BioMedBridges & ELIXIR
Software Tools Registry

Jon Ison, PhD
Tools Workshop
16-18 Oct, 2013. CBS, Copenhagen

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:**
  - Biomedical research domains (molecules to systems)
  - Biomedical researchers (biologists, bioinformaticians, clinicians)
  - Types of software (webservice to scripts)
  - Analytical tools and data services
  - 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
<table>
<thead>
<tr>
<th>Collection</th>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBI databases</td>
<td>EMBL tools</td>
<td></td>
<td>Protein and nucleotide sequence rep</td>
</tr>
<tr>
<td></td>
<td>EMBOSB</td>
<td></td>
<td>Retrieve data from a dictionary of m.</td>
</tr>
<tr>
<td></td>
<td>Clustal Omega</td>
<td></td>
<td>Multiple sequence alignment using g.</td>
</tr>
<tr>
<td></td>
<td>ClustalW2</td>
<td></td>
<td>General purpose global multiple seq.</td>
</tr>
<tr>
<td></td>
<td>ClustalW2 Phylogeny</td>
<td></td>
<td>Phylogenetic tree generation from a</td>
</tr>
<tr>
<td></td>
<td>Phylogeny</td>
<td></td>
<td>Search a protein structure for likely...</td>
</tr>
<tr>
<td></td>
<td>DataCenter</td>
<td></td>
<td>Phylogenetic tree construction</td>
</tr>
<tr>
<td></td>
<td>EMBL</td>
<td></td>
<td>Structural analysis of protein</td>
</tr>
<tr>
<td></td>
<td>Global Aligner</td>
<td></td>
<td>Multiple sequence alignment</td>
</tr>
<tr>
<td></td>
<td>Global Phylogeny</td>
<td></td>
<td>Phylogenetic tree construction</td>
</tr>
<tr>
<td></td>
<td>OAE</td>
<td></td>
<td>Data retrieval</td>
</tr>
<tr>
<td></td>
<td>OAE resolver</td>
<td></td>
<td>Find a gene, transcript etc. in EMB W.</td>
</tr>
<tr>
<td></td>
<td>EMBL Backtrans</td>
<td></td>
<td>Back-Translate a protein sequence t.</td>
</tr>
<tr>
<td></td>
<td>EMBL Backtrans</td>
<td></td>
<td>Back-Translate a protein sequence t.</td>
</tr>
<tr>
<td></td>
<td>EMBL capstd</td>
<td></td>
<td>Identification of potential CpG islands</td>
</tr>
<tr>
<td></td>
<td>EMBL reshold</td>
<td></td>
<td>Plot potential isochore features</td>
</tr>
<tr>
<td></td>
<td>EMBL matchcr</td>
<td></td>
<td>Waterman-Eggert local alignment of.</td>
</tr>
<tr>
<td></td>
<td>EMBL needle</td>
<td></td>
<td>Needleman-Wunsch global alignement</td>
</tr>
<tr>
<td></td>
<td>EMBL nanoprep</td>
<td></td>
<td>Identification of potential CpG islands</td>
</tr>
</tbody>
</table>

**Main menu**
- Partner contact details
- Scientific use-cases

**Grid menu**
- Show attributes
- Clear filters

**Search box**
- Free-text search over all information

**Customisable information grid**
- Faceted search and browsing
<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioMart</td>
<td>Database</td>
<td>Protein and nucleotide sequence records</td>
<td>Sequence composition, sequence analysis, protein detection</td>
</tr>
<tr>
<td>ChEBI</td>
<td>Database</td>
<td>Retrieve data from a dictionary of molecules</td>
<td>Small molecules, small molecule annotation</td>
</tr>
<tr>
<td>Clustal Omega</td>
<td>Tool</td>
<td>Multiple sequence alignment using global methods</td>
<td>Sequence alignment, multiple sequence alignment construction</td>
</tr>
<tr>
<td>ClustalW</td>
<td>Tool</td>
<td>General purpose global multiple sequence alignment</td>
<td>Sequence alignment, multiple sequence alignment construction</td>
</tr>
<tr>
<td>ClustalW Phylogeny</td>
<td>Tool</td>
<td>Phylogenetic tree generation from alignments</td>
<td>Phylogenetics</td>
</tr>
<tr>
<td>CSS</td>
<td>Tool</td>
<td>Search a protein structure for likeliness</td>
<td>Protein structure alignment</td>
</tr>
<tr>
<td>Dalilite</td>
<td>Tool</td>
<td>Pairwise structure alignment</td>
<td>Pairwise structure alignment</td>
</tr>
<tr>
<td>DASTy</td>
<td>Tool</td>
<td>Visualization and integration of protein structures</td>
<td>Visualize protein structure, protein alignment</td>
</tr>
<tr>
<td>DbClustal</td>
<td>Tool</td>
<td>Global multiple sequence alignment</td>
<td>Sequence alignment, multiple sequence alignment construction</td>
</tr>
<tr>
<td>dbfetch</td>
<td>Tool</td>
<td>Dbfetch provides an easy way to retrieve database</td>
<td>Data search and retrieval</td>
</tr>
<tr>
<td>eB-Eye</td>
<td>Tool</td>
<td>Easy and uniform access to the database</td>
<td>Data retrieval</td>
</tr>
<tr>
<td>EG ID Resolver</td>
<td>Tool</td>
<td>Find a gene, transcript etc in EG web</td>
<td>Translation</td>
</tr>
<tr>
<td>EMBOSS backtrans</td>
<td>Tool</td>
<td>Back-translate a protein sequence to DNA</td>
<td>Translation</td>
</tr>
<tr>
<td>EMBOSS backtras</td>
<td>Tool</td>
<td>Back-translate a protein sequence to DNA</td>
<td>Translation</td>
</tr>
<tr>
<td>EMBOSS cogplot</td>
<td>Tool</td>
<td>Identification of potential CpG islands</td>
<td>Sequence composition, CpG island and isochore detection</td>
</tr>
<tr>
<td>EMBOSS matcher</td>
<td>Tool</td>
<td>Waterman-Eggert local alignment</td>
<td>Sequence alignment, isochore sequence alignment construction</td>
</tr>
<tr>
<td>EMBOSS needle</td>
<td>Tool</td>
<td>Needleman-Wunsch global alignment</td>
<td>Sequence alignment, multiple sequence alignment construction</td>
</tr>
<tr>
<td>EMBOSS newspaper</td>
<td>Tool</td>
<td>Identification of potential CpG islands</td>
<td>Sequence composition, CpG island and isochore detection</td>
</tr>
</tbody>
</table>

**Software attribute**

1 attribute / column

**Software information**

1 software / row
Filter rows (facets)
Control what rows you see

Filter columns (attributes)
Control what columns you see

Filter rows (software)
Control what rows you see
Filters available on all columns
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
Links & Contacts

- Registry UI
  - http://tinyurl.com/bmbtoolsui
- 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!)
- Kristoffer Rapacki, Søren Brunak, CBS-DTU, Denmark (ELIXIR tools node)
- Matus Kalas, Inge Jonassen, University of Bergen, Norway (ELIXIR)
- Steve Pettifer, Carole Goble, Rob Haines – Manchester University, UK (ELIXIR)
- EBI External Services Team