

MetaboLights: Towards a new COSMOS of metabolomics data management

Kenneth Haug¹, Reza M. Salek^{1,2}, Pablo Conesa¹, Paula de Matos¹,
Eamonn Maguire³, Tejasvi Mahendrakar¹, Philippe Rocca-Serra³,
Susanna-Assunta Sansone³, Julian L. Griffin² and Christoph Steinbeck^{1,*}
¹*European Bioinformatics Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridgeshire, CB10 1SD*; ²*Elsie Widdowson Laboratory,
Fulbourn Road, Cambridge, CB1 9NL, UK, University of Cambridge,
Department of Biochemistry, Cambridge CB2 1QW* ; ³*Oxford e-Research
Centre, University of Oxford, Oxford, UK.*
*steinbeck@ebi.ac.uk

Abstract: Exciting funding initiatives are emerging in Europe and the US for metabolomics data production, storage, dissemination and analysis. This is based on a rich ecosystem of resources around the world, which has been build during the past ten year, including but not limited to resources such as MassBank in Japan and the Human Metabolome Database (HMDB) in Canada. Now, the European Bioinformatics Institute (EBI) has launched MetaboLights a database for metabolomics experiments and the associated meta-data (<http://www.ebi.ac.uk/metabolights>). It is the first comprehensive, cross-species, cross-platform metabolomics database maintained by one of the major open access data providers in molecular biology. In October, the European COSMOS consortium will start its work on Metabolomics data standardization, publication and dissemination workflows. The NIH in the US is establishing 6-8 metabolomics services cores as well as a national metabolomics repository. This paper reports about MetaboLights as a new resource for Metabolomics research, summarises the related developments and outlines how they may consolidate the knowledge management in this third large omics field next to proteomics and genomics.

1 Introduction

Metabolomics has become an important phenotyping technique for molecular biology and medicine. It assesses the molecular state of an organism or collections of organisms through the comprehensive quantitative and qualitative analysis of all small molecules in cells, tissues, and body fluids. Metabolic processes are at the core of physiology. Consequently, metabolomics is ideally suited as a medical tool to characterize disease

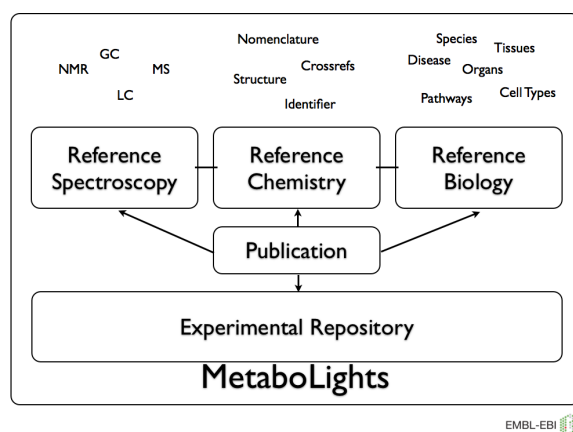
states in organisms, as a tool to assessment of organism for their suitability in, for example, renewable energy production or for biotechnological applications in general. In addition application of metabolomics in environmental science, toxicology, food and medical industry is well established, growing and documented. Metabolomics studies generate large amounts of analytical data (Giga- to Terabytes depending on the size of the study) and therefore impose significant challenges for biomedical and life science e-infrastructures to cope with such data volumes and ensure that the data is captured, stored and disseminated based on open and widely accepted community standards. Years after the first standardisation exercises [FRGea07, TFSea08], metabolomics is now reaching the state of a mature analytical technique as indicated by the establishment of 6-8 Regional Comprehensive Metabolomics Resource Cores (RCMRCs) by the NIH in the United States. In addition, we are now facing a rich ecosystem of specialised metabolomics databases as well as the first general metabolomics repositories and databases emerging. In Europe, the COSMOS consortium of 14 leading laboratories in metabolomics will begin its work on standards, data management and dissemination in Metabolomics. Here, we outline these developments and show how they may consolidate the knowledge management in this third large omics field next to proteomics and genomics.

2 MetaboLights – A cross-species repository for metabolomics experiments

The European Bioinformatics Institute (EBI) has recently launched MetaboLights, a database for metabolomics experiments and the associated metadata. It is the first comprehensive, cross-species, cross-platform metabolomics database maintained by one of the major open access data providers in molecular biology. MetaboLights lives at <http://www.ebi.ac.uk/metabolights>. For their convenience, users can also use metabolights.org, metabolights.net and metabolights.eu. The EBI ensures long-term stability and maintenance of the resource. Like all other EBI resources, the MetaboLights database is completely open to the public, including open access to the data. Data are made available in publicly accepted open standards compliance with MIBBI (The Minimum Information for Biological and Biomedical Investigations) [TFSea08]. The software is open source and adheres to the promotion of open source file formats, such as mzML and nmrML. One of the main submission channels

for MetaboLights' use is the ISA Tools Suite [SRSFea12]. MetaboLights is not intended to replace specialist resources for Metabolomics. Rather, it will build on prior art and collaborate. We are dedicated to close collaboration with all major parties involved in the creation of this prior art, such as the Metabolomics Society, Metabomeeting and the Metabolomics Standards Initiative (MSI). MetaboLights aims to agree on formal data sharing agreements with major resources such as the Human Metabolome Database, the Golm Metabolome Database and the Rikken Metabolomics Platform. Currently we house selection of experimental raw data and their associated metadata for different platforms such as NMR, GC-MS and LC-MS.

Figure 1: MetaboLights general outline



3 Outlook

In October, the European COSMOS (COordination of Standards in Meta-bOlogicS) consortium will start its work on Metabolomics data standardization, publication and dissemination workflows. It is the aim of COSMOS to develop efficient policies to ensure that Metabolomics data is

1. Encoded in open standards to allow barrier-free and widespread analysis.
2. Tagged with a community-agreed, complete set of metadata (mini-

mum information standard).

3. Supported by a communally developed set of open source data management and capturing tools.
4. Disseminated in open-access databases adhering to the above standards.
5. Supported by vendors and publishers, who require deposition upon publication
6. Properly interfaced with data in other biomedical and life science e-infrastructures, such as
 - ELIXIR (<http://www.elixir-europe.org/>),
 - BioMedBridges (<http://www.biomedbridges.eu/>),
 - EU-OPENSOURCE (<http://www.eu-openscreen.de/>) and
 - BBMRI (<http://www.bbmri.eu/>).

During 2012, MetaboLights' repository layer will be expanded by a reference layer with chemical, spectroscopic and biological reference information about individual metabolites (Figure 1).

The NIH in the US is establishing 6-8 metabolomics services cores as well as a national metabolomics repository. Together with similar initiatives in Australia, Japan and hopefully more emerging over time, this opens the door for a global network of metabolomics data collection, exchange and dissemination.

References

- [FRGea07] Oliver Fiehn, Don Robertson, Jules Griffin, and et al. The metabolomics standards initiative (MSI). *Metabolomics*, 3(3):175–178, 2007.
- [SRSFea12] Susanna-Assunta Sansone, Philippe Rocca-Serra, Dawn Field, and et al. Toward interoperable bioscience data. *Nat Genet*, 44(2):121–126, January 2012.
- [TFSea08] Chris F Taylor, Dawn Field, Susanna-Assunta Sansone, and et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nature Biotechnology*, 26(8):889–896, 2008.