

MetaboLights: towards a new COSMOS of metabolomics data management

Christoph Steinbeck · Pablo Conesa · Kenneth Haug · Tejasvi Mahendrakar · Mark Williams · Eamonn Maguire · Philippe Rocca-Serra · Susanna-Assunta Sansone · Reza M. Salek · Julian L. Griffin

Received: 9 July 2012 / Accepted: 30 August 2012 / Published online: 25 September 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

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 years, including but not limited to resources such as MassBank in Japan and the Human Metabolome Database in Canada. Now, the European Bioinformatics Institute has launched MetaboLights, a database for metabolomics experiments and the associated metadata (<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 communication 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.

Keywords Metabolomics · Databases · ISA-Tab · ISA commons

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 for assessment of organisms 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 are captured, stored and disseminated based on open and widely accepted community standards. Years after the first standardisation exercises (Fiehn et al. 2007; Taylor et al. 2008), metabolomics is now reaching the state of a mature analytical technique as indicated by the establishment of 6–8 Regional Comprehensive

C. Steinbeck (✉) · P. Conesa · K. Haug · T. Mahendrakar · M. Williams
European Bioinformatics Institute (EMBL-EBI),
Wellcome Trust Genome Campus, Hinxton, Cambridgeshire
CB10 1SD, UK
e-mail: steinbeck@ebi.ac.uk

E. Maguire · P. Rocca-Serra · S.-A. Sansone
Oxford e-Research Centre, University of Oxford, Oxford, UK

R. M. Salek · J. L. Griffin
Elsie Widdowson Laboratory, Fulbourn Road, Cambridge
CB1 9NL, UK

R. M. Salek · J. L. Griffin
Department of Biochemistry, University of Cambridge,
Cambridge CB2 1QW, UK

Metabolomics Resource Cores (RCMRCs) by the NIH in the United States (<http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-11-016.html>). In addition, we are now facing a rich ecosystem of specialised metabolomics databases, such as (Wishart et al. 2007; Kopka et al. 2005; Smith et al. 2005; Skogerson et al. 2011) as well as the first general metabolomics repositories (<http://www.ebi.ac.uk/metabolights>) 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 (EMBL-EBI) has recently launched MetaboLights, a database for metabolomics experiments and the associated metadata. It aims to become the first comprehensive, cross-species, cross-platform metabolomics database maintained by one of the major open access data providers in molecular biology. The EBI ensures long-term stability and maintenance of the resource. Deposited datasets are assigned a stable identifier of the form MTBLS1 (the first dataset ever deposited in MetaboLights). These identifiers, like other stable identifiers in bioinformatics, can be used to mark datasets in publications or merge data in systems biology applications.

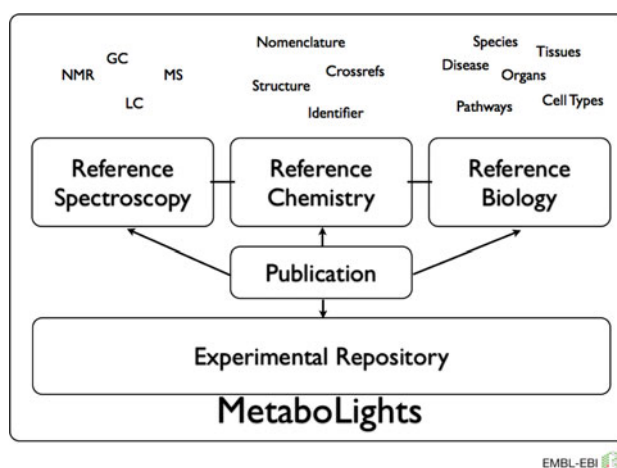


Fig. 1 MetaboLights general outline with repository and reference layer. The reference layer is work currently in progress

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 compliant with community standards (BioSharing: http://biosharing.org/standards_view), including Minimum Information for Biological and Biomedical Investigations (MIBBI) checklists (Taylor et al. 2008). The software is open source and adheres to the promotion of open source file formats, such as mzML and nmrML. MetaboLights will ultimately consist of a reference layer on top of the repository layer. The reference layer will contain information about individual metabolites and their chemical, analytical and biological properties. The repository

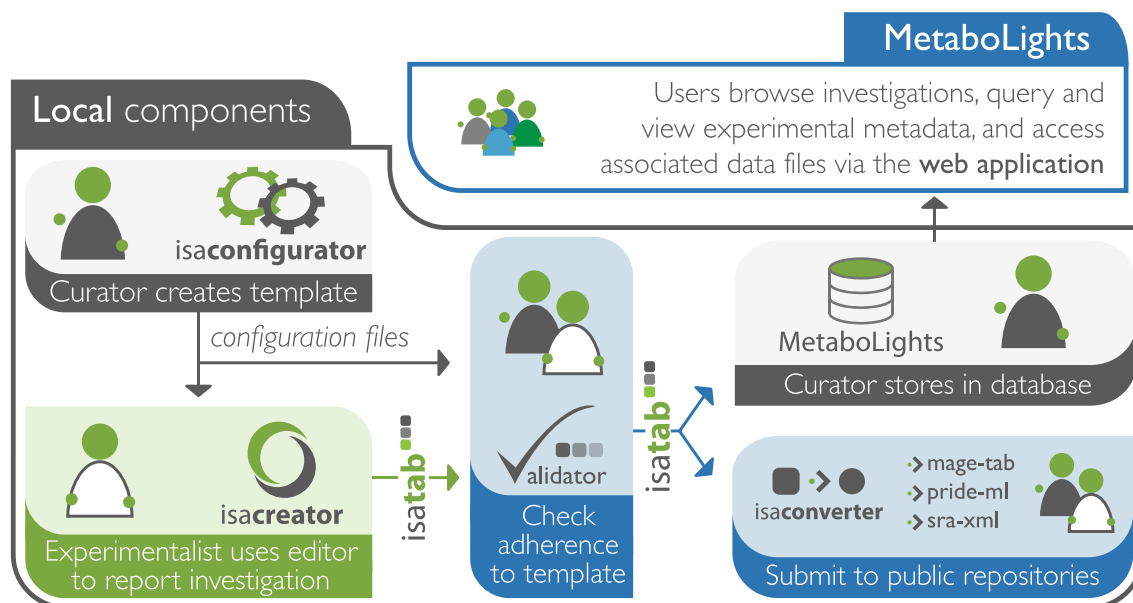


Fig. 2 MetaboLights data submission workflow

The screenshot shows the MetaboLights search interface. At the top, there is a search bar with the text "Metabolights" and a search button. To the right of the search bar are links for "home", "browse", "download", "submit", and "log in". Below the search bar, a green banner indicates "2 Search result(s) showing 1 to 2".

On the left side, there is a "Filter by" section with two categories: "Organism" and "Technology".

- Organism filters:**
 - Arabidopsis thaliana (thale cress)
 - Bovinae
 - Caenorhabditis elegans
 - Homo sapiens (Human)
 - Saccharomyces cerevisiae
- Technology filters:**
 - mass spectrometry
 - NMR spectroscopy

The search results are displayed in two boxes:

Result 1: High-resolution extracted ion chromatography, a new tool for metabolomics and lipidomics using a second-generation orbitrap mass spectrometer
Study Public Release Date: 14-Feb-2012
Organism: Homo sapiens (Human)
Study Factors: healthy volunteers: [Accepts Healthy Volunteers Indicator]; blood plasma: [plasma ringer gluc:plasma:plasma ringer palm:plasma ringer cer:plasma ringer gluc cer palm:plasma ringer]
Assays: mass spectrometry (60)
Study Identifier: MTBLS4
 Submitted by Albert Koulman

Result 2: NMR based metabolomics of Human Type 2 Diabetes urine samples
Study Public Release Date: 14-Feb-2012
Organism: Homo sapiens (Human)
Study Factors: Metabolic syndrome: [type 2 diabetes mellitus:Control Group]; Gender: [Female Gender(Female_Gender):Male Gender(Male_Gender)]
Assays: NMR spectroscopy (132)
Study Identifier: MTBLS1
 Submitted by Reza Salek

Fig. 3 MetaboLights search results

later, which has been launched and is fully operational, contains primary research data from published metabolomics studies, annotated with meta data (Fig. 1). One of the main submission channels for MetaboLights is the ISA Tools Suite (Fig. 2) (Sansone et al. 2012).

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, Metabolic Profiling Forum (Metabomeeting) and the Metabolomics Standards Initiative (MSI). MetaboLights aims at 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 a selection of experimental raw data and their associated metadata for different platforms such as NMR, GC-MS and LC-MS (Fig. 3). The repository layer is generally open to any data that was used in a metabolomics study. That could include, for instance, flux data (temporal measurements with ^{13}C), spatial maps, and IR and Raman fingerprint data.

2.1 Call for submitting data

MetaboLights is now ready for receiving metabolomics datasets. We have, for example, recently received the

validation dataset measured by O'Callaghan *et al.* for validating their PyMS software (O'Callaghan et al. 2012). We think that this is the way forward for sharing gold standard datasets for validating metabolomics software. Generally, we hope, and will work towards this with journal editors, that the submission of datasets used to justify findings in publications will be submitted to the MetaboLights or one of the emerging collaborating repositories. Interested readers are encouraged to go to <http://www.ebi.ac.uk/metabolights/presubmit> and submit their data. The MetaboLights team is happy to assist in this process.

3 Conclusion and outlook

Here, we have reported the publication of MetaboLights, the first 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, readers can use the URL's metabolights.org, metabolights.net and metabolights.eu. In October, the European COSMOS (COordination of Standards in MetabOlogicS) 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 are

1. Encoded in open standards to allow barrier-free and widespread analysis.
2. Tagged with a community-agreed, complete set of metadata (minimum 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, Bio-MedBridges, EU-OPENSOURCE and BBMRI).

COSMOS will also strive to harmonize the European agenda with efforts in US, where the NIH 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.

Acknowledgments The authors gratefully acknowledge funding of this work by the BBSRC MetaboLights Grant BB/I000933/1 and the European Commission COSMOS Grant EC312941. The authors are also extremely grateful to the participants of the initial metabolites planning workshops at the European Bioinformatics Institute (EMBL-EBI) as well as to those collaborators who contributed Metabolomics datasets in the early stages of the MetaboLights launch.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Fiehn, O., Robertson, D., Griffin, J. et al. (2007). The metabolomics standards initiative (MSI). *Metabolomics*, 3(3), 175–178.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmuller, E., Dormann, P., Weckwerth, W., Gibon, Y., & Stitt, M. (2005). GMD@ CSB. DB: the Golm metabolome database. *Bioinformatics*, 21(8), 1635–1638.
- O’Callaghan, S., DeSouza, D. P., Isaac, A., Wang, Q., Hodkinson, L., Olshansky, M., Erwin, T., Appelbe, B., Tull, D. L., Roessner, U., Bacic, A., McConville, M. J., Likic, V. A. (2012). PyMS: a Python toolkit for processing of gas chromatography–mass spectrometry (GC–MS) data. Application and comparative study of selected tools. *BMC Bioinformatics*, 13(1), 115.
- Sansone, S.-A., Rocca-Serra, P., Field, D., et al. (2012). Toward interoperable bioscience data. *Nature Genetics*, 44(2), 121–126.
- Skogerson, K., Wohlgemuth, G., Barupal, D. K., & Fiehn, O. (2011). The volatile compound BinBase mass spectral database. *BMC Bioinformatics*, 12, 321.
- Smith, C. A., O’Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., Siuzdak, G. (2005). METLIN: A metabolite mass spectral database. *Therapeutic Drug Monitoring*, 27(6), 747–751.
- Taylor, C. F., Field, D., Sansone, S.-A. et al. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nature Biotechnology*, 26(8), 889–896.
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.-A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D. D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G. E., Macinnis, G. D., Weljie, A. M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B. D., Vogel, H. J., Querengesser, L. (2007). HMDB: The human metabolome database. *Nucleic Acids Research*, 35(Database), D521–D526.