

3 The metabolomics standards initiative (MSI)

4 Oliver Fiehn · Don Robertson · Jules Griffin ·
5 Mariet van der Werf · Basil Nikolau · Norman Morrison ·
6 Lloyd W. Sumner · Roy Goodacre · Nigel W. Hardy ·
7 Chris Taylor · Jennifer Fostel · Bruce Kristal ·
8 Rima Kaddurah-Daouk · Pedro Mendes ·
9 Ben van Ommen · John C. Lindon · Susanna-Assunta Sansone

10 Received: 12 January 2007 / Accepted: 8 June 2007
11 © Springer Science+Business Media, LLC 2007

12 **Abstract** In 2005, the Metabolomics Standards Initiative
13 has been formed. An outline and general introduction is
14 provided to inform about the history, structure, working
15 plan and intentions of this initiative. Comments on any of
16 the suggested minimal reporting standards are welcome to
17 be sent to the open email list Msi-workgroups-feed-
18 back@lists.sourceforge.net
19

Keywords Standardization · Metabolomics · 20
Metabonomics · Metabolite profiling · Databases · 21
Ontology 22
Metabolomics and metabonomics techniques have suffi- 23
ciently matured over the past decade to enable utilization of 24
these methods for multiple and diverse projects and studies 25

A1 O. Fiehn (✉)
A2 Genome Center, University of California, Davis, Davis, CA
A3 95616, USA
A4 e-mail: ofiehn@ucdavis.edu
A5 D. Robertson
A6 Molecular Profiling, Pfizer Global Research and Development,
A7 Ann Arbor, MI 48105, USA
A8 J. Griffin
A9 Department of Biochemistry, University of Cambridge, Tennis
A10 Court Road, Cambridge CB2 1QW, UK
A11 M. van der Werf · B. van Ommen
A12 TNO Quality of Life, Zeist, The Netherlands
A13 B. Nikolau
A14 Iowa State University, Ames, IA, USA
A15 N. Morrison
A16 School of Computer Science, University of Manchester,
A17 Kilburn Building, Oxford Road, Manchester M13 9PL, UK
A18 N. Morrison
A19 NERC Environmental Bioinformatics Centre,
A20 Oxford Centre for Ecology and Hydrology, Oxford
A21 OX1 3SR, UK
A22 L. W. Sumner
A23 The Samuel Roberts Noble Foundation, Ardmore, OK, USA

A24 R. Goodacre
A25 School of Chemistry, The University of Manchester,
A26 Manchester, UK
A27 N. W. Hardy
A28 Department of Computer Science, University of Wales, Penglais,
A29 Aberystwyth, Ceredigion, Wales SY23 3DB, UK
A30 C. Taylor · S.-A. Sansone
A31 EMBL-EBI The European Bioinformatics Institute, Wellcome
A32 Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
A33 J. Fostel
A34 National Institute for Environmental Health Science, Research
A35 Triangle Park, NC 27709-2233, USA
A36 B. Kristal
A37 Department of Neurosurgery, Harvard Medical School, Brigham
A38 and Women's Hospital, Boston, MA, USA
A39 R. Kaddurah-Daouk
A40 Department of Psychiatry and Behavioral Sciences, Duke
A41 University, Durham, NC, USA
A42 P. Mendes
A43 Virginia Bioinformatics Institute, Blacksburg, VA, USA
A44 J. C. Lindon
A45 Department of Biomolecular Medicine, Imperial College of
A46 London, London SW7 2AZ, UK

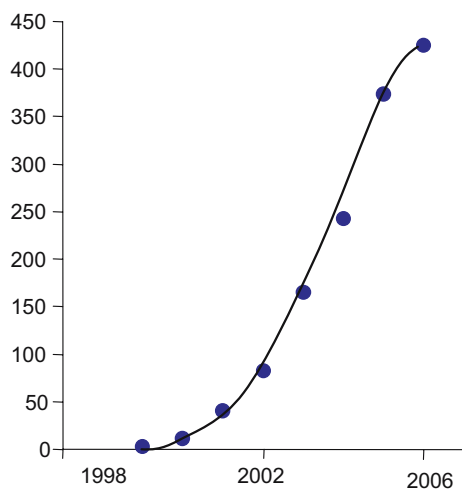


Fig. 1 Number of publications per year based on ISI Web-of-Science database query using the search phrases *metabolom** OR *metabonom**

in biology and biomedicine. The number of publications in peer-reviewed journals has risen steadily every year to well over 400 reports in 2006, 75% of which were original research papers or published meeting abstracts. Although the number of publications (Fig. 1) is still small compared to genomics and proteomics (which see about ten times the number of reports), the need for reporting standards has been recognized early on.

Metabolomic data can be an important source of new knowledge, as long as it can be shared in a way to permit use and re-use. Because the biological milieu can change rapidly, and can be highly responsive to the environment, history and handling of the subject, details of the experiment are critical to interpretation of the data and to its re-use in meta-analyses of data from multiple experiments. There is, therefore, a need for data reporting standards, such as checklists, outlining the minimal information content that should be reported, common syntax, defining the transmission formats that facilitate the exchange of information, and common semantics, adding the interpretive layer to the data (Field and Sansone 2006). Data reporting standards support an unambiguous description of the study and its execution, the subjects, and the biological materials and data derived from them. Standards make information more accessible and enable the extraction of maximum value from data sets, by supporting efficient querying, accurate data analysis, and facilitate data exchange (Quackenbush 2004). At least, reporting standards should be usable for minimum requirements in the peer-reviewing process of journal publications.

Led by the Imperial College in London (UK), a consortium for metabonomic toxicity had started working in 2001 on generating databases of ^1H NMR spectra of body

fluids from animal models, involving six major pharmaceutical corporations (Lindon et al. 2003). Due to the collaborative nature of this consortium, standards were needed how to exchange data and communicate results, leading to the SMRS initiative (Standard Metabolic Reporting Structure). The SMRS group published their detailed discussion in their forum (SMRS XXXX) and in a peer-reviewed paper in 2005 (Lindon et al. 2005). Simultaneously, a plant focused consortium worked on evolving a generic data model to provide a basis for the design of systems for data storage and exchange, called Architecture for Metabolomics (ArMet) (Jenkins et al. 2004). Both efforts were complementary, because ArMet focused on the conceptual framework how to organize metabolomic data and supporting information (called 'metadata') whereas the SRMS group initiated detailed recommendations of which parameters and data were necessary to be reported. Following a series of MetaboMeetings (XXXX) initiated at The European Bioinformatics Institute, participants of both initiatives joined with further interested scientists to form the Metabolomics Standards Initiative (MSI). The MSI took up the work end of 2005 after an inaugural workshop (Castle et al. 2006) hosted by the U.S. National Institutes of Health (NIH) and the Metabolomics Society.

Standardization efforts at the MSI are being carried out by an international community of volunteers who are working to generate broad community consensus around the proposed standards. The reports presented in this issue of *Metabolomics* are designed as a means to spotlight ongoing work and an important mile stones rather than a mechanism to report final descriptions of the various standards under development. It is hoped that these reports will spark additional discussion, community input, and raise awareness of this effort. Progress on these objectives and dissemination of updated versions of MSI documents are publicly available at the initiative's public webpage [<http://msi-workgroups.sourceforge.net/>]. The aim of the reporting standards proposed here is not to prescribe how an experiment is carried out but rather to provide a common mechanism for describing the work so that the data can be made available to others for evaluation, or to support an extension or repeat of the work as desired, or published in a public repository. In metabolomics, more so than in other 'omics areas', the collection of data about the subject during and after sample preparation is critical to support interpretation of the results and also to facilitate use of the data in other analyses that may uncover new relationships between biological state and the metabolome.

The effort is coordinated by the Metabolomics Society who appointed an MSI Oversight Committee that helps to coordinate the work and to ensure that approaches and data

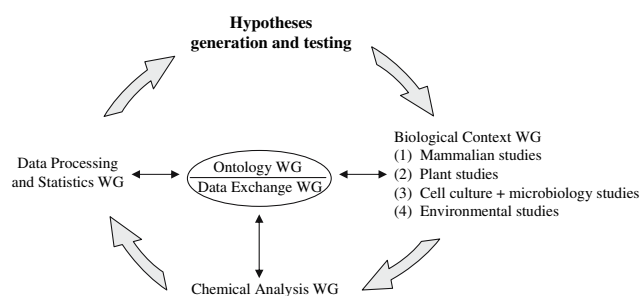


Fig. 2 Organization of the MSI working groups (WG) along the general scientific cycle of generating and testing hypotheses

relationships shared by the various working groups are maintained. The MSI is further structured into five working groups (Fig. 2):

1. Biological context metadata with the following subgroups:
 - (1) mammalian studies, divided between human and animal studies
 - (2) plant studies
 - (3) cell cultures and microbiology
 - (4) environmental studies
2. Chemical analysis
3. Data processing
4. Ontology
5. Data exchange

The structure of the working groups follows the scientific workflow (Lockhart and Winzeler 2000), and starts with the experimental design according to which the biological samples are generated ('Biological Context') and continues with data acquisition ('Chemical Analysis'), and data evaluation ('Data Processing and Statistics') in order to falsify or validate the starting hypotheses, to refine existing hypotheses or even to generate completely novel hypotheses as result of a metabolomic study. In addition, two further working groups are concerned with facilitating re-use of data and general consistency, the Data Exchange working group and the Ontology working group. Both groups span the breadth of parameters that is suggested by the other working groups and hence, are leaning on the results and input from these groups.

For all groups there is certain degree of overlap with other existing initiatives in functional genomics, namely the Human Proteome Organization Proteomics Standards Initiative (Taylor et al. 2006) and the Microarray Gene Expression Data Society (Ball and Brazma 2007). Interaction with these and related initiatives and communities is ensured by both participating in workshops and collaborating at the umbrella level under several synergistic projects. These include: (a) MCheck (XXXX), a response to

the current proliferation of Minimal Information checklists; (b) Functional Genomics Object Model (Jones et al. 2006), to underpin the development of tools, XML formats and database schemata; (c) the Ontology for Biomedical Investigations (Whetzel et al. 2006; OBI XXXX), to build a cross-domain ontology as a resource for the annotation of investigations. The aim of such collaborations is to facilitate data integration across biological or cellular domains and instrument equipments, but also to avoid duplication of efforts and to define common standards (Quackenbush 2006).

The MSI acknowledges and appreciates all input into this effort so far, and actively seeks for greater participation (Fiehn et al. 2006) for improving the current stage of its Standards documents. Apart from ensuring a large community consensus by reaching out to a greater number of scientists in academia and industry, these MSI reporting standards are thought to be eventually endorsed and supported by funding agencies, database repositories and scientific journals. Ultimately, it is envisioned that regulatory bodies may adopt these recommendations as a way to standardize reports of experimental data submitted by their diverse constituencies with the goal of simplifying the task of reviewers thus accelerating the process. Indeed, the US FDA has already utilized the early efforts of MSI in the preparation of their draft (and yet to be published) metabolomics best practice document.

Acknowledgments This research was supported in part by the Intramural Research Program of the NIH, and NIEHS, and partially supported by a research grant MCB-0520140 from the U.S. National Science Foundation.

References

- Ball, C. A., & Brazma, A. (2006). MGED standards: Work in progress. *OMICS*, 10, 138–144.
- Castle, L. A., Fiehn, O., Kaddurah-Daouk, R., & Lindon, J. C. (2006). Metabolomics standards workshop and the development of international standards for reporting metabolomics experimental results. *Briefings in Bioinformatics*, 7, 159–165.
- Fiehn, O., Kristal, B., van Ommen, B., Sumner, L. W., Sansone, S. A., Taylor, C., Hardy, N., & Kaddurah-Daouk, R. (2006). Establishing reporting standards for metabolomic and metabonomic studies: A call for participation. *OMICS*, 10, 158–63.
- Field, D., & Sansone, S. A. (2006). A special issue on data standards. *OMICS*, 10, 84–93.
- Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A. R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R., Hall, R., Kopka, J., Lane, G. A., Lange, B. M., Liu, J. R., Mendes, P., Nikolau, B. G., Oliver, S. G., Paton, N. W., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgard, J., Sumner, L. W., Wang, T., Walsh, S., Wurtele, E. S., & Kell, D. B. (2004). A proposed framework for the description of plant metabolomics experiments and their results. *Nature Biotechnology*, 22, 1601–1605.

- 202 Jones, A. R., Miller, M., Aebersold, R., Apweiler, R., Ball, C. A.,
203 Brazma, C., Gref, G. D., Hardy, N., Hermjakod, H., Hubbard, S.
204 J., Hussey, P., Igra, M., Jenkins, H., Julian Jr., R. K., Laursen, K.,
205 Oliver, S. G., Paton, N. W., Sansone, S. A., Sarkans, U.,
206 Stoeckert Jr., C. J., Taylor, C. F., Whetzel, P. L., White, J. A.,
207 Spellman, P., & Pizarro, A. (2006). The Functional Genomics
208 Experiment model (FuGE): An extensible framework for
209 standards in functional genomics. *Nature Biotechnology* (*com-*
210 *munity consultation phase*: <http://www.nature.com/nbt/consult>).
211 Lindon, J. C., Nicholson, J. K., Holmes, E., Antti, H., Bollard, M. E.,
212 Keun, H., Beckonert, O., Ebbels, T. M., Reilly, M. D.,
213 Robertson, D., Stevens, G. J., Luke, P., Breau, A. P., Cantor,
214 G. H., Bible, R. H., Niederhauser, U., Senn, H., Schlotterbeck,
215 G., Sidemann, U. G., Laursen, S. M., Tymiak, A., Car, B. D.,
216 Lehman-McKeeman, L., Colet, J. M., Loukaci, A., & Thomas,
217 C. (2003). Contemporary issues in toxicology—The role of
218 metabonomics in toxicology and its evaluation by the COMET
219 project. *Toxicology and Applied Pharmacology*, *187*, 137–146.
220 Lindon, J. C., Nicholson, J. K., Holmes, E., Keun, H. C., Craig, A.,
221 Pearce, J. T. M., Bruce, S. J., Hardy, N., Sansone, S. A., Antti,
222 H., Jonsson, P., Daykin, C., Navarange, M., Beger, R. D.,
223 Verheij, E. R., Amberg, A., Baunsgaard, D., Cantor, G. H.,
224 Lehman-McKeeman, L., Earll, M., Wold, S., Johansson, E.,
225 Haselden, J. N., Kramer, K., Thomas, C., Lindberg, J., Schuppe-
226 Koistinen, I., Wilson, I. D., Reily, M. D., Robertson, D. G., Senn,
227 H., Krotzky, A., Kochhar, S., Powell, J., van der Ouderaa, F.,
Plumb, R., Schaefer, H., & Spraul, M. (2005). Summary
228 recommendations for standardization and reporting of metabolic
229 analyses. *Nature Biotechnology*, *23*, 833–838.
230 Lockhart, D. J., & Winzler, E. A. (2000). Genomics, gene expression
231 and DNA arrays. *Nature*, *405*, 827–836.
232 MetaboMeetings. [<http://www.thempf.org/conferences.html>].
233 MICheck. [<http://micheck.sourceforge.net/>].
234 OBI. [<http://obi.sourceforge.net/>].
235 Quackenbush, J. (2004). Data standards for ‘omic’ science. *Nature*
236 *Biotechnology*, *22*, 613–614.
237 Quackenbush, J. (2006). Standardizing the standards. *Molecular*
238 *Systems Biology*, *2*, 2006.
239 SMRS. [<http://www.smrsgroup.org/>].
240 Taylor, C. F., Hermjakob, H., Julian Jr., R. K., Garavelli, J. S.,
241 Aebersold, R., & Apweiler, R. (2006). The work of the Human
242 Proteome Organisation’s Proteomics Standards Initiative (HUPO
243 PSI). *OMICS*, *10*, 145–151.
244 Whetzel, P. L., Brinkman, R. R., Causton, H. C., Fan, L., Field, D.,
245 Fostel, J., Fragoso, G., Gray, T., Heiskanen, M., Hernandez-
246 Bousard, T., Morrison, N., Parkinson, H., Rocca-Serra, P.,
247 Sansone, S. A., Schober, D., Smith, B., Stevens, R., Stoeckert Jr.,
248 C. J., Taylor, C., White, J., & Wood, A. (2006). Development of
249 FuGO: An ontology for functional genomics investigations.
250 *OMICS*, *10*, 199–204.
251
252