BioMedBridges
Software Tools Registry

Jon Ison, PhD
e-IRG Workshop
22-23 May, 2013, Trinity College, Dublin

www.biomedbridges.eu
Ten new biomedical sciences research infrastructures: stronger through common links

- FP7-funded cluster project
- 21 partners in 9 countries
- Bring together 10 emerging biological and medical research infrastructures on the ESFRI roadmap
- Each of the ten RIs again has up to 50 partners!
From Molecules to Medicine...
BioMedBridges Objectives

- **e-Infrastructure**
  - construct the e-infrastructure to allow interoperability between data and services in the biological, medical, translational and clinical domains

- **Data and services**
  - provide the computational ‘data and service’ bridges between individual biomedical infrastructures, to and linking the basic biological research and data to the clinical research and associated data.
BioMedBridges: Building data and service bridges across European infrastructures
BioMedBridges is about:

- Adding value to existing data by linking it
  - Creating links between available data that were not linked before will hugely increase the potential for new discoveries

- Bringing together different communities in the biological and medical sciences
  - create a common understanding of and approach to data (standards, formats, etc.... and how to make it linkable!)
What does BioMedBridges want to achieve?

- “Building data and service bridges in the life sciences”
  - What is needed for discovery?
  - What do researchers want/need?
  - What are the use cases and real world problems?
  - What data is involved?
  - What are the technical issues?
Registry Purpose

- Tool discovery - find, understand, compare, select
  - not (inter)operability in 1st instance

- “On ramp” for the working bioinformatician
  - but emphatically not just for geeks!

- Expose concise, practical information

- Highly streamlined, customisable, intuitive UI
Registry Scope

- Include applications for biological / clinical research
  - analytical tools and data resources
  - not system software (operating systems, IDEs etc.)

- Any software types that interest BioMB
  - Web services (REST-style, SOAP-based, RDF-based)
  - Web UI
  - Desktop GUI
  - Command-line tools
  - Shell scripts etc.
Biomedical software registry landscape

SADI: RDF

Biocatalogue: Webservices

AppDB: GRID-enabled

BMB Registry:

BioSWR: Moby Services

Other collections:

Middleware / cloud services
Tool Discovery

○ Primary purpose: to facilitate tool discovery across:
  o Biomedical research domains (molecules to systems)
  o Biomedical researchers (biologists, bioinformaticians, clinicians)
  o Types of software (webservices to scripts)
  o Analytical tools and data services
  o EU infrastructures (BMB participants and beyond)

○ Plans to participate in a pan-scientific registry (SciencePAD)
Design Principles

- Practical
  - support tool discovery use-cases

- Highly streamlined UI
  - faceted filtering / tree-based browsing
  - search box for free-text search

- Concise, valuable tool information
  - sufficient to handle necessary software complexity
  - emphasis on attribution, impact

- Use controlled vocabularies
  - SWO for general software concepts
  - ... which imports EDAM for bioinformatics-specific concepts
  - others (for biological context)

- Develop in sensible / agile way
  - Regular contact with users

- Implementation
  - JQWidgets widget set (jQuery / JavaScript), RESTful API
Software attribute
1 attribute / column

Software information
1 software / row

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Topic</th>
<th>Function</th>
<th>Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMBoss Water</td>
<td>Smith-Waterman local alignment of DNA sequences</td>
<td>Sequence alignment</td>
<td>Pairwise sequence alignment</td>
<td>ELIXIR</td>
</tr>
<tr>
<td>EMMA DB</td>
<td></td>
<td></td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>eMouseAtlas</td>
<td>The EMA Anatomy Atlas of Mouse Dissection Images</td>
<td></td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Empress</td>
<td>Database of SOPs, developed by the HubMap Project</td>
<td></td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>ENA Browser</td>
<td>Retrieve sequence and associated metadata</td>
<td>Nucleic acid sequence database</td>
<td>Retrieval</td>
<td>ELIXIR</td>
</tr>
<tr>
<td>Ensembl mouse</td>
<td>Mouse division of Ensembl</td>
<td>Genomeics</td>
<td></td>
<td>Infrafrontier</td>
</tr>
<tr>
<td>Ensembl</td>
<td>Genome databases for vertebrates</td>
<td>Genomics</td>
<td></td>
<td>ELIXIR</td>
</tr>
<tr>
<td>Ensembl Genomes</td>
<td>Extending Ensembl across the taxonomic hierarchy</td>
<td>Genomics</td>
<td></td>
<td>ELIXIR</td>
</tr>
<tr>
<td>Europe PMC</td>
<td>Literature information for European proteins</td>
<td>Literature and reference search</td>
<td>Literature and reference search</td>
<td>ELIXIR</td>
</tr>
<tr>
<td>Europhenome</td>
<td>Raw and annotated mouse phenotype</td>
<td></td>
<td></td>
<td>Infrafrontier</td>
</tr>
<tr>
<td>FASTA</td>
<td>Sequence similarity searching against databases</td>
<td>Data search and retrieval</td>
<td>Data search and retrieval</td>
<td>ELIXIR</td>
</tr>
<tr>
<td>FASTM</td>
<td>Peptide fragment searches using the FASTA model</td>
<td></td>
<td></td>
<td>ELIXIR</td>
</tr>
<tr>
<td>Flash Anonymizer</td>
<td>A comprehensive and efficient, open-source</td>
<td>(security)</td>
<td>Anonymization tool</td>
<td>BBMRI</td>
</tr>
<tr>
<td>Gene Expression Atlas</td>
<td>Enriched database of summary statistics</td>
<td>Functional genomics</td>
<td>Data retrieval</td>
<td>ELIXIR</td>
</tr>
<tr>
<td>GeneWise</td>
<td>Comprehensive sequence analysis tool for genes</td>
<td></td>
<td>Sequence alignment, Pairwise sequence alignment</td>
<td>ELIXIR</td>
</tr>
<tr>
<td>Genview</td>
<td>Genview is a Visualization Tool for large datasets</td>
<td>Visualization of large datasets</td>
<td></td>
<td>BBMRI</td>
</tr>
<tr>
<td>Grid BWA</td>
<td>Edito BWApaired-End Alignment, Aligner</td>
<td>Mapping</td>
<td>Short read mapping</td>
<td>ELIXIR</td>
</tr>
</tbody>
</table>
Filter columns (attributes)
Control what columns you see

Filter rows (software)
Control what rows you see
Filters may be set on any column

better faceting is a high priority!
Benefits

- Developers get credit for their work
- Helps scientists find tools
- Helps to avoid wasteful duplication of coding efforts
- Institutions & infrastructures get advert for their services
- Generates traffic to tool websites
- Save time (money) finding the right tools for the job
Development Vision

**Short-term**
- Basic features
- Better faceting
- BioMB software
- Handling complexity
- 1st BETA release

**Medium term**
- Basic content
- Free-text search
- Minimum info, standard
- Fine annotation (to info. standard)
- 1st MAJOR release

**Longer term**
- More features
- Standards & compatibility
- Standards & compatibility
- BioCatalogue import
- BioCatalogue import
- Federated content model
- And on ...

**Sustainability**
- Curation & QA
- Feature stability
- Feature stability
- DRCAT import
- Service versioning
- Social network / good-will

**And on ...**
- More content
- More features
- More content
- ... and more!
- ... and more!

**Service interoperability**
- Compatibility
- Service monitoring
- Service versioning
- ... and more!
Links & Contacts

○ Registry UI
  - http://wwwdev.ebi.ac.uk/fgpt/toolsui/

○ Code
  - https://github.com/jongithub/biomedbridges

○ jQuery UI Widgets
  - http://www.jqwidgets.com/

○ Contacts
  - Jon Ison (jison@ebi.ac.uk)
  - Julie McMurry (jmcmurry@ebi.ac.uk)

○ Partner contacts
  - crucial – please keep in touch / get involved

○ Workshops
  - AGM
  - Imperial College, UK (April 17)
  - CBS-DTU, Copenhagen (mid Sep)
Acknowledgements

- BioMedBridges partners (providers of feedback!)

- Matus Kalas, Inge Jonassen, University of Bergen, Norway (ELIXIR)

- Kristoffer Rapacki, Søren Brunak, CBS-DTU, Denmark (ELIXIR tools node)

- Steve Pettifer, Carole Goble, Rob Haines – Manchester University, UK (ELIXIR)

- EBI External Services Team