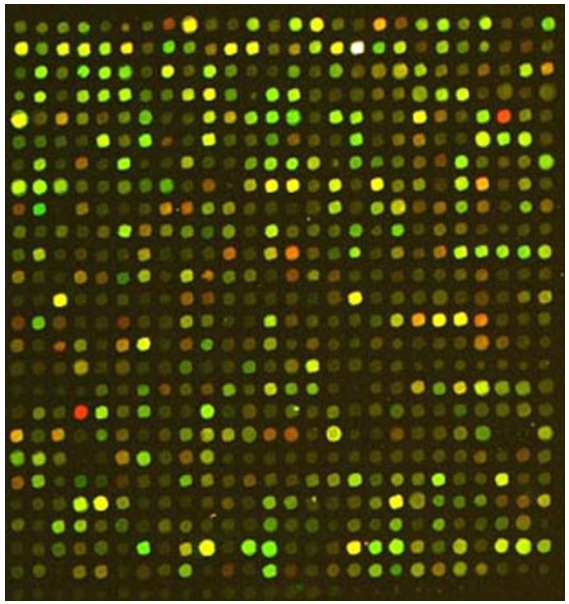
E-mail: olivier.hanotte@nottingham.ac.uk

Understanding the process: Genotype - Phonotype

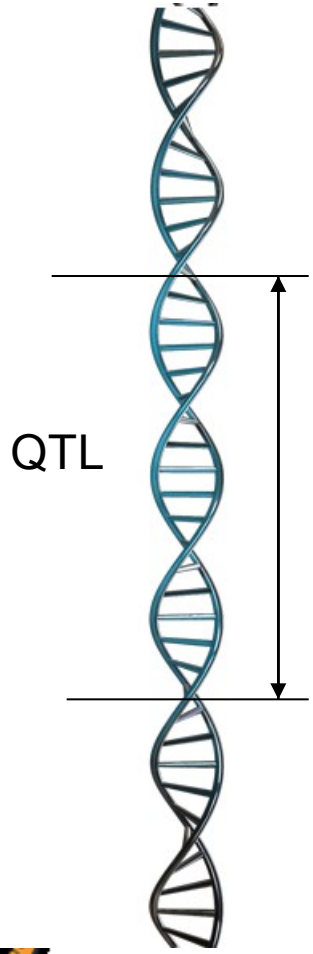


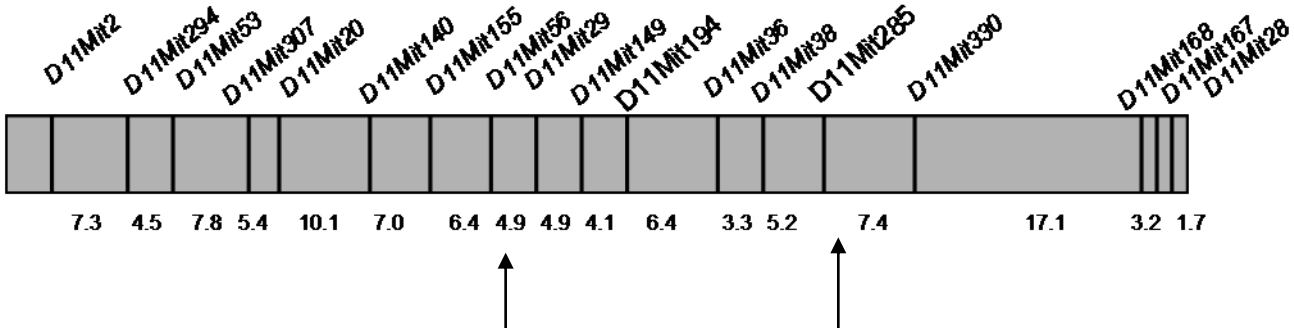
QTL + Microarrays





Quantitative Trait Loci (QTL)

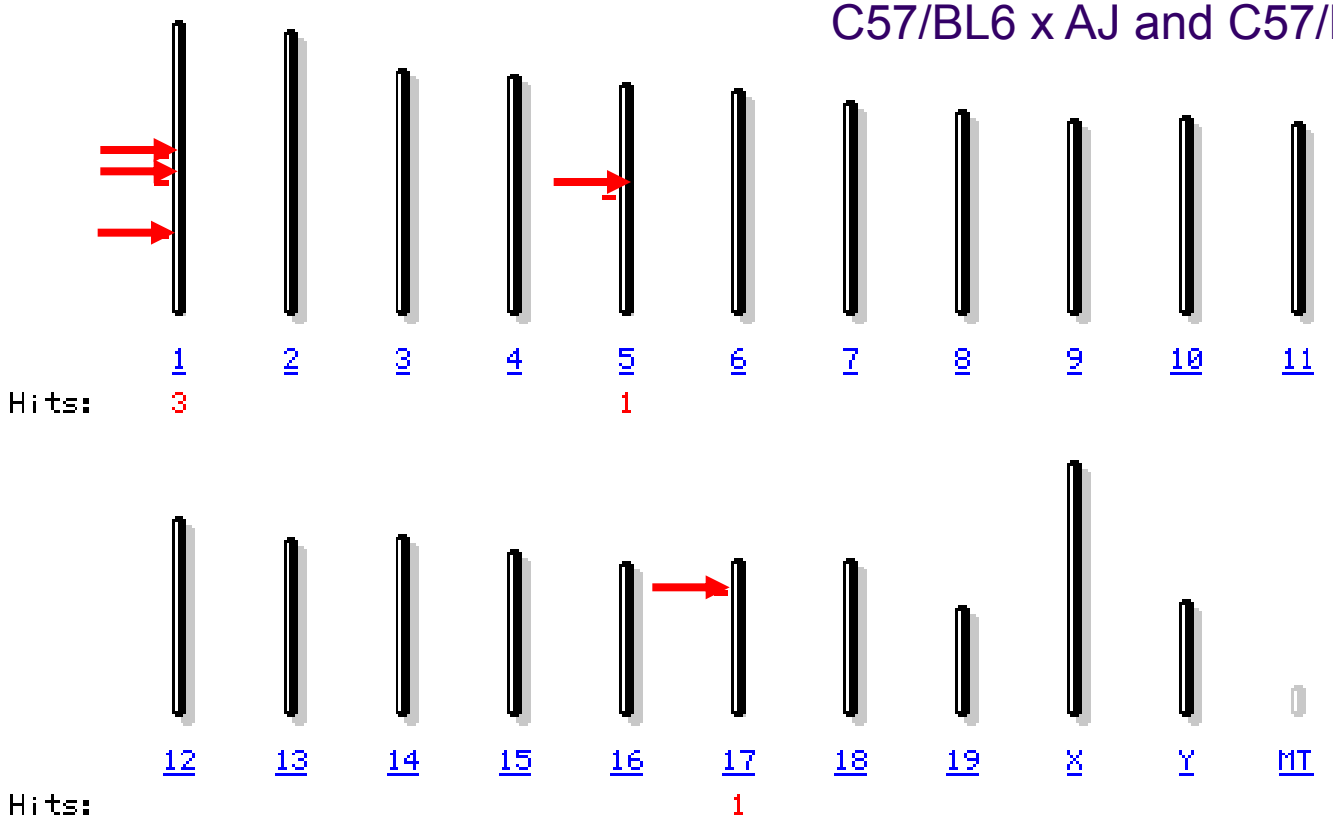




- Regions of chromosomes have distinctive base pair sequences, called markers
- Markers can be assembled into correct order to find regions of chromosomes
- QTL studies can be used to identify markers that correlate with a disease
- QTLs can span
  - small regions containing few genes
  - encompass almost entire chromosomes containing 100's of genes

Trypanosoma infection response (Tir) QTL

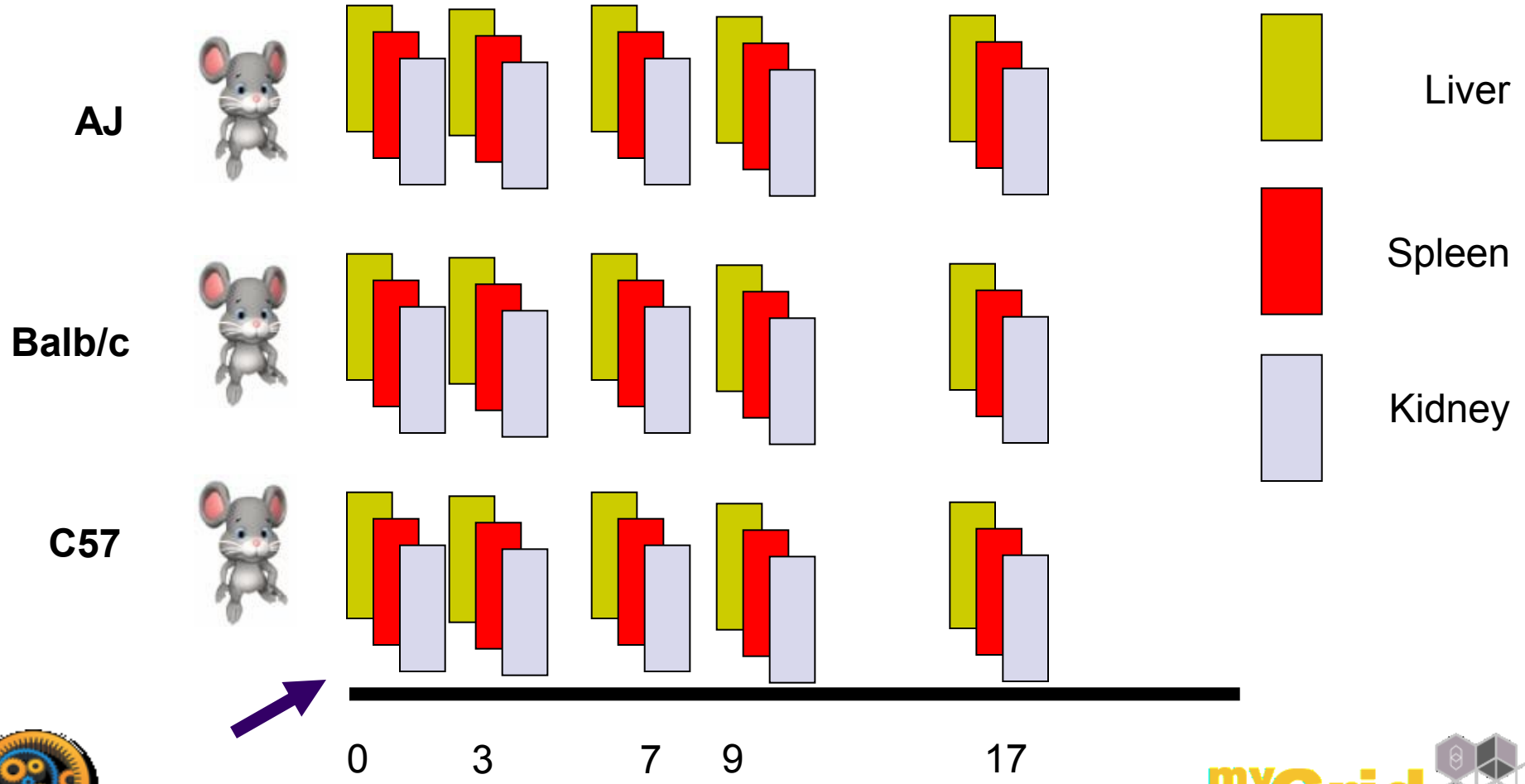
C57/BL6 x AJ and C57/BL6 x BALB/C



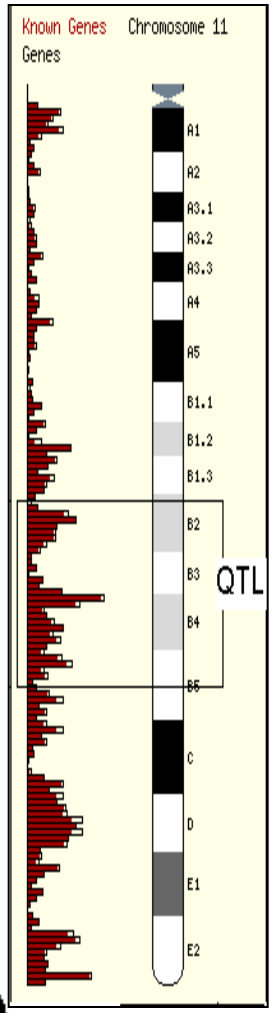
Iraqi et al Mammalian Genome 2000 11:645-648
Kemp et al. Nature Genetics 1997 16:194-196

The experiment

A total of 225 microarrays

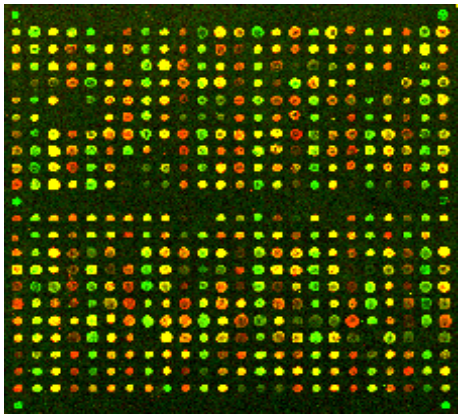


Huge amounts of data



QTL region on chromosome

200+ Genes



Microarray

1000+ Genes

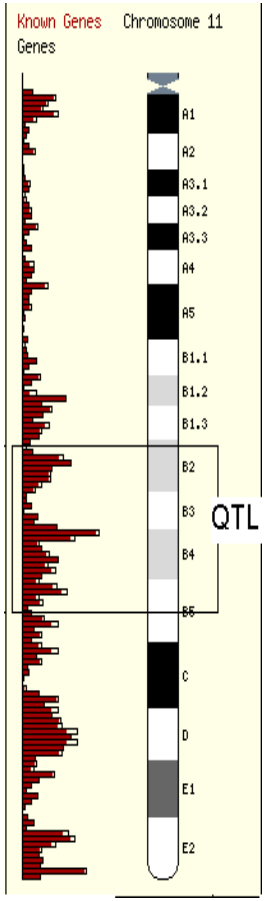
How do I look at ALL the genes systematically?

MANCHESTER 1824

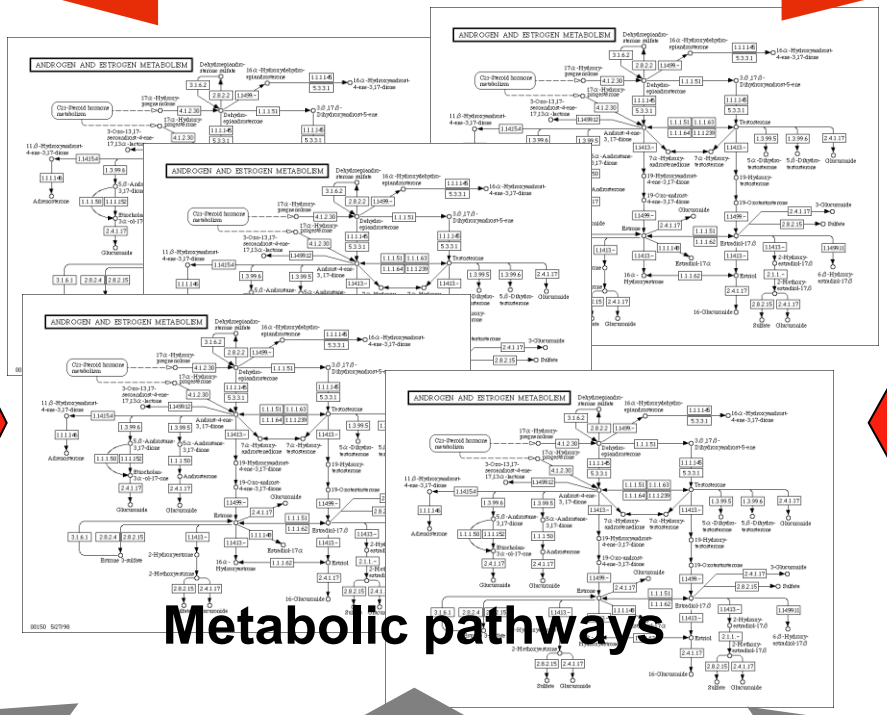
The University of Manchester

Genotype

Phenotype

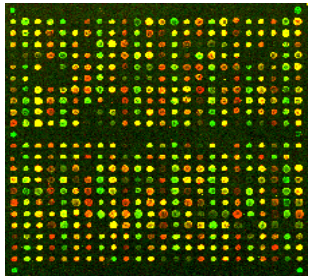


200



Metabolic pathways

?



Microarray + QTL

Phenotypic response investigated using microarray in form of expressed genes or evidence provided through QTL mapping

Genes captured in microarray experiment and present in QTL (**Quantitative Trait Loci**) region

myGrid

MANCHESTER 1824

The University of Manchester

Data analysis

- Identify pathways that have differentially expressed genes (from microarray studies)
- Identify pathways from Quantitative Trait genes (QTg)
- Track genes through pathways that are suspected of being involved in resistance/susceptibility

myGrid

MANCHESTER 1824

The University of Manchester

Trypanosomiasis Resistance Results

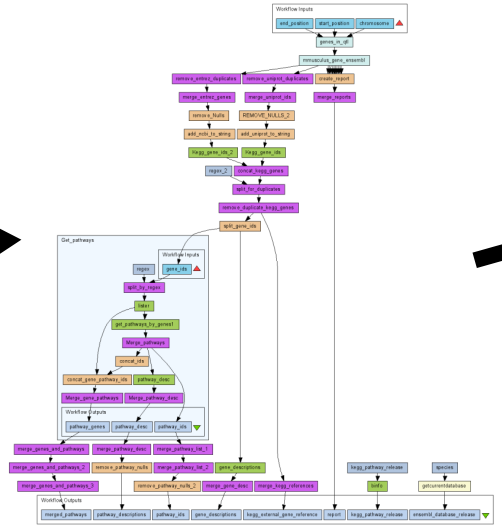
- DAXX gene identified in the workflows
- **Daxx** gene not found using manual investigation methods
- Sequencing of the Daxx gene in **Wet Lab** (at Liverpool) showed mutations that are thought to change the structure of the protein
- These mutations were also published in scientific literature, noting its effect on the **binding of Daxx protein to p53 protein**
- **p53 plays direct role in cell death and apoptosis, one of the Trypanosomiasis phenotypes**

myGrid

Reuse, Recycle, Repurpose Workflows

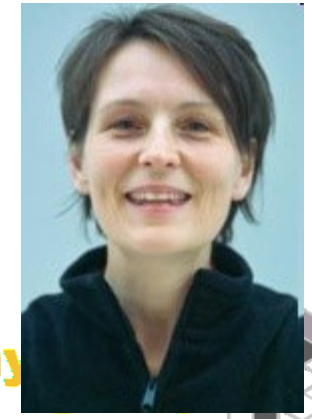
Dr Paul Fisher

Identify QTg and pathways implicated in resistance to Trypanosomiasis in the mouse model



Dr Jo Pennock

Identify the QTg and pathways of colitis and helminth infections in the mouse model



PubMed ID: 20687192

Same Host, another Parasite...but the SAME Method

- Mouse whipworm infection - parasite model of the human parasite - *Trichuris trichuria*

Understanding Phenotype

- Comparing resistant vs susceptible strains – Microarrays

Understanding Genotype

- Mapping quantitative traits – Classical genetics QTL

Joanne Pennock, Richard Grencis
University of Manchester

Workflow Results

- Identified the biological pathways involved in sex dependence in the mouse model, previously believed to be involved in the ability of mice to expel the parasite.
- Manual experimentation: Two year study of candidate genes, processes unidentified
- Workflow experimentation: Two weeks study – identified candidate genes

Joanne Pennock, Richard Grencis
University of Manchester

"Traditional" Hypothesis-Driven Analyses

'*Cherry Pick*' genes

200 genes

Pick the genes involved in immunological process

40 genes

Pick the genes that I am most familiar with

2 genes

What about the other 198 genes? What do they do?

Biased view

Workflow Success

- Workflow analysed each piece of data *systematically*
  - Eliminated user bias and premature filtering of datasets
- The size of the QTL and amount of the microarray data made a manual approach impractical
- Workflows capture exactly where data came from and how it was analysed
- Workflow output produced a manageable amount of data for the biologists to interpret and verify
  - "make sense of this data" -> "does this make sense?"

MANCHESTER 1824

The University of Manchester

Sharing and Reusing Workflows

myexperiment

myGrid

Workflow Repository

MANCHESTER 1824

my experiment

About | Mailing List | Publications

Logout | Give us Feedback | Invite

Home | Users | Groups | Workflows | Files | Packs | Services | Topics

Workflows | Search

Home > Workflows

Workflows

Search filter terms

Filter by type

- Taverna 2 1152
- Taverna 1 645
- RapidMiner 223
- Kepler 43
- Bioclipse Scrip... 34
- LONI Pipeline 26
- GWorkflowDL 24
- BioExtract Server 17
- Tesla 11
- Galaxy 10

Filter by tag

- example 230
- mygrid 104
- localworker 103
- bioinformatics 102
- graph 91

« previous | 1 | 2 | 3 | ... | 230 | next »

Sort by: Rank

Showing 2293 results. Use the filters on the left and the search box below to refine the results.

Search

Taverna 2

Original Uploader

Paul Fisher

Pathways and Gene annotations for QTL region (v7)

View

Download (v7)

Created: 19/11/09 @ 18:18:52 | **Last updated:** 07/09/12 @ 18:23:36

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 searog for pathways in the KEGG path...

Rating: 4.6 / 5 (10 ratings) | **Versions:** 7 | **Reviews:** 1 | **Comments:** 7 |

New/Upload

Workflow | GO

Katy Wolstencroft

- My Profile [edit]
- My Messages
- My Memberships (4)
- My History
- My News

3 new friendship requests

- mihaiionita_me
- Pankaj chauhan
- Hanny

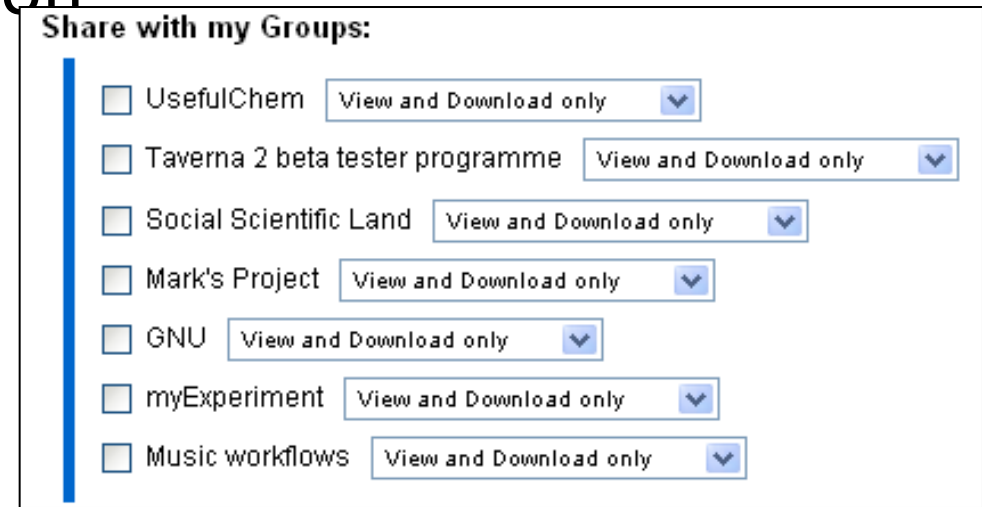
4 new group requests

- From Rolando Milian (for Group: Tutorial)
- From Charlyb (for Group: Msc Tutorial)
- From Mkh (for Group: MIB_Tutorial)
- From Mateenraj

Just Enough Sharing….

myExperiment can provide a central location for workflows from one community/group

- You specify:
  - Who can look at your workflow
  - Who can download and run your workflow
  - Who can modify your workflow
- Ownership and attribution



Community myExperiments

myexperiment

About | Mailing List | Publications

Log in | Register | Give us Feedback

Home | Users | Groups | Workflows | Files | Packs | Services | Topics

Workflows | Search

Home > Workflows

New/Upload

Workflow | GO

Workflows

Search filter term

Filter by type

- Taverna 2
- Taverna 1
- RapidMiner
- Kepler
- Bioclipse Scrip...
- LONI Pipeline
- GWorkflowDL
- BioExtract Server
- Tesla
- KNIME

Filter by tag

- example
- mygrid
- localworker

BioVeL

Biodiversity Virtual e-Laboratory

on myexperiment

Log in | Register | Give us Feedback | Invite

Home | Users | Groups | Workflows | Files | Packs | Services | Topics

All | Search

Home > Groups > BioVeL

| Members (46) |

Group for sharing workflows
FP7-283359 BioVeL

For more information visit htt

Created at: Saturday 06 Augus

Unique name: biovel

PANACEA MyExperiment

Log in | Register | Give us Feedback | Invite

Home | Users | Groups | Workflows | Files | Packs | Services | Topics

Workflows | Search

Home » Workflows

New/Upload

Workflow | GO

Workflows

Search filter terms

« previous | 1 | 2 | 3 | ... | 6 | next »

Sort by: Rank

Showing 59 results. Use the filters on the left and the search box below to refine the results.

Search

Filter by type

- Taverna 2 59

Filter by tag

- panacea 19
- basicxces 17
- freeling 17
- example 15
- tagging 15
- crawled 11
- bilingual 7
- cgp 5

Taverna 2

bilingual word aligner for crawled data (v3)

View

Download (v3)

Created: 19/04/11 @ 15:18:06 | **Last updated:** 03/06/11 @ 10:18:35

Credits: Marcpoch

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Original Uploader

Marcpoch

This is a word alignment workflow using hunalign and giza++.

Log in / Register

Username or Email:

Password:

Remember me:

Log in

Need an account?
Click here to register

Forgot Password?

Popular Tags
25 tags

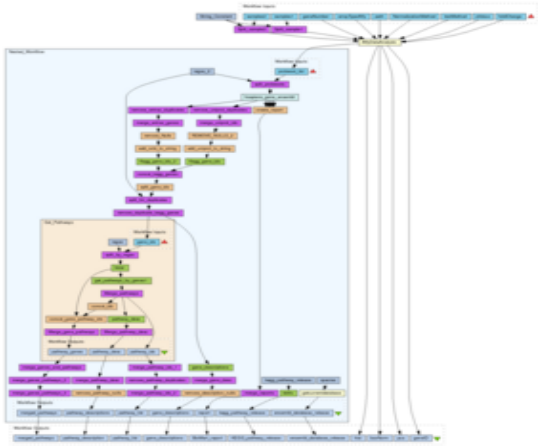
Workflow Entry: Microarray CEL file to candidate pathways

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 retruned by this workflow are a list of the top differentialy expressed genes (size dependant on the number specified as input – geneNumber), which are then used to find the candidate pathways which may be influencing the observed changes in the microarray data. By identifying the candidate pathways, more detailed insights into the gene expression data can be obtained.

CEL

Download

Download Scalable Diagram (SVG)

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

Ratings

Hover and click to rate

Current: 4.50/5 (2 ratings)

You rated it:

Breakdown

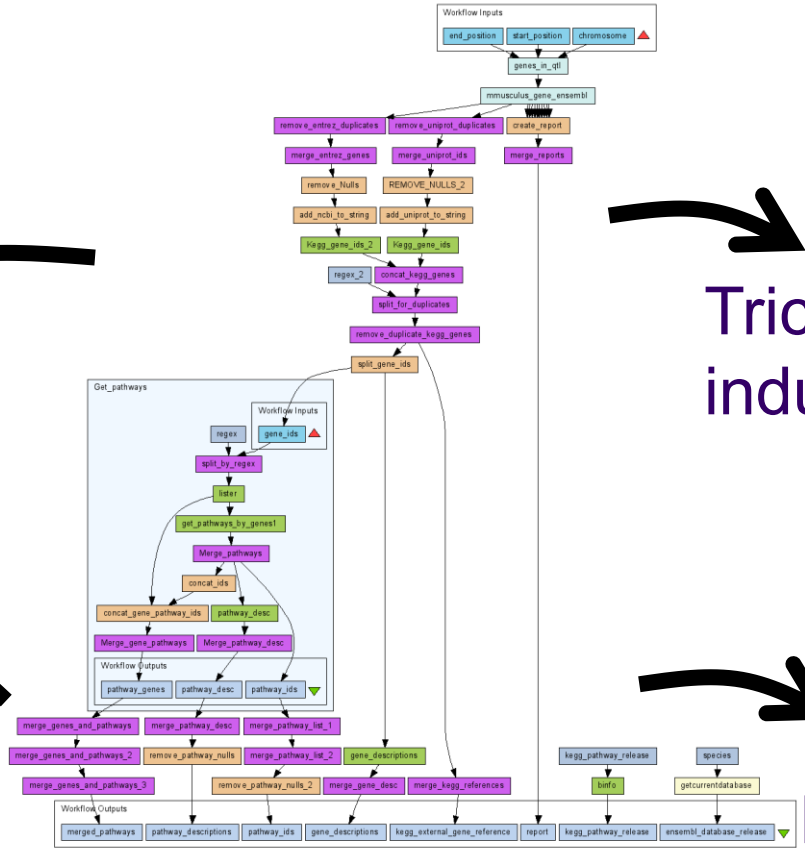
Paul Fisher - 5/5
Jits - 4/5

Statistics

Reuse, Reuse, Reuse

Atopic Dermatitis

Trichuriasis induced Colitis



Blood Pressure

Epilepsy

MANCHESTER 1824

The University of Manchester

FINDING AND USING A MYEXPERIMENT WORKFLOW: DEMO

myGrid

Workflow engine features

- Implicit iterations
  - With customisable list handling
- Parallelisation
  - Run as soon as data is available
- Streaming
  - Process partial iteration results early
- Retries, failover, looping
  - For stability and conditional testing

Data and Provenance

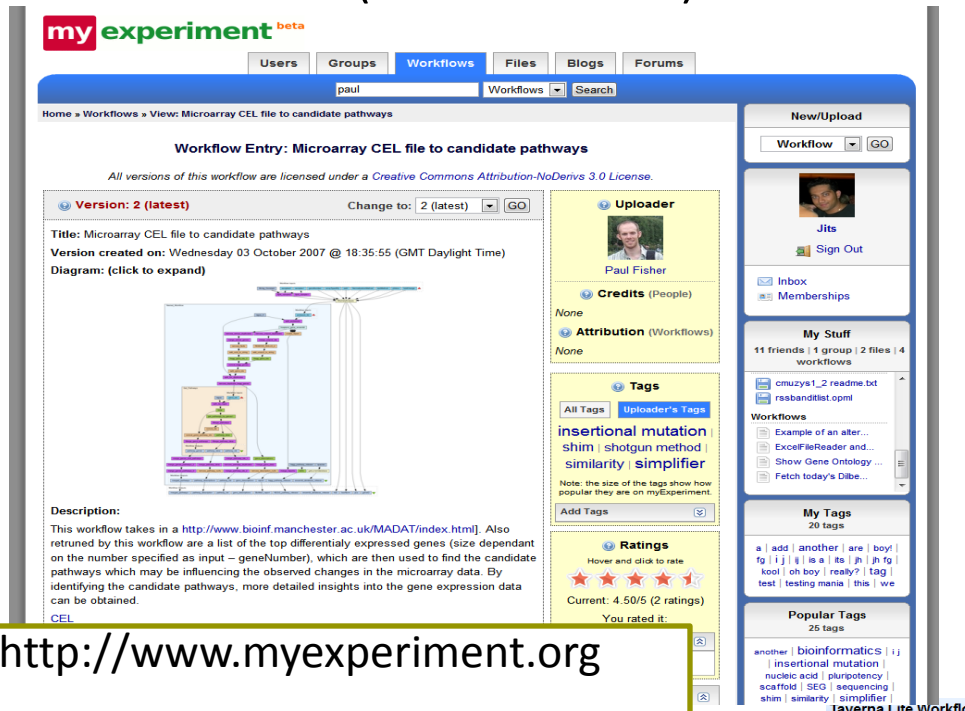
- Workflows can generate vast amount of data - how can we manage and track it?
- We need to manage data AND metadata AND experimental provenance
- Scientists need to check back over past results, compare workflow runs and share workflow runs with colleagues
- Scientists need to look at intermediate results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

Data and Provenance Handling

- **Provenance** captured for workflow runs
  - **Trace** execution steps, view **intermediate values** while running
  - Export as Open Provenance Model (OPM) / RDF
  - Proof and **origin** of produced outputs
  - Extensible **annotations**
- Wf4Ever: reproducible **research objects**
  - Workflow/data as a scientific publication → preservation
  - Need to capture more service data and metadata

Spectrum of Users

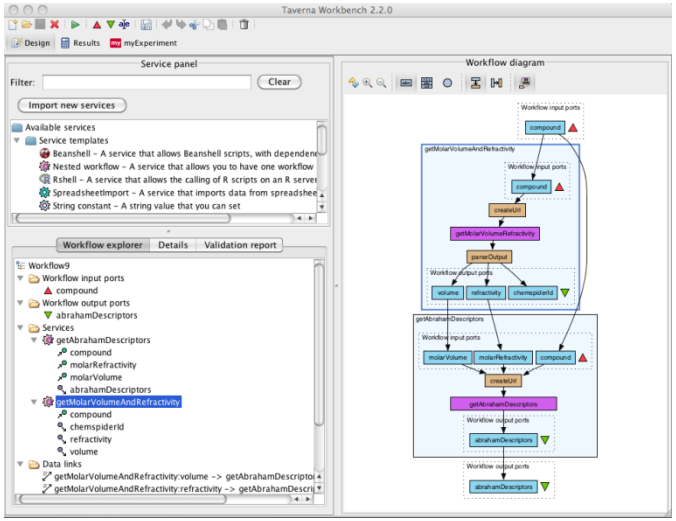
Advanced users design and build workflows (informaticians)



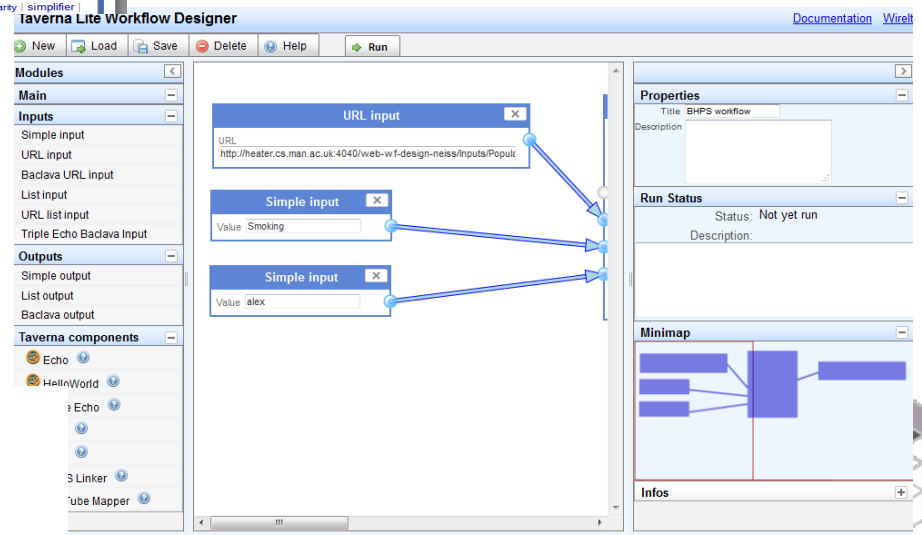
http://www.myexperiment.org

Load Data:

Run Workflow



Intermediate users reuse and modify existing workflows



Others "replay" workflows through a web interface or Taverna Lite

TAVERNA SERVER

Taverna Server

- Running workflows remotely
  - Through other client software
  - Via a web interface
- Tapping into remote computing resources
  - Execution on servers, grids or clouds

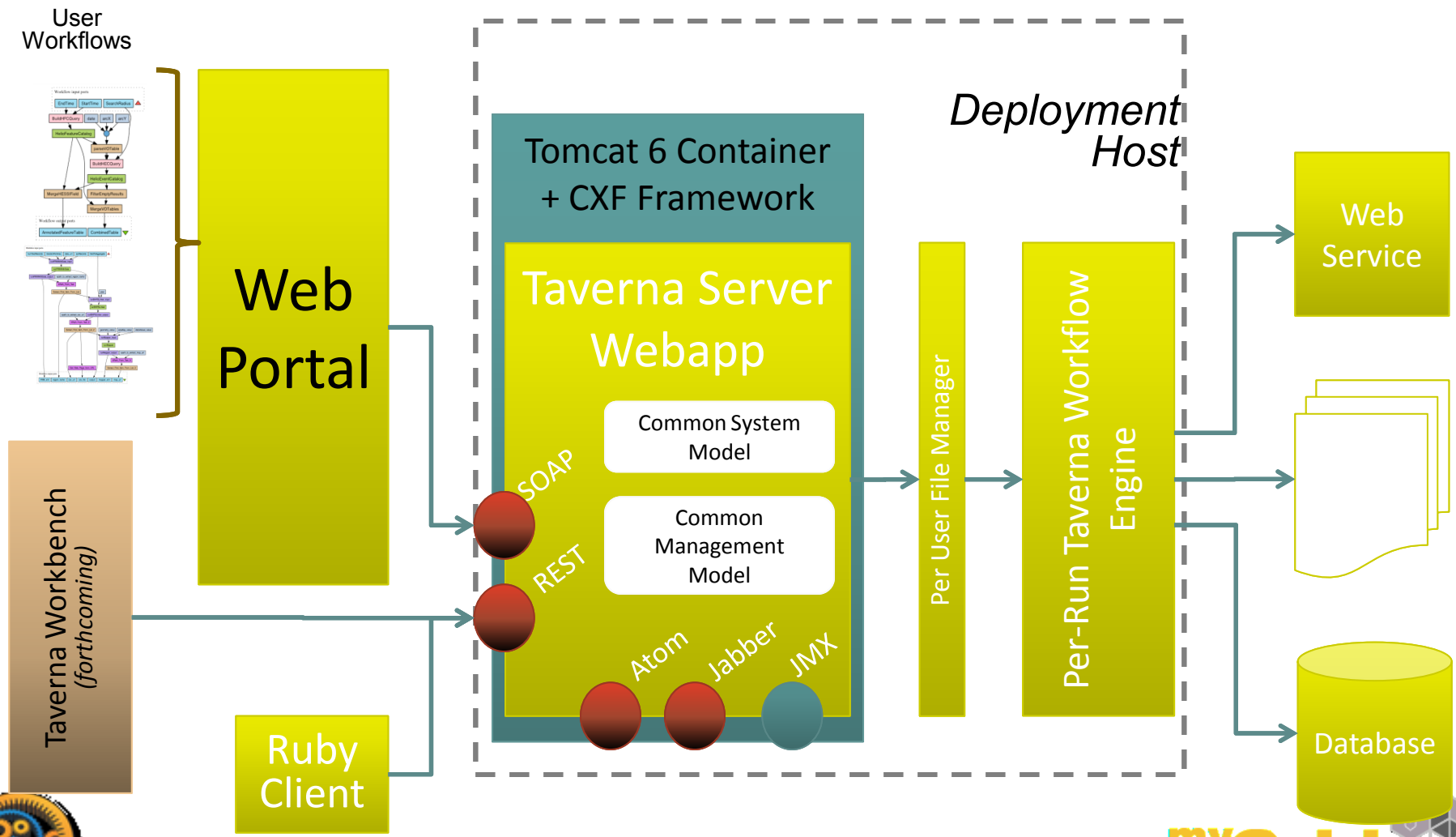
Limitations of the Desktop workbench

- You have to install it and learn how to use it
- Although computation could happen at remote service locations, data and computation can also happen locally
- High throughput experiments take a lot of compute and a lot of time
- Long running workflows need uninterrupted execution

Data Limitations with the Desktop Workbench

- Running the Workbench is limited by:
  - Local disk space for storing data
  - Network speeds for up/download
  - Firewall access

Taverna Server



MANCHESTER 1824

The University of Manchester

Taverna Server in Use

- T2Web, running myExperiment workflows through web interface
- HELIO - Heliophysics Integrated Observatory
- SCAPE - SCalable Preservation Environment (digital archives)
- BioVel – Biodiversity Virtual e-laboratory
- Cloud analytics for the life sciences – Taverna on the cloud
- Running Taverna through Galaxy

myGrid

T2 Web

Marco Roos
Kostas Karasavvas

Workflow: BioAID_ProteinDiscovery - Mozilla Firefox

www.mybiobank.org/t2web/workflow/74

netherlands bioinformatics centre

BioAID_ProteinDiscovery

workflow by Marco Roos

Leiden University Medical Cente

Configure Workflow Inputs

Description: Fill in a search query, similar to pubmed. For advanced queries look up the Lucene syntax (http://lucene.apache.org/java/2_9_1/queryparsersyntax.html).

Enter Query:

"transmembrane

Upload file?

Enter maxHits_parameter:

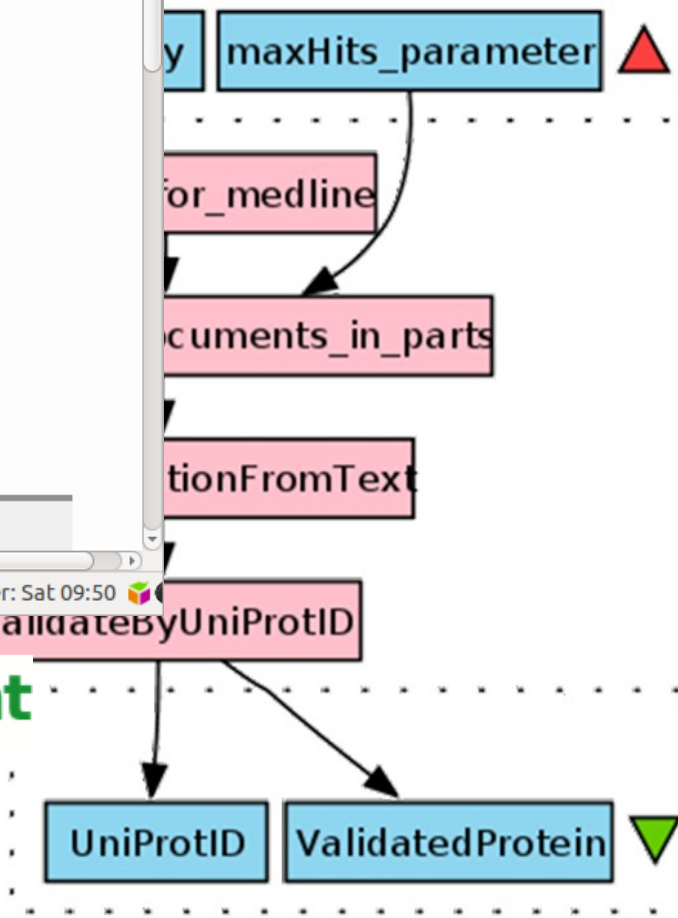
5

Upload file?

Execute

Workflow Description

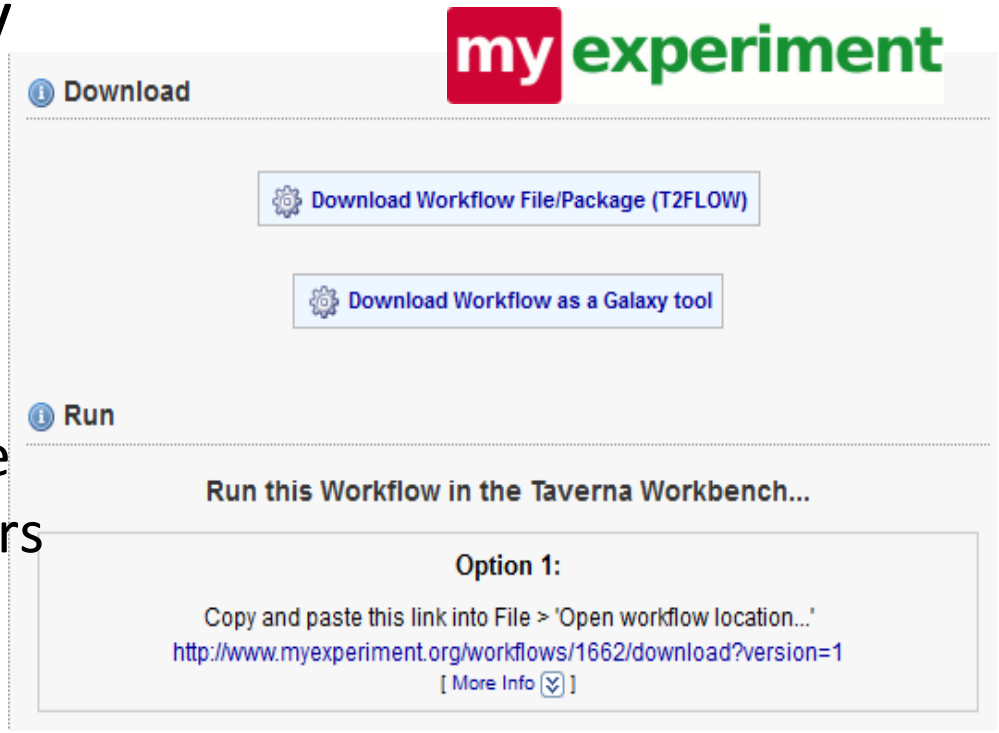
zotero UK: Sat 16:50 US Pacific: Sat 08:50 Hong Kong: Sat 23:50 GMT/UTC: Sat 15:50 Denver: Sat 09:50



myExperiment workflow ID

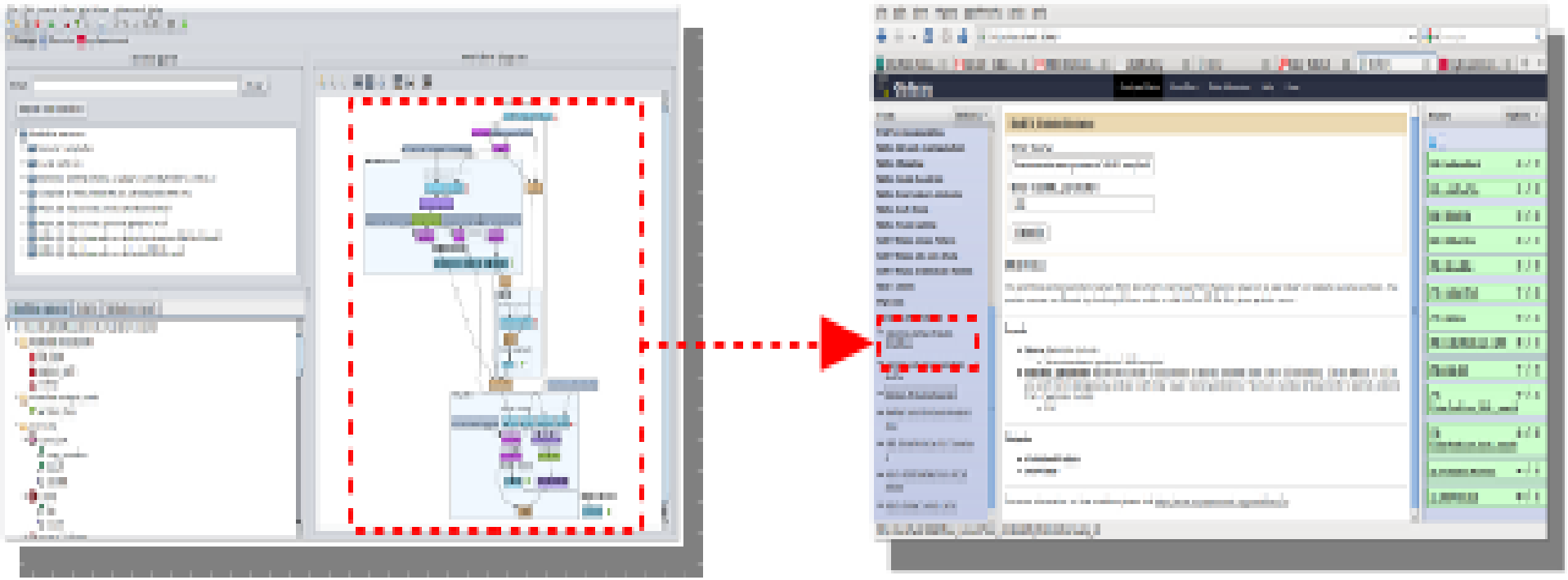
Running Taverna Through Galaxy

- Workflow interoperability
  - The methods are more important than the platform
  - Workflows in Galaxy and Taverna already exist
  - Any Taverna workflow can be made available to Galaxy users
  - Discover and import from myExperiment



Running Taverna through Galaxy

Kostas Karasavvas, NBIC



- Connect the Taverna and Galaxy communities
- Galaxy specialises in genomics, next gen sequencing etc
- Taverna can access more ‘downstream’ analysis services – e.g. pathway analyses, literature, GO enrichment etc

MANCHESTER 1824

The University of Manchester

Cloud Analytics for the Life Sciences

- Workflows for genetic diagnostics (for the NHS)
  - Exome and whole genome
  - SNP analysis and annotation
- Execution on the cloud
  - Secure execution and results handling
  - Elastic to cope with demand
  - Pay-as-you-go – cheap at the point of use

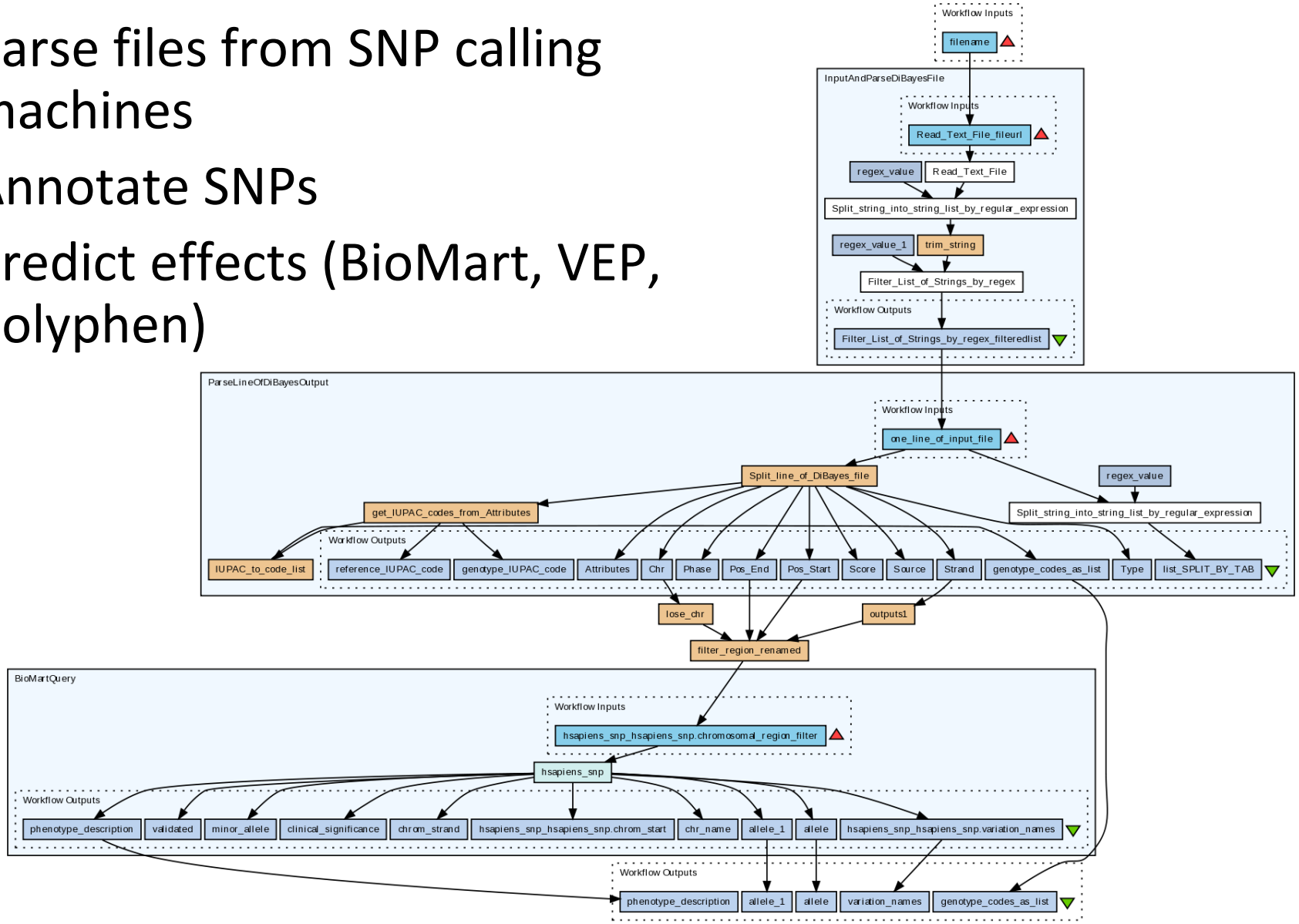
Cytocell

eagle genomics

myGrid

A Typical Workflow

- Parse files from SNP calling machines
- Annotate SNPs
- Predict effects (BioMart, VEP, polyphen)



MANCHESTER 1824

The University of Manchester

A Typical Workflow

ElasticView (v 0.1 Alpha) Powered by Eagle Genomics Ltd.

madhu Logout »

My Workspace

Data

Workflows

- Hello, World!
- Simple pass thr...
- Simple SNP work...
  - madhuruns

My Activity

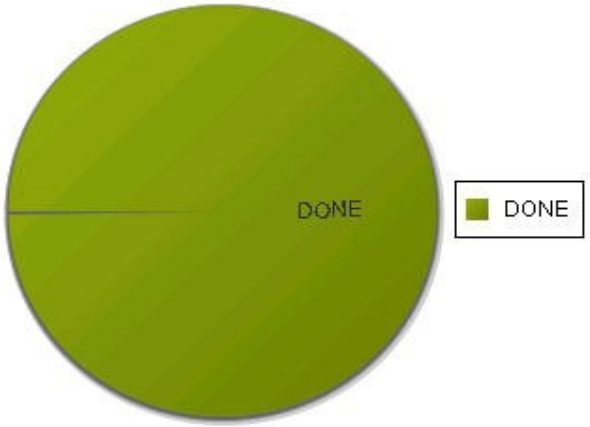
Getting Started | madhuruns

Inputs

Progress

Overall Run Progress

InputAndParseDiBayesFile		1/1 done
ParseLineOfDiBayesOutput		1/1 done
BioMartQuery		1/1 done



Results

Advantages

- Workflows are reusable
- Cloud computing infrastructure manages large data and processes – no need for big local resources
- Genomic analyses easy to run in parallel
- Simple submission through web interface for researchers
- Selecting ready-made workflows
- Simple and limited configuration of workflows
- Collaboration with industry – commercialisation of the services

MANCHESTER 1824

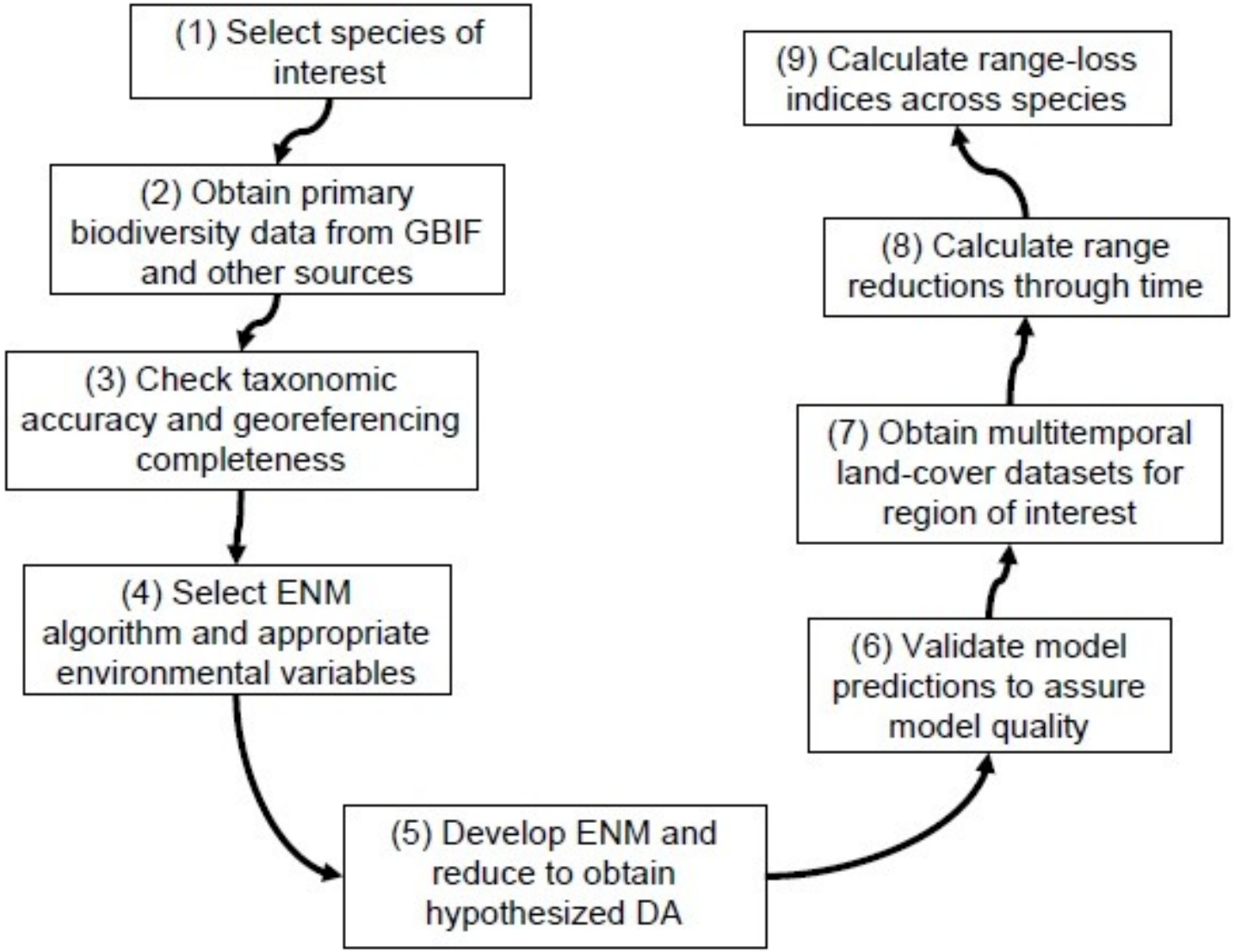
The University of Manchester

BioVel: Biodiversity Virtual e-Laboratory

- A network of expert scientists who develop, support, and use workflows and services in biodiversity
- Workflows, including:
  - Phylogenetics
  - Metagenomics
  - Ecological niche modelling
    - Species distribution modelling
    - Models how environmental niches of a species shift due to the changing climate.

myGrid

Case Study: Ecological Niche Modelling



Interaction Service: Communicating with your Remote Workflow

- Service suspends workflow execution to wait for further input from the user
- Interaction through the web interface
- Messages between workflow engine and web page via ATOM feeds, using Javascript

TAVERNA SERVER DEMO

MANCHESTER 1824

The University of Manchester

A RECAP ON TAVERNA WORKFLOWS

myGrid

Summary

Taverna Advantages

- Allows complex analysis pipelines
- Access to local and remote services (>8000 in biology)
- New services ‘added’ instantly
- Workflows can be shared and run in any Taverna instance
- Can be used for any areas of bio or non-bio research

Issues and Problems

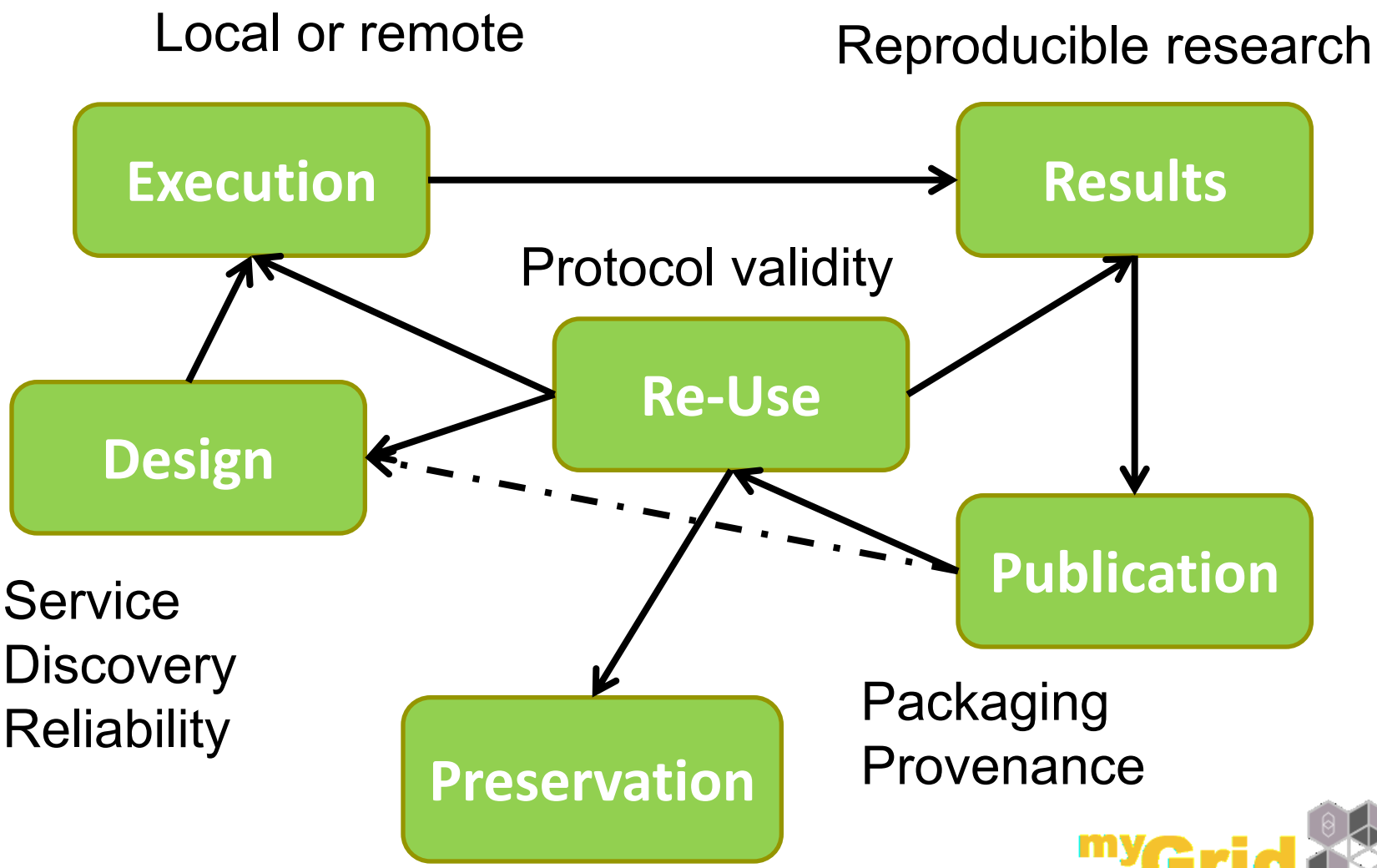
- Transferring large data over networks
  - Take services to data (like in the cloud example)
  - Pass by reference, rather than by value
  - Transfer only what you need for analysis
- Service incompatibility
  - shims – sharing and reusing
  - Creating integrated sets of services → components
- Services changing and vanishing
  - Use BioCatalogue and myExperiment to identify alternatives and find similar methods

Components

- A set of services designed to be compatible by
  - Consistent annotation to help understand how they work
  - Combining with shims to provide uniform (or predictable) input and output formats

- Hiding the complexity of public web services

Taverna Workflows Supporting *in silico* Science



Taverna 3 roadmap

- OSGi plugin system
- Workflow language: Scufl2
  - Making programmatic interaction easier
  - Compound format; embedding metadata, dependencies, independent API for creating/inspecting workflows
- Components
  - Finding/sharing command line tool descriptions
  - Richer way of finding compatible services

Summary – Workflow Advantages

- Informatics often relies on data integration and large-scale data analysis
- Workflows are a mechanism for linking together resources and analyses
- Automation
- Large data manipulation
- Promote reproducible research
- myExperiment allows you to reuse workflows and benefit from others work
- Easy to find and use successful analysis methods

More Information

- Taverna
  - http://www.taverna.org.uk
- myExperiment
  - http://www.myexperiment.org
- BioCatalogue
  - http://www.biocatalogue.org

MANCHESTER 1824

The University of Manchester

Acknowledgements

- myGrid consortium, in particular
  - Paul Fisher
  - Carole Goble
  - Alan Williams
  - Stian Soiland
  - Khalid Belhajjame
  - Rob Haines
  - Donal Fellows
  - Helen Hulme
- Trypanosomiasis project
  - Andy Brass
  - Paul Fisher
  - Harry Noyes

myGrid

MANCHESTER
1824

The University of Manchester