

MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data

Kenneth Haug¹, Reza M. Salek^{1,2,3}, Pablo Conesa¹, Janna Hastings¹, Paula de Matos¹, Mark Rijnbeek¹, Tejasvi Mahendraker¹, Mark Williams¹, Steffen Neumann⁴, Philippe Rocca-Serra⁵, Eamonn Maguire⁵, Alejandra González-Beltrán⁵, Susanna-Assunta Sansone⁵, Julian L. Griffin^{2,3} and Christoph Steinbeck^{1,*}

¹European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, ²MRC HNR, Elsie Widdowson Laboratory, Fulbourn Road, Cambridge CB1 9NL, ³Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, Cambridge CB2 1GA, UK, ⁴Department of Stress- and Developmental Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle, Germany and ⁵Oxford e-Research Centre, University of Oxford, 7 Keble Road, Oxford OX1 3QG, UK

Received August 15, 2012; Revised September 22, 2012; Accepted October 1, 2012

ABSTRACT

MetaboLights (<http://www.ebi.ac.uk/metabolights>) is the first general-purpose, open-access repository for metabolomics studies, their raw experimental data and associated metadata, maintained by one of the major open-access data providers in molecular biology. Metabolomic profiling is an important tool for research into biological functioning and into the systemic perturbations caused by diseases, diet and the environment. The effectiveness of such methods depends on the availability of public open data across a broad range of experimental methods and conditions. The MetaboLights repository, powered by the open source ISA framework, is cross-species and cross-technique. It will cover metabolite structures and their reference spectra as well as their biological roles, locations, concentrations and raw data from metabolic experiments. Studies automatically receive a stable unique accession number that can be used as a publication reference (e.g. MTBLS1). At present, the repository includes 15 submitted studies, encompassing 93 protocols for 714 assays, and span over 8 different species including *human*, *Caenorhabditis elegans*, *Mus musculus* and *Arabidopsis thaliana*. Eight hundred twenty-seven of the metabolites identified in these studies have been mapped to ChEBI. These studies cover a variety of techniques, including NMR spectroscopy and mass spectrometry.

INTRODUCTION

Metabolomics is the systematic study of the small molecular metabolites in a cell, tissue, biofluid or cell culture media that are the tangible result of cellular processes or responses to an environmental stress (1,2). The identification and quantification of such metabolites provide unique insights into the metabolic processes that are taking place in the cellular environment. Metabolic profiles taken from body fluids have the potential to act as biomarkers for many different diseases, an approach that has already shown value in, for example, heart disease and diabetes (3), the effects of diet (4) and interactions with the environment (5). Metabolomics technologies yield many insights into basic biological research in areas such as systems biology and metabolic modeling (6), pharmaceutical research (7), nutrition (8) and toxicology (9). However, to harness the full potential of metabolomics, researchers need access to data and knowledge to compare, contrast and make inferences from the results they obtain in their experiments (10). The metabolome is the total complement of metabolites present in a biological sample under given genetic, nutritional or environmental conditions. Since such conditions can vary dramatically, it is clear that databases will need to collect numerous experiments together for a given species to accurately reflect the underlying diversity and complexity. In recent years, several instrument or species-specific dedicated metabolomics reference databases have been created. Examples include the Human Metabolome Database [HMDB, <http://www.hmdb.ca>, (11)], the Biological

*To whom correspondence should be addressed. Tel: +44 1223 492640; Fax: +44 1223 494468; Email: steinbeck@ebi.ac.uk

The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors.

© The Author(s) 2012. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

Magnetic Resonance Data Bank [BMRB, <http://www.bmrw.wisc.edu>, (12)], METLIN [<http://metlin.scripps.edu>, (13)], LIPIDMAPS [<http://www.lipidmaps.org>, (14)] and more general databases such as KNApSAcK (<http://kanaya.aist-nara.ac.jp/KNApSAcK/>). However, the various metabolomics communities worldwide have not had a global open repository to share experimental data and associated metadata across species and platforms. MetaboLights will (i) provide a single point of access to worldwide data and knowledge in metabolomics, (ii) facilitate the development and adoption of a common data sharing format, (iii) ensure data traceability and reproducibility and (iv) progressively promote interoperability across existing resources.

MetaboLights consists of two distinct layers: a *repository*, enabling the metabolomics community to share findings, data and protocols for any form of metabolomics study, and a *reference layer* of curated knowledge about metabolites (forthcoming). MetaboLights is not intended to replace specialist resources but is specifically designed to build on prior art and extensively collaborate with the existing databases to ensure that data are exchanged and that assimilation efforts target gaps in worldwide available knowledge. 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) (15). MetaboLights is working towards the setup of formal data sharing agreements with major resources such as the HMDB, the Golm Metabolome Database (16), MetabolomeExpress (17) and the Riken Metabolomics Platform (18). MetaboLights contains references to identified metabolites in existing databases, such as HMDB and ChEBI (19), and does not duplicate compound information residing in these external databases. Rather, it uses programmatic access to retrieve relevant data to display a unified metabolite-centric view to our users. In the future, such metabolite-centric views will be extended to show metabolites in the context of pathways, harnessing the Reactome database of biochemical pathways (20). In this article, we report on the structure and content of the MetaboLights repository and describe on-going work in the development of the reference layer.

DATABASE DESCRIPTION

The MetaboLights repository can be accessed at <http://www.ebi.ac.uk/metabolights> and <http://metabolights.org>.

Database content

We store and display an extensive set of associated information for studies in MetaboLights. This includes submitter and author information, publication references, the study design, protocols applied, names of data files included, platform information and metabolite information. The metabolite information includes a description, external database identifiers, formula and intensity or concentration, and where the metabolite was identified in the sample.

At present, the repository includes 15 submitted studies, of which 10 are publicly visible. These studies encompass 93 protocols for 714 assays, and span over 8 different types of organism including human, *Caenorhabditis elegans*, *Mus musculus* and *Arabidopsis thaliana*. Eight hundred twenty-seven of the metabolites identified in these studies have been mapped to ChEBI and 136 to HMDB. Thirty-eight users are currently registered.

Technical architecture

The MetaboLights repository is based on open source freely available software and tools. The web application runs on an Apache Tomcat server and the database backend is an Oracle database, but other standard SQL databases like MySQL and PostgreSQL can be used.

At the core of the database implementation is the ISA framework (21). The main database schema is powered by the ISA BioInvestigation Index (BII), which contains user information and all searchable metadata for the studies. Currently, there are 72 tables in this database schema. Any data-files that are associated with a study are stored on a traditional file system, and only their reference is stored in the database. Each study has a separate folder on the file systems containing the study metadata and associated files. This ensures a relative small database schema, but individual studies can be very large depending on the size of attached data files.

Searching for data

The online search facility provides the ability to search using free text through most of the underlying data fields, including the study description, study title, protocols, metabolites and authors. Currently, we support free-text searching and you can combine multiple search terms, for example 'human urine' will give you all studies where you find the terms 'human' and 'urine' are used. The search result page, as illustrated in Figure 1, shows general study information like the submitter of the study, the study title, public release date, organism(s), study design and platform.

It is possible to further refine the search result using 'search facets'. Search facets give the user the ability to limit the search results to a selection of species, platform and metabolite. For example, if you select a specific organism from the filter, the search results are limited to show only studies containing this organism. The search mechanism in MetaboLights is implemented using a text index (Lucene index) so no direct backend database queries are performed during a general search. This ensures a fast search facility.

Figure 1 shows the search results page when searching for 'human' across all of MetaboLights. To see the details of a study, the user can simply click on the study title. Example of what is displayed in the study details are in Supplementary Figures S1–S4. These images show screenshots of the web interface of MetaboLights with study data loaded for an NMR-based metabolomics study, MTBLS1. The Study details page consists of four tabs. The first tab (Supplementary Figure S1) shows information about the submitters, the relevant dates, study title and

The screenshot displays the MetaboLights web application interface. At the top, the 'Metabolights' logo is on the left, followed by a search bar containing the text 'human' and a search button. To the right of the search bar are navigation links: 'home', 'browse', 'download', 'submit', 'help', and 'log in'. Below the search bar, a light blue banner indicates '3 Search result(s) showing 1 to 3'. On the left side, there are two filter boxes. The first, titled 'Organism', has a checked checkbox for 'Homo sapiens (Human)'. The second, titled 'Technology', has unchecked checkboxes for 'mass spectrometry' and 'NMR spectroscopy'. The main content area shows three search results, each in a light blue box. The first result is titled 'NMR based metabolomics of Human Type 2 Diabetes urine samples', with a public release date of 15-Feb-2012. It lists the organism as Homo sapiens (Human), study factors including metabolic syndrome and gender, and assays including NMR spectroscopy (132). The study identifier is MTBLS1, submitted by Reza Salek. The second result is titled 'Annotation of the human adult urinary metabolome and metabolite identification using ultra high performance liquid chromatography coupled to a LTQ-Orbitrap mass spectrometer', with a public release date of 06-Aug-2012. It lists the organism as Homo sapiens (Human) and assays including mass spectrometry (148). The study identifier is MTBLS20, submitted by Christophe Junot. The third result is titled 'High-resolution extracted ion chromatography, a new tool for metabolomics and lipidomics using a second-generation orbitrap mass spectrometer', with a public release date of 15-Feb-2012. It lists the organism as Homo sapiens (Human), study factors including healthy volunteers and blood plasma, and assays including mass spectrometry (60). The study identifier is MTBLS4, submitted by Albert Koulman.

Figure 1. Searching for 'human' in the MetaboLights web application.

description, organisms, study design, publications and the experimental factors. The next tab (Supplementary Figure S2) details the protocols used during this study, from how the sample was collected through to the metabolite identification. Next, we have the data tab (Supplementary Figure S3). Here, we show data files for this study, detailed for technology platform used and experimental factors. Finally, we have metabolite identification (Supplementary Figure S4). Each identified metabolite has an external database reference, for example a ChEBI or HMDB identifier. Metabolites identified with a ChEBI accession show additional molecule description. The identified metabolite tab details which sample the compound was identified in. Unknown compounds are listed without a database reference.

Browsing data

Users can browse studies in MetaboLights using the 'browse' link. This will give a complete list of all the public studies currently available. If the user is registered and currently logged in to MetaboLights, additional private studies may be displayed. These private studies are either under the users control or have been directly shared from other users. To limit the number of studies in the browsing list, the user can activate the same facets available for a general search.

Downloads and programmatic access

MetaboLights software components are open source and all data are free to download and use for any purpose. All public studies are downloadable as ISA-Tab (22) metadata files with associated data files directly from the online study details page, and from the MetaboLights download page <http://www.ebi.ac.uk/metabolights/download>. A direct bulk download using ftp is available from <ftp://ftp.ebi.ac.uk/pub/databases/metabolights/>, organized into sub-folders for public studies. There are no web services for programmatic access available at present. However, this functionality is scheduled for a future release of the repository.

Submitting data

MetaboLights accepts experimental descriptions in ISA-Tab format, which can be created by the ISAcreeator editor tool. MetaboLights also offers different templates for the ISAcreeator tool to accommodate the description of different types of metabolomics experiments. ISAcreeator is a standalone Java desktop application that enables researchers to report experimental information, associate raw and processed data files, and submit the collated information to the MetaboLights database. Building on the OSGI plugin architecture, the ISAcreeator has been

extended to create a 'Metabolite Identification' add-on to capture relevant information for all small molecules identified in a study, with a link to a relevant chemical database (Figure 2). MetaboLights also accepts studies that have unknown or incomplete metabolite identification. This information has the potential to facilitate the identification of unknown metabolites in the future.

Currently, we accept all data formats for 'raw' instrumental data, converted open source file formats and any processed data, but we strongly recommend that processed data should be made available in open formats, such as mzML (23) for MS data.

MetaboLights implements metadata guidelines according to the recommendations of the Metabolomics Standards Initiative (MSI). The MSI defined a set of metabolomics reporting standards by harnessing and coordinating the efforts of several pre-existing international initiatives. MSI developed checklists and standards that have subsequently been adopted by the community, including minimum metadata reporting recommendations (24).

To facilitate high quality data submissions for NMR or MS experiments, there is a guided submission process to help meet MSI recommendations and extensively use community-developed controlled vocabularies and ontologies. ISAcreator also provides advanced mechanisms for mapping to and uploading information from existing spreadsheets. Figure 3 illustrates the ISA components in a typical data creation scenario.

An R package has been developed to facilitate data analysis (Supplementary Method). The Risa module, available in the next BioConductor release, includes functionality to process mass spectrometry data relying on the xcms package (25), and to save analysis results back to ISA archives.

Installing a local copy of the MetaboLights repository

To install MetaboLights locally, you require a SQL database (MySQL, PostgreSQL or Oracle), a subversion client (svn) and an Apache Tomcat server. The MetaboLights Repository source code can be found at <http://sourceforge.net/projects/metabolomes>, here you will also find more details regarding how to install MetaboLights locally. The ISAcreator Metabolite Identification plugin can be found at: <https://github.com/EBI-Metabolights/ISAcreatorPlugins>. The ISA framework is also open source and is available at: <https://github.com/ISA-tools>. Figure 4 shows the principal components of a local MetaboLights repository installation.

Access and privacy policy

MetaboLights grants free access and reuse of the public data it stores to everyone. Only registered users can upload and share study data. To facilitate deposition of research data not yet publicly visible, the submitter can set a data embargo for a period of up to 60 months, which can be lifted on results publication or extended upon request. Submitters can also request for access to their private data to be granted to specific other registered users. This feature may be particularly useful in facilitating collaborations and the peer review process.

Feedback

To facilitate user feedback, we have created a SourceForge tracker for logging issues, available at <http://sourceforge.net/projects/metabolomes>. There is also an online contact form, <http://www.ebi.ac.uk/metabolights/contact>, and contact email address, metabolights-help@ebi.ac.uk.

Row No.	identifier	chemical_formula	description	chemical_shift	multiplicity	taxid
28	CHEBI:60645	C5H10O3	a-Hydroxyisovalerate	[0.96 .. 1.01]		NEWT:9606
29	CHEBI:24898	C6H13NO2	isoleucine	[1.01 .. 1.03]	d 1.01	NEWT:9606
30	CHEBI:35932	C6H11O3	a-keto-b-methyl-N v...	[1.03 .. 1.09]	d 1.075	NEWT:9606
31	CHEBI:16530	C5H8O3	2-oxoisovalerate	[1.09 .. 1.13]	d 1.11	NEWT:9606
32	CHEBI:16236	C2H6O	Ethanol	[1.09 .. 1.13]	t 1.11	NEWT:9606
33				[1.13 .. 1.16]		NEWT:9606
34				[1.16 .. 1.18]		NEWT:9606
35	CHEBI:20067	C4H8O3	b-Hydroxybutyrate	[1.18 .. 1.22]	d 1.20	NEWT:9606
36			Unknown_t 1.205	[1.18 .. 1.22]	t 1.205	NEWT:9606
37			Unknown_d 1.235	[1.22 .. 1.26]	d 1.235	NEWT:9606
38	CHEBI:30860	C4H6O4	methylmalonate	[1.22 .. 1.26]	d 1.255	NEWT:9606
39			Unknown_s 1.27	[1.26 .. 1.32]	s 1.27	NEWT:9606
40	CHEBI:28358	C3H6O3	lactic acid	[1.32 .. 1.35]	d 1.335	NEWT:9606
41			Unknown_s 1.35	[1.35 .. 1.38]	s 1.35	NEWT:9606
42	CHEBI:50129	C4H8O3	a-hydroxyisobutyrate	[1.35 .. 1.38]	s 1.36	NEWT:9606
43				[1.38 .. 1.43]		NEWT:9606
44				[1.43 .. 1.47]		NEWT:9606
45	CHEBI:16449	C3H7NO2	alanine	[1.47 .. 1.50]	d 1.485	NEWT:9606
46				[1.50 .. 1.55]		NEWT:9606
47	HMDR00701	C8H15NO2	caprovaline	[1.55 .. 1.60]		NEWT:9606

Figure 2. Part of Study MTBLS1 in ISAcreator with the Metabolite Identification Plugin active.

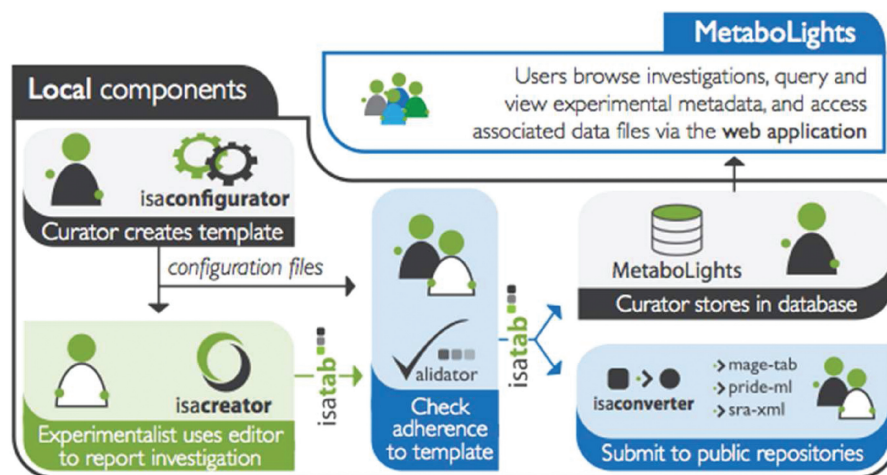


Figure 3. A typical workflow, using the ISA framework, for reporting information and submitting it to the MetaboLights database.

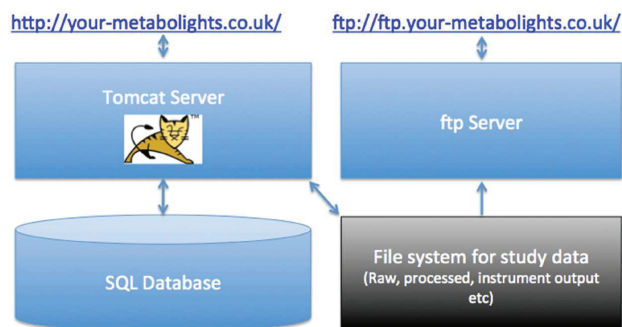


Figure 4. Simple technical architecture of a local MetaboLights repository, including web (http) and ftp access.

DISCUSSION

The MetaboLights repository was launched on the 28 June 2012 at the 8th International Conference of the Metabolomics Society in Washington, DC, USA. The repository is now accepting study submissions from a growing number of active users worldwide with submission privileges. For the latest statistics on current studies and submitters, please see <http://www.ebi.ac.uk/metabolights/stats>.

The requirement by a growing number of publishers and funding agencies to deposit data associated with journal publications to public repositories is expected to motivate a substantial number of future submissions. As more datasets become available, Metabolights will become an invaluable resource for those wishing to develop new algorithms for the processing of metabolomic data. The creation of a long-term institution-backed, as it will be maintained by EBI after the grant ends, public repository such as MetaboLights at EMBL-EBI allows laboratories across the globe to collaborate on projects through data sharing, and thereby to begin to collaboratively generate the large datasets needed to address how the environment, genome and diet influence the metabolome of a species.

Future work

The MetaboLights team is now actively specifying the *MetaboLights Reference Layer*, which will be launched in Summer 2013. The Reference Layer will be a comprehensive knowledge base organized around a metabolite-centric view, and will include elements such as reference spectra of various types, biological reference data, protocols, cross-references to other resources and advanced search and download functionality. There will be comprehensive manually curated data, including chemical structures and characteristics from ChEBI, metabolic pathways, reference spectroscopy and chromatography. Furthermore, there will be information about the reference biology, metabolites and their occurrence and concentration in species, organs, tissues and cellular compartments in various conditions, both healthy and diseased. Publication references and protocols will also be available. This will enable experimentalists to get a comprehensive Metabolomic view on known metabolites.

We are also substantially enhancing our online help capabilities with online video instructions as well as detailed scenarios for completing new studies for submission. A new section with 'Gold Standard Studies' will be included for easy reference. These studies can be used as templates for similar experiments.

In October 2012, the European COOrdination of Standards in MetabOmicS (COSMOS) consortium, comprising 14 European partners, will start its work on Metabolomics data standardization, publication and dissemination workflows. The MetaboLights database is a key component in this effort. It is the aim of the COSMOS project to develop efficient policies to ensure that Metabolomics data are encoded in open standards, tagged with a community-agreed and complete set of metadata, supported by a communally developed set of open source data management and capturing tools, disseminated in open-access databases adhering to these standards, supported by vendors and publishers, who require deposition upon publication, and properly interfaced with data in other biomedical and life science

e-infrastructures [such as ELIXIR (26), BioMedBridges (<http://www.biomedbridges.eu>), EU-OPENSOURCE (<http://www.eu-openscreen.de>) and BBMRI (<http://bbmri.eu>)].

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–4 and Supplementary Methods.

ACKNOWLEDGEMENTS

The MetaboLights project team thanks the following persons for their invaluable contributions: Rafael Alcántara, Masanori Arita, Mike Beale, Nick Bond, Kees van Bochove, Jildau Bouwman, Scarlet Brockmoeller, Steve Bryant, Hong Cao, Juan Castrillo, Jenny Cham, Cecilia Castro, Yajing Chu, Tim Ebbels, Michael Eiden, Oliver Fiehn, Andrew Gibbs, Roy Goodacre, Martin Hornshaw, Jan Hummel, Albert Koulman, Peter Meadows, Pablo Moreno, Theo Reijmers, Francis Rowland, Linda Scoriels, Mark Seymour, Tim Smith, Anthony Taylor, Chris Taylor, Michael Wakelam, Jane Ward and David Wishart.

FUNDING

The development of MetaboLights is funded by the BBSRC [BB/I000933/1]; ISA framework by the BBSRC [BB/I025840/1, BB/I000771/1, BB/J020265/1 to S.A.S.]; NERC, EU [EC 312941 to S.A.S.]; University of Oxford e-Research Centre [to S.A.S.]. COSMOS is funded by European Commission Grant [EC312941]. Funding for open access charge: BBSRC [BB/I000933/1].

Conflict of interest statement. None declared.

REFERENCES

- Fiehn, O. (2002) Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.*, **48**, 155–171.
- German, J.B., Hammock, B.D. and Watkins, S.M. (2005) Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics*, **1**, 3–9.
- Pearson, H. (2007) Meet the human metabolome. *Nature*, **446**, 8.
- Cheng, K.K., Benson, G.M., Grimsditch, D.C., Reid, D.G., Connor, S.C. and Griffin, J.L. (2010) Metabolomic study of the LDL receptor null mouse fed a high-fat diet reveals profound perturbations in choline metabolism that are shared with ApoE null mice. *Physiol. Genomics*, **41**, 224–231.
- Veldhoen, N., Ikononou, M.G. and Helbing, C.C. (2012) Molecular profiling of marine fauna: integration of omics with environmental assessment of the world's oceans. *Ecotoxicol. Environ. Saf.*, **76**, 23–38.
- Kell, D.B. (2004) Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.*, **7**, 296–307.
- Xu, E.Y., Schaefer, W.H. and Xu, Q.W. (2009) Metabolomics in pharmaceutical research and development: metabolites, mechanisms and pathways. *Curr. Opin. Drug Discovery Dev.*, **12**, 40–52.
- Gibney, M.J., Walsh, M., Brennan, L., Roche, H.M., German, B. and van Ommen, B. (2005) Metabolomics in human nutrition: opportunities and challenges. *Am. J. Clin. Nutrition*, **82**, 497–503.
- Kaddurah-Daouk, R., Kristal, B.S. and Weinshilboum, R.M. (2008) Metabolomics: a global biochemical approach to drug response and disease. *Annu. Rev. Pharmacol. Toxicol.*, **48**, 653–683.
- Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G. and Kell, D.B. (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.*, **22**, 245–252.
- Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S. *et al.* (2007) HMDB: the human metabolome database. *Nucleic Acids Res.*, **35**, D521–D526.
- Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.
- Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan, R. and Siuzdak, G. (2005) METLIN: a metabolite mass spectral database. *Ther. Drug Monit.*, **27**, 747–751.
- Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E.A., Glass, C.K., Merrill, A.H., Murphy, R.C., Raetz, C.R.H., Russell, D.W. *et al.* (2007) Lmsd: lipid maps structure database. *Nucleic Acids Res.*, **35**, D527–D532.
- Sansone, S.A., Fan, T., Goodacre, R., Griffin, J.L., Hardy, N.W., Kaddurah-Daouk, R., Kristal, B.S., Lindon, J., Mendes, P., Morrison, N. *et al.* (2007) The metabolomics standards initiative. *Nat. Biotechnol.*, **25**, 846–848.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M. *et al.* (2005) Gmd@Csb.Db: the Golm metabolome database. *Bioinformatics*, **21**, 1635–1638.
- Carroll, A.J., Badger, M.R. and Harvey Millar, A. (2010) The MetabolomeExpress Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. *BMC Bioinformatics*, **11**, 376.
- Akiyama, K., Chikayama, E., Yuasa, H., Shimada, Y., Tohge, T., Shinozaki, K., Hirai, M.Y., Sakurai, T., Kikuchi, J. and Saito, K. (2008) PRIME: a web site that assembles tools for metabolomics and transcriptomics. *In Silico Biol.*, **8**, 339–345.
- de Matos, P., Alcantara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S. and Steinbeck, C. (2010) Chemical entities of biological interest: an update. *Nucleic Acids Res.*, **38**, D249–D254.
- Vastrik, I., D'Eustachio, P., Schmidt, E., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L. *et al.* (2007) Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.*, **8**, R39.
- Sansone, S.A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, P., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L. *et al.* (2012) Toward interoperable bioscience data. *Nat. Genet.*, **44**, 121–126.
- Rocca-Serra, P., Brandizi, M., Maguire, E., Sklyar, N., Taylor, C., Begley, K., Field, D., Harris, S., Hide, W., Hofmann, O. *et al.* (2010) ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics*, **26**, 2354–2356.
- Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W.H., Rompp, A., Neumann, S., Pizarro, A.D. *et al.* (2011) mzML—a community standard for mass spectrometry data. *Mol. Cell. Proteomics*, **10**, R110.000133.
- Goodacre, R., Broadhurst, D., Smilde, A.K., Kristal, B.S., Baker, J.D., Beger, R., Bessant, C., Connor, S., Calmani, G., Craig, A. *et al.* (2007) Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, **3**, 231–241.
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.*, **78**, 779–787.
- Crosswell, L.C. and Thornton, J.M. (2012) ELIXIR: a distributed infrastructure for European biological data. *Trends Biotechnol.*, **30**, 241–242.