

A roadmap for the establishment of standard data exchange structures for metabolomics

Nigel W. Hardy · Chris F. Taylor

Received: 18 January 2007 / Accepted: 25 June 2007 / Published online: 6 September 2007
© Springer Science+Business Media, LLC 2007

Abstract This briefing lays out the initial objectives and intent of the Data Exchange Working Group of the Metabolomics Standards Initiative (MSI), which operates under the aegis of the Metabolomics Society. The MSI has now reached a significant milestone in publishing a complete set of draft reporting requirements, simultaneously providing to the Data Exchange Working Group a set of drivers that function, for our purposes, as a series of specifications of the data and metadata that the products endorsed or developed by the Data Exchange Working Group should be able to effectively contain. We here describe our view of the existing resources that can be put to use, the ‘functionality gaps’ that need to be filled and the way in which we envisage these existing and planned resources fitting together to provide a framework that firstly, answers the needs of the metabolomics community and secondly, addresses the issues raised by the prospect of integrating metabolomics workflow descriptions and data with that from other domains, such as proteomics, genomics and transcriptomics.

Keywords Data exchange · Data format · Metabolomics Standards Initiative

1 Introduction

The current surge in the development and application of metabolomics techniques has greatly increased the rate of production of metabolomics data. Taken as a whole, those data and the metadata needed to describe and contextualize them have also become significantly more diverse. There is a growing recognition in equivalent fields such as transcriptomics and proteomics (Brazma et al. 2001; Taylor et al. 2003) and across the scientific community in general (National Research Council (US) 2003) that the dissemination and archiving of experimental data can accelerate the rate of scientific advance; this climate has fostered a number of initiatives to develop standards for metabolomics data (Bino et al. 2004; Jenkins et al. 2004; Lindon 2005) in addition to a plethora of technology specific and general data formats. In 2005 the Metabolomics Society established the Metabolomics Standards Initiative¹ comprising a series of specialist working groups with an oversight committee monitoring, coordinating and reviewing their efforts. Working groups on Biological Context, Chemical Analysis and Data Processing were tasked to produce reporting standards making explicit the information to provide in an experimental report; their drafts are reported elsewhere in this volume (Fiehn et al. 2007b; Goodacre et al. 2007; Griffin et al. 2007; Morrison et al. 2007; Sumner et al. 2007; van der Werf et al. 2007). The Ontology working group (Sansone et al. 2007) (also reporting in this volume) seeks to support the activities of the other MSI working groups by developing standardised lists of descriptive terms, providing a common semantic framework to enable metabolomics user communities to annotate experiments in a concise, consistent and unam-

N. W. Hardy (✉)
Department of Computer Science, University of Wales,
Aberystwyth, Penglais, Aberystwyth,
Ceredigion SY23 3DB, UK
e-mail: nwh@aber.ac.uk

C. F. Taylor
The European Bioinformatics Institute, Hinxton, UK

¹ <http://msi-workgroups.sourceforge.net/>

biguous manner. The Data Exchange Working Group's original stated aim was to "define data exchange formats and produce a schema for such operations that cover all aspects of the metadata, the analytical data (both spectroscopic and chromatographic) and the data analysis". The overall work plan for the group has been to review and evaluate existing initiatives which might be potential partial solutions prior to commencing design work based of the developing reporting standards This paper provides a progress report. The group welcomes input to its continuing work via the authors or the e-mail list Msi-workgroups-feedback@lists.sourceforge.net which receives comments on any aspect of the work of the MSI.

2 Context

Data may be generated and stored using a variety of technologies, some open and some proprietary. A data exchange standard provides a means by which data can be transmitted between producers and storage, with guarantees of common semantics and demonstrable validity. In particular it should support relevant reporting requirements (without limiting data sets to those requirements) and be able to demonstrate the degree of compliance to them shown by a particular data set.

As scientific investigations progress, data are frequently accumulated using computer systems. Various data formats and structures are required for this. Data will also subsequently be deposited in a range of repositories, including those of the originating laboratories, collaborators, funding organizations, scientific publishers and regulatory authorities. Again, formats and data structures will be required for this, though perhaps with different design objectives, including efficient data retrieval. In defining modular schemata and exchange formats the MSI intends to promote appropriate and semantically consistent transmission of data, but crucially, without constraining how it is managed at either end. Many alternate existing and novel LIMS and archival systems should be able to support the data, though the design of novel systems may be informed to a greater or lesser extent by the transmission standard and this may therefore be an additional benefit.

3 Proposed deliverables of the work group

The work group initially proposed four specific deliverables and this remains our intent. The first deliverable is a detailed specification of the data for exchange. This may be called a schema or a data model. A data model describes data precisely but independently of any particular implementation. Current software engineering practice involves

(automated) transformation of models into code; this is known as Model-Driven Architecture (MDA) (Miller and Mukerji 2003). A data model may be implemented in more than one way: by transformation into database tables and the associated code as a relational database implementation for storage; as a specification of an XML (Extensible Markup Language) document type for transmission; and as a library for a programming language (perhaps Java, C or C++). This first deliverable (the model) will therefore provide the basis for the format (see below) and will also be made available to software developers, who will work with it. The proposal is that the model should be represented in UML (the Unified Modeling Language). This is now the most widely used modeling language; it is open and supported by many tools; and it is more than adequate for modeling the emerging requirements we have to hand. Moreover, pre-existing components developed for other purposes will most likely be modeled in UML, permitting immediate reuse by the MSI.

The second deliverable is an implementation of the model as an exchange format intended for general application. XML² is currently the accepted basis for such formats. It is an open standard. Support for reading and generating XML as data streams or files is now almost universal in commonly used programming languages, data analysis packages and office automation software. While extensions appear regularly, the core language definition has been extremely stable (contributing to its rapid adoption) and can be expected to have a long life. Modularity is discussed below and XML (in contrast to its predecessor SGML, the Standard Generalized Markup Language) supports this well. A commonly cited argument against its adoption is the size of data files. XML does carry a size overhead due to the tags (the $\langle \rangle$ bracketed element delimiters familiar to users of HTML). Two responses to this are (i) that metabolomics data transmission is neither time critical nor bandwidth limited and (ii) that the World Wide Web Consortium is expected to respond with a binary representation³ which will reduce the size of documents. A second argument against its adoption is the difficulty of generating files. This is discussed below.

The standard will require documentation, not withstanding the fact that both UML and XML are self documenting, particularly in the syntactic sense. The way in which they reflect the reporting requirements and how users should use them will be the third deliverable. This will go hand in hand with the fourth which is demonstration data sets available on the web and a simple data entry system for creation of data sets that utilizes the formats developed.

² <http://www.w3.org/TR/xml11/>

³ <http://www.w3.org/XML/EXI/>

4 Existing standards and initiatives

Implicit in the constitution of the work group is the intention that re-use (or at least extension and adaptation) of existing standards from related fields is extremely desirable. Since establishment of the MSI the opportunities for such collaboration have increased. Early work in standardization in transcriptomics (Spellman et al. 2002) and proteomics (Taylor et al. 2003) as well as metabolomics (Jenkins et al. 2004; Lindon et al. 2005) and other fields, more or less explicitly recognized that future unification or integration of standards would remove the necessity to recast data from cross-disciplinary studies for reporting to different recipients and was therefore highly desirable in the longer term. At the same time, the need to clearly codify the needs of each discipline was recognized and separate development was the pragmatic approach. That work has begun to identify the commonalities and it is now appropriate that new developments should seek maximum commonality.

Existing standards, which may be purposed either for data transport or as preferred referents, range from the general to the specific. Broad-scope models such as ArMet, FuGE and MAGE, which are focused on data transport, contrast with standards for identifying chemical moieties such as InChI or ChEBI. Previous work on content, such as SMRS and MIAMET, may offer some contribution to gross structure. Table 1 is a working list of entities to consider. The group would welcome additional suggestions.

Cross-disciplinary standards will not be without costs. Specialization and extension of general standards for use in metabolomics or even for some part of metabolomics still involves substantial development costs and establishing and maintaining general agreement across a wider community can be costly. For the scientist generating cross-disciplinary data sets, while adherence to a narrower set of standards may imply less cost, the resulting complexity of the broader standards may mean that there is little or no saving.

FuGE (Jones et al. 2006) is described as “a framework for creating data standards for high-throughput biological experiments”. It is being used as a basis for data formats in transcriptomics (Ball and Brazma 2006) and the HUPO Proteomics Standards (Taylor et al. 2006). Its stated aims as a framework encompass metabolomics and initial investigation provides a *prima facie* case for its use in metabolomics. The workgroup is therefore committed to pursuing its application as the basis for the MSI standards.

One potentially important feature of FuGE is its ability to describe and robustly reference external data artefacts, such as files encoded in vendor-specific data formats. Of particular interest in this context is the Systems Biology Markup language (Hucka et al. 2003; Kell 2006). The

Table 1 Existing initiatives for consideration

Data exchange formats/models	
FuGE	(Jones et al. 2006)
MAGE	(Spellman et al. 2002)
ArMet	(Jenkins et al. 2004)
SpML	http://www.psidev.info/index.php?q=node/90
analysisXML	http://www.psidev.info/index.php?q=node/85
MzData	http://www.psidev.info/index.php?q=node/80
MzXML	http://sashimi.sourceforge.net/
MzML	http://www.psidev.info/index.php?q=node/80
CCPN	(Fogh et al. 2005)
LSID	http://www.omg.org/docs/formal/04-12-01.pdf
CDISC/SEND	http://www.cdisc.org/models/send/v2/
NMR-STAR	http://bmrb.protein.osaka-u.ac.jp/formats.html
PMML	(Wettschereck et al. 2003)
Reporting specifications	
SMRS	(Lindon et al. 2005)
MIAMET	(Bino et al. 2004)
ArMet	(Jenkins et al. 2004)
MIAPE	http://www.psidev.info/index.php?q=node/91
MIAME	(Brazma et al. 2001)
MIGS	http://darwin.nox.ac.uk/gsc/gcat/migs
CDISC/SEND	http://www.cdisc.org/models/send/v2/
NMR-STAR	http://bmrb.protein.osaka-u.ac.jp/formats.html
MIRIAM	(Novere et al. 2005)
MIBBI	(Taylor et al. <i>in press</i>)

ability to reference a particular SBML instance document, encoding a particular ‘model’ of metabolism, offers the opportunity to meaningfully link captured experimental data to an SBML-encoded metabolic model, which may have functioned as a ‘hypothesis’ to be tested. For example, were a particular SBML-encoded model, having been used in an *in silico* experiment, to make certain predictions that were then tested by ‘wet’ experimentation, that SBML-encoded model and both data sets (captured in FuGE) could be robustly linked (and therefore transmitted) together.

5 Emerging general principles

At this stage a number of general features of a data model are emerging which are now briefly reviewed.

The data model and format must be modular: arguments for this include the following. In general terms, modularity is typically regarded as good engineering. Pre-existing reporting standards and formats all show effective modularity, reflecting the inherent modularity in the underlying concepts. This analysis, reflected also in the structure of the MSI (Fiehn et al. 2007a) and draft reporting standards in-

cluded elsewhere in this volume, shows that multiple alternative reporting structures will be required. The most obvious alternative reporting structures concerns biological context metadata, where different application fields have differing requirements. The most obvious candidates for multiple reporting structures are (1) chemical analysis, where aliquots of a sample may be subjected to a number of techniques, and (2) data processing, where a data set may be analysed in multiple ways. There are a number of important benefits for the development and use of the standard accruing to modular design; it will permit parallel development of new resources and straightforward re-use of existing (supported) standards, which makes for more rapid implementation; it will allow for parts of the standard to appear over time without significant impact on the process as a whole; and integration with other initiatives will be more straightforwardly supported. Some module variants might, for example be entirely satisfied by existing standards. These could be represented in the same underlying format – assumed to be XML – or some other character based or binary format. Any module which may optionally be implemented using, for example, an existing non-XML format can be incorporated ‘in-line’ or by reference to external files from within XML using existing XML formalisms. Finally, data capture can be performed in a modular fashion. Previously assembled descriptions of standard procedures or standard biological lines may be re-used. Data may be added to a set as an experimental timeline proceeds.

In summary, the format is expected to consist of a suite of well decoupled modules. But in addition to modularity, the model must effectively represent the concatenated experimental workflow. This concept is explicit in the ArMet model (Jenkins et al. 2004) and is reflected in the SMRS recommendations (Lindon et al. 2005). Moreover a clear distinction between materials and processes will be important. This is reflected in a number of related ontologies; is strongly represented in FuGE and is evident in ArMet. One lesson learned from ArMet as published (and which has been mitigated in subsequent unpublished implementations) was a failure to make this distinction in all cases. Processes applied to materials have two aspects. There is a description of the process, which might represent a Standard Operating Procedure for example, and there is the description of its application, which should at least include information about the time and date, the operator and perhaps any deviations from the standard in this instance. Anecdotally, failure to distinguish procedure from application and intended from actual parameter values can often be a cause of confusion in the interpretation of metabolomics data. Later versions of ArMet increasingly reflect the distinction and support for it is a fundamental feature of the FuGE framework.

It is clear that validation of the completeness of a dataset must be flexible. It will be necessary for recipients of a dataset to be able to assess its acceptability according to local criteria with the assistance of model driven checking. The SMRS recommendations (Lindon 2005; Lindon et al. 2005) consider this in some detail. Submission to public databases; as journal supplementary data and to regulatory authorities are distinguished, along with the need in some circumstances to withhold some detail for reasons of confidentiality. It may be that in some circumstances editors and reviewers would consider that certain missing data would not render a dataset inadequate to support claims made. In this context dataset validation may be considered as a report on deficiencies (if any) rather than a blunt pass/fail test. In all cases, the model and derived formats must be extensible, and the inclusion of well characterized and identified additional data should not render a data set invalid.

Collection and assembly of complex data in XML form has proved difficult in other initiatives. Under ArMet, the benefits of form and table based support tools have been identified (Jenkins et al. 2005). In transcriptomics, the problems with the complexity of MAGE-ML has led to the development of MAGE-TAB (Rayner et al. 2006), a spreadsheet-based, MIAME-supportive format. The work group must consider data entry issues and their acceptability aspects.

6 Conclusions

The workgroup is committed to four deliverables: (i) a data model, in UML to accommodate the reporting requirements; (ii) an exchange format in XML which implements the model; (iii) user documentation on the use of the standard and (iv) example data sets using the standard.

The workgroup has developed a broad appreciation of existing standards and developed contacts with other communities. It is committed to maximal integration with standards in related fields and in particular expects to realize this through exploitation of the FuGE framework. With the first drafts of reporting standards now available development of a standard can begin. First, an overall modular structure must be established. This will, in part, involve a synthesis of several reporting standards. Secondly, detailed representational issues for data must be decided. The reporting standards are precise in terms of individual data items and their data types in some aspects, while in others, representations and possible values are not mandated. This work will involve close collaboration with the Ontology workgroup. Finally, we can render the fruits of our labors as test interfaces as a prelude to the ultimate release of the standard resources required by

the community for the exchange and archiving of their valuable data and its integration with that from other domains.

Acknowledgements Hardy acknowledges the support of the EU Framework VI project “META-PHOR”: Food-ST-2006-03622 and The University of Wales, Aberystwyth. Taylor acknowledges the support of the European Molecular Biology Laboratory.

References

- Ball, C. A., & Brazma, A. (2006) MGED standards: Work in progress. *OMICS: A Journal of Integrative Biology*, *10*, 138–144.
- Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B. J., Mendes, P., Roessner-Tunali, U., Beale, M. H., Trethewey, R. N., Lange, B. M., Wurtele, E. S., & Sumner, L. W. (2004). Potential of metabolomics as a functional genomics tool. *Trends Plant Science*, *9*, 418–425.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C. P., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., & Vingron, M. (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nature Genetics*, *29*, 365–371.
- Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., Ommen, B. v., Lindon, J. C., & Sansone, S.-A. (2007a) The Metabolomics Standards Initiative (MSI). *Metabolomics*, *3*, this issue.
- Fiehn, O., Sumner, L. W., Rhee, S. Y., Ward, J., Dickerson, J., Lange, B. M., Lane, G., Roessner, U., & Last, R. (2007b) Minimum reporting standards for plant biology context information in metabolomic studies. *Metabolomics*, *3*, this issue.
- Fogh, R. H., Boucher, W., Vranken, W. F., Pajon, A., Stevens, T. J., Bhat, T. N., Westbrook, J., Ionides, J. M. C., & Laue, E. D. (2005) A framework for scientific data modeling and automated software development. *Bioinformatics* *21*, 1678–1684.
- Goodacre, R., Baker, J. D., Beger, R., Broadhurst, D., Craig, G. C. A., Kell, D., Manetti, B. K. C., Newton, J., Paternostro, G., Sjöström, M., Smilde, A., Trygg, J., & Wulfert, F. (2007) Data analysis standards in metabolomics. *Metabolomics*, *3*, this issue.
- Griffin, J. L., Nicholls, A. W., Daykin, C., Heald, S., Keun, H., Schuppe-Koistinen, I., Griffiths, J. R., Cheng, L., Rocca-Serra, P., Rubtsov, D. V., & Robertson, D. (2007) Standard reporting requirements for biological samples in metabolomics experiments: Mammalian/in vivo experiments. *Metabolomics*, *3*, this issue.
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J. H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Le Novere, N., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., & Wang, J. (2003) The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, *19*, 524–531.
- Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A. R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R. J., Hall, R., Kopka, J., Lane, G. A., Lange, B. M., Liu, J. R., Mendes, P., Nikolau, B. J., Oliver, S. G., Paton, N. W., Rhee, S., Roessner-Tunali, U., Saito, K., Smedsgaard, 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*, 1601–1606.
- Jenkins, H., Johnson, H., Kular, B., Wang, T., & Hardy, N. (2005) Toward supportive data collection tools for plant metabolomics. *Plant Physiology*, *138*, 67–77.
- Jones, A. R., Pizarro, A., Spellman, P., & Miller, M. (2006) FuGE: Functional genomics experiment object model. *OMICS: A Journal of Integrative Biology*, *10*, 179–184.
- Kell, D. B. (2006) Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discovery Today*, *11*, 1085–1092.
- Lindon, J. C. (Ed.) (2005). Standardisation of reporting methods for metabolic analyses: A draft policy document from the standard metabolic reporting structures (SMRS) Group. http://www.smrs-group.org/documents/SMRS_policy_draft_v2.3.pdf
- Lindon, J. C., Nicholson, J. K., Holmes, E., Keun, H. C., Craig, A., Pearce, J. T. M., Bruce, S. J., Hardy, N., Sansone, S. A., Antti, H., Jonsson, P., Daykin, C., Navarange, M., Beger, R. D., Verheij, E. R., Amberg, A., Baunsgaard, D., Cantor, G. H., Lehman-McKeeman, L., Earll, M., Wold, S., Johansson, E., Haselden, J. N., Kramer, K., Thomas, C., Lindberg, J., Schuppe-Koistinen, I., Wilson, I. D., Reily, M. D., Robertson, D. G., Senn, H., Krotzky, A., Kochhar, S., Powell, J., van der Ouderaa, F., Plumb, R., Schaefer, H., & Spraul, M. (2005) Summary recommendations for standardization and reporting of metabolic analyses. *Nature Biotechnology*, *23*, 833–838.
- Miller, J., & Mukerji, J., (Eds.) (2003). MDA Guide Version 1.0.1. In: <http://www.omg.org/cgi-bin/doc?omg/03-06-01>, p. 62.
- Morrison, N., Bearden, D., Bundy, J. G., Collette, T., Currie, F., Davey, M. P., Haigh, N. S., Hancock, D., Jones, O. A. H., Rochfort, S., Sansone, S. -A., Stys, D., Teng, Q., Field, D., Viant, & M. R. (2007) Standard reporting requirements for biological samples in metabolomics experiments: Environmental context. *Metabolomics*, *3*, this issue.
- National Research Council (US) (2003). Sharing publication-related data and materials: Responsibilities of authorship in the life sciences. National Academies Press (<http://darwin.nap.edu/books/0309088593/html>)
- Novere, N. L., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., Crampin, E. J., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J. L., Spence, H. D., & Wanner, B. L. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM) 23, 1509.
- Rayner, T., Rocca-Serra, P., Spellman, P., Causton, H., Farne, A., Holloway, E., Irizarry, R., Liu, J., Maier, D., Miller, M., Petersen, K., Quackenbush, J., Sherlock, G., Stoeckert, C., White, J., Whetzel, P., Wymore, F., Parkinson, H., Sarkans, U., Ball, C., & Brazma, A. (2006) A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics*, *7*, 489.
- Sansone, S. A., Schober, D., Atherton, H. J., Fiehn, O., Jenkins, H., Rocca-Serra, P., Rubtsov, D. V., Spasic, I., Soldatova, L., Taylor, C., Tseng, A., & Viant, M. R. (2007) Metabolomics Standards Initiative – Ontology working group – work in progress. *Metabolomics*, *3*, this issue.
- Spellman, P. T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S., Bernhart, D., Sherlock, G., Ball, C., Lepage, M., Swiatek, M., Marks, W., Goncalves, J., Markel, S., Shojatalab, D. I., Pizarro, A., White, J., Hubley, R., Deutsch, E., Senger, M.,

- Aronow, B. J., Robinson, A., Bassett, D., Stoeckert, C. J. Jr., & Brazma, A. (2002). Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biology* 3: research0046.0041-0046.0049.
- Sumner, L. W., Amberg, A., Barrett, D., Beger, R., Beale, M. H., Daykin, C., Fan, T. W. -M., Fiehn, O., Goodacre, R., Griffin, J. L., Higashi, R., Kopka, J., Lindon, J. C., Lane, A. N., Marriott, P., Nicholls, A. W., Reily, M. D., & Viant, M. (2007) Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3, this issue.
- Taylor, C. F., Hermjakob, H., Julian, R. K., Garavelli, J. S., Aebersold, R., & Apweiler, R. (2006) The Work of the Human Proteome Organisation's Proteomics Standards Initiative (HUPO PSI). *OMICS: A Journal of Integrative Biology*, 10, 145–151.
- Taylor, C. F., Paton, N. W., Garwood, K. L., Kirby, P. D., Stead, D. A., Yin, Z. K., Deutsch, E. W., Selway, L., Walker, J., Riba-Garcia, I., Mohammed, S., Deery, M. J., Howard, J. A., Dunkley, T., Aebersold, R., Kell, D. B., Lilley, K. S., Roepstorff, P., Yates, J. R., Brass, A., Brown, A. J. P., Cash, P., Gaskell, S. J., Hubbard, S. J., & Oliver, S. G. (2003) A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nature Biotechnology*, 21, 247–254.
- Taylor, C. F., Field, D., Sansone, S.-A., Apweiler, R., Ashburner, M., Ball, C. A., Binz, P.-A., Booth, T., Brazma, A., Brinkman, R. R., Clark, A. M., Deutsch, E. W., Fiehn, O., Fostel, J., Ghazal, P., Gray, T., Grimes, G., Hardy, N. W., Hermjakob, H., Randall, K. J. J., Kane, M., Kinsinger, C., Kolker, E., Kuiper, M., Novère, N. L., Leebens-Mack, J., Lewis, S. E., McNally, R., Mehrle, A., Morrison, N., Quackenbush, J., Robertson, D. G., Rocca-Serra, P., Rodriguez, H., Smith, B., Snape, J., Sterk, P., & Wiemann, S. (in press). Promoting coherent minimum reporting requirements for biological and biomedical investigations: The MIBBI project. *Nature Biotechnology*
- van der Werf M., Takors R., Smedsgaard J., Nielsen J., Ferenci T., Portais J. C., Wittmann C., Hooks M., Tomassini A., Oldiges M., Fostel J., & Sauer U. (2007) Standard reporting requirements for biological samples in metabolomics experiments: Microbial and in vitro biology experiments. *Metabolomics*, 3, this issue.
- Wettschereck, D., Jorge, A., & Moyle, S. (2003). Visualization and evaluation support of knowledge discovery through the predictive model markup language. In: Knowledge-Based Intelligent Information and Engineering Systems, Pt 1, Proceedings. Springer-Verlag Berlin, Berlin, pp. 493–501.