Open bridges for life-science data

17-18 November 2015

Janet Thornton
Meeting the Grand Challenges

- Health
  - Changing demographics
  - Healthy ageing – extending healthspan
  - Infectious diseases

- Agricultural
  - Feeding a growing population
  - Food security
  - Managing in changing climates

- Environment
  - Maintaining biodiversity
  - Biofuels
From Molecules to Medicine...
Impacts for Medicine

1. Molecular Basis of Disease
2. Human Variation and disease risk
3. Cancer Genomics
4. Tracking the source of Infectious Diseases
5. Tracking food-borne pathogens
Impact on Agriculture and the Environment

- Coping with increasing populations and changing climatic conditions
- Developing better crop strains
- Improving farm animal strains
- Mapping diversity
- Biofuels

BioMedBridges
Biomedical research requires increasingly sophisticated infrastructure

- Early studies were completed without any formal infrastructure (single use) or only using basic, local infrastructure
- Research that is done now and in the future requires robust infrastructure that is
  - large-scale
  - non-trivial
  - and interconnected across disciplines, data types and resources, and countries

BioMedBridges
A very diverse community
Ten new biomedical sciences research infrastructures: stronger through common links

- Computational ‘data and service’ bridges between the BMS RIs
- Interoperability between data and services in the biological, medical, translational and clinical domains
- Link basic biological research data with clinical research and associated data
Disciplines covered
BioMedBridges

Providing data and service bridges between the biomedical sciences research infrastructures

www.biomedbridges.eu
BioMedBridges has delivered:

○ **Data bridges**
  - Creating links between available data that were not linked before will hugely increase the potential for new discoveries

○ **Interoperability bridges**
  - standards, formats, ontologies... and how to make it linkable!

○ **Social bridges**
  - Connecting the biomedical research infrastructures
Beyond data availability and accessibility!

“Building data and service bridges in the life sciences”

- What is needed for scientific discovery?
- What do researchers want/need?
- What are the use cases and real world problems?
- What data is involved?
- What are the technical issues?
Standards description and harmonization

- To link data between different domains, they must use:
  - common identifiers
  - harmonised content, syntax and semantics

- BioMedBridges has created:
  - an online dictionary of common molecular identifiers of the BMS research infrastructures
  - a registry of standards used
  - A service registry
Bridging communities

- **Images**: cell→tissue→organism: enable drug target and biomarker discovery for human disease
- **Translation**: between mouse and human phenotypes
- **Personalised medicine**: treatment options
- **Biomolecular volume data**: analyse and annotate
- **Linking**: disease-related data and terminology; biosample discovery
Ethics requirements, data protection and secure access

- BioMedBridges has built a security framework that:
  - addresses the ethical, legal and regulatory issues resulting from sharing data and providing access to biomaterials
  - is in compliance with national and European regulations, privacy rules and access requirements
Technical bridges

• “Secure Access” and “Standards” facilitate the effective sharing of data
• “Technical integration” provides the e-infrastructure

Standards description & harmonisation

Technical integration

Secure access
Five inter-related use cases

- To test the technical and data developments of the construction work packages
- Each use case coordinated by a particular ESFRI Research Infrastructure

Interoperability of large scale image data sets

Improving the link between mouse models and human data

Personalized Medicine

Integrating structural data

Integrating disease-related data and terminology
Speakers

- Jeffrey BARRETT, Centre for Therapeutic Target Validation
- Ewan BIRNEY, European Bioinformatics Institute (EMBL-EBI)
- Phil BOURNE, National Institutes of Health, BD2K
- Søren BRUNAK, ELIXIR Denmark
- Kate BUSHBY, Newcastle University, RD Connect
- Edwin CUPPEN, Hubrecht Institute
- Jim DAVIES, University of Oxford, Genomics England
- Jan ELLENBERG, European Molecular Biology Laboratory, Euro-BioImaging
- Helen FIRTH, University of Cambridge, Deciphering Developmental Disorders
- Alex GUTTERIDGE, Pfizer
- Andrew LEACH, GSK
- Peter LUIJTEN, University Medical Center Utrecht, Center for Translational Molecular Medicine
- Ruth MARCH, Innovative Medicines & Early Development, AstraZeneca
- Willem OUWEHAND, University of Cambridge, Blueprint Epigenome
- Janet THORNTON, European Bioinformatics Institute (EMBL-EBI)
- Mathias UHLÉN, SciLifeLab
Workshops

- Unlocking low-resolution structural biology data (INSTRUCT)
- Translational research infrastructure: submit, use, manage and combine imaging, clinical and molecular data (EATRIS/BBMRI/Euro-BioImaging)
- Data sharing for advancing health (ECRIN-ERIC/BBMRI-ERIC)
- Challenges in interoperability and semantic data integration - Lessons learned from BioMedBridges and OpenPhacts (ELIXIR)
- New resources and tools to support the use of mouse data for the study of human diseases (INFRAFRONTIER)
- Interoperability of cellular phenotypic data derived from microscopy images (Euro-BioImaging)
- Security challenges and solutions for sharing sensitive data in an open data environment (BBMRI/ECRIN)
- Metagenomics: bridging between environment and life (EMBRC/ELIXIR)
- An integrated platform connecting databases, registries and omics data for rare disease research (RD Connect)
- Journal data policies and practices: practical considerations for researchers (EMBL-EBI/ELIXIR/NPG Scientific Data)
- Better data-level metrics: quality and impact (F1000 Research/CERN/CASRAI)
- Challenges in international Ethics Review equivalency (Global Alliance for Genomics and Health)