

# Proposed reporting requirements for the description of NMR-based metabolomics experiments

Denis V. Rubtsov,<sup>a,\*</sup> Helen Jenkins,<sup>b</sup> Christian Ludwig,<sup>c</sup> John Easton,<sup>d</sup> Mark R. Viant,<sup>e</sup> Ulrich Günther,<sup>c</sup> Julian L. Griffin,<sup>a</sup> and Nigel Hardy<sup>b</sup>

<sup>a</sup>Department of Biochemistry, University of Cambridge, The Hopkins Building, Building O, Downing Site, Cambridge, CB2 1QW, UK

<sup>b</sup>Department of Computer Science, University of Wales, Aberystwyth, Ceredigion, SY23 3DB, Wales, UK

<sup>c</sup>The Henry Wellcome Building for Biomolecular NMR Spectroscopy, University of Birmingham, Birmingham, B15 2TT, UK

<sup>d</sup>School of Engineering, University of Birmingham, Birmingham, B15 2TT, UK

<sup>e</sup>School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK

Received 22 June 2006; Accepted 2 October 2006

The amount of data generated by NMR-based metabolomic experiments is increasing rapidly. Furthermore, diverse techniques increase the need for informative and comprehensive meta-data. These factors present a challenge in the dissemination, interpretation, reviewing and comparison of experimental results using this technology. Thus, there is a strong case for unification and standardisation of the data representation for both academia and industry. Here, a systems analysis of an NMR-based metabolomics experiment is presented in order to reveal the reporting requirements. An in-depth analysis of the NMR component of a metabolomics experiment has been produced, and a first round of data standard development completed. This has focussed on both one- and two-dimensional <sup>1</sup>H NMR experiments, but is also applicable to higher dimensions and other nuclei. We also report the modelling of this schema using Unified Modelling Language (UML), and have extended this to a proof-of-concept implementation of the standard as an XML schema.

**KEY WORDS:** metabolomics; NMR; standardisation; reporting requirements; data model.

## 1. Introduction

The use of Nuclear Magnetic Resonance (NMR) spectroscopy based metabolomics has increased dramatically in the last few years in a range of fields including functional genomics, toxicology, and environmental and nutritional studies (Griffin *et al.*, 2001; Nicholson *et al.*, 2002, 2005; Lindon *et al.*, 2003, 2004; Viant *et al.*, 2003; Griffin and Shockcor, 2004). One- and two-dimensional (1D and 2D) <sup>1</sup>H NMR of solution state biofluids or tissue extracts have become some of the most popular tools used for metabolomics, benefiting from being high throughput in nature, relatively cheap on a per sample basis and potentially non invasive. With improvements in automation and flow probe technology, sample throughput for metabolite rich fluids such as urine and plasma is set to increase. Using such approaches, biofluid analysis has recently been used to generate predictive pattern recognition models for detecting early stage hepato- and renal toxicity following the acquisition of data on 150 model hepatic and renal toxins (Lindon *et al.*, 2003).

As NMR spectroscopy is a high-throughput technique, the amount of data generated by this approach is increasing rapidly. The context-dependent nature of

information in metabolomics studies adds complexity to the problem of systematically describing the experiment results. That is, meaningful interpretation of the results of metabolomics experiments is possible only in a specific experimental context that needs to be captured. The presence of many potential, and not always obvious, sources of experimental variation makes it difficult to extract the relevant biological information contained within metabolomic data and requires a detailed experiment description. This presents a challenge to the dissemination, interpretation, reviewing and comparison of these experimental results. Moreover, since several different NMR experiments are used in metabolomics, comprehensive and metadata-rich description plays a crucial role in facilitating adequate cross-comparison and assessment of results.

Thus there is a strong case for unification and standardisation of data representation for NMR-based metabolomics. Indeed, this problem is not limited to metabolomics in general, or NMR-based metabolomics in particular, but also affects the reporting of other functional genomics datasets. In answer to this demand several initiatives have emerged, including the MGED (Microarray Gene Expression Data Society) for transcriptomics (<http://www.mged.org>) (Spellman *et al.*, 2002), the Proteomics Standards Initiative (<http://www.psdev.sourceforge.net/gps/index.html>) for proteomics and the FuGE (Functional Genomics Experi-

\*To whom correspondence should be addressed.  
E-mail: dvr22@cam.ac.uk.

ment) project (<http://www.fuge.sourceforge.net/index.php>) for functional genomics. For metabolomics, SMRS (Standardisation of Reporting Methods for Metabolic Analysis) (Lindon *et al.*, 2005), MIAMET (Minimum Information about a Metabolomics Experiment) (Bino *et al.*, 2004) and ArMet (Architecture for Metabolomics) (Jenkins *et al.*, 2004) were all developed in parallel and are now serving as inputs to the Metabolomics Standardisation Initiative (MSI) that is being orchestrated by the Metabolomics Society (<http://www.metabolomicsociety.org/mstandards.html>). However, these latter initiatives have not yet produced detailed reporting requirements for NMR-based metabolomics experiments.

There are several online databases that allow deposition of NMR experiment results, such as NMR ShiftDB for organic structures (<http://www.nmrshiftdb.org/>) and BioMagResBank (BMRB) (<http://www.bmrb.wisc.edu/>). These databases are mainly built to facilitate deposition of NMR spectra together with various amounts of associated metadata. In addition, data exchange formats for NMR data sets are available from both the CCPN project (Fogh *et al.*, 2002; Vranken *et al.*, 2005), that offers a data model for macromolecular NMR and related areas, and JCAMP-DX (Davies and Lampen, 1993; Lampen *et al.*, 1999). Also, a more general XML (eXtensible Markup Language) (<http://www.w3.org/TR/xml11>) format for analytical chemistry, that is currently in pre-release form, has been developed by the AnIML (Analytical Information Markup Language) (<http://www.animl.sourceforge.net/>) initiative. While these existing initiatives contain valuable content for handling metabolomics data sets, none have been developed based on a detailed systems analysis of an NMR-based metabolomics experiment.

The aim of this work was to perform a systems analysis of NMR-based metabolomics experiments in order to reveal their minimal reporting requirements. This will represent suggested core reporting requirements with the option of user-defined extra information. The results of such a systems analysis will not only enable the development of databases and data handling tools specifically for NMR-based metabolomics experiments, but will also enable proper assessment of the appropriateness of the pre-existing data models and data exchange formats for use in metabolomics data handling. We have produced an in-depth analysis of the NMR component of a metabolomics experiment, and finished the first draft of a data reporting standard. This has focussed on both 1D and 2D  $^1\text{H}$  NMR experiments, but is also applicable to higher dimensions and other nuclei. We also report the modelling of this schema using Unified Modelling Language (UML) (Booch *et al.*, 1999), and have extended this to a proof-of-concept implementation of the standard as an XML schema.

## 2. Scope of the proposed reporting requirements

A typical work flow for metabolomics experiments is depicted in figure 1. This diagram divides the work flow into three major parts:

- (1) The source of sample material (experimental design; selection criteria for cell tissue or biofluid, cultivation or housing of biological source material; collection of samples from the biological source material and extraction of the metabolites that they contain).
- (2) The production of data sets (preparation of extracted samples for analysis by an analytical instrument; chemical analysis using a particular analytical technology; FID and spectral processing; spectral quantitation).
- (3) Statistical analysis and data mining of the data sets to provide answers to the original experimental questions.

The distinction between FID and spectral processing and spectral quantitation can be described as follows. FID and spectral processing is performed upon the raw output data from the analytical instrument. It involves transformation of a raw data set into a representation of the metabolome of the sample usually by mathematical or algorithmic means, i.e. production of an NMR spectrum from a FID (Free Induction Decay). Spectral quantitation is performed upon the data sets that result from FID and spectral processing and aims to summarise them or annotate them with speculative values by either automatic or manual means, e.g. techniques such as “bucketing” (also known as “binning”) and “peak-picking” perform spectral quantitation.

Of these activities, those involved in the production of data sets (contained within the dotted box within figure 1) are dependent on the analytical technology that

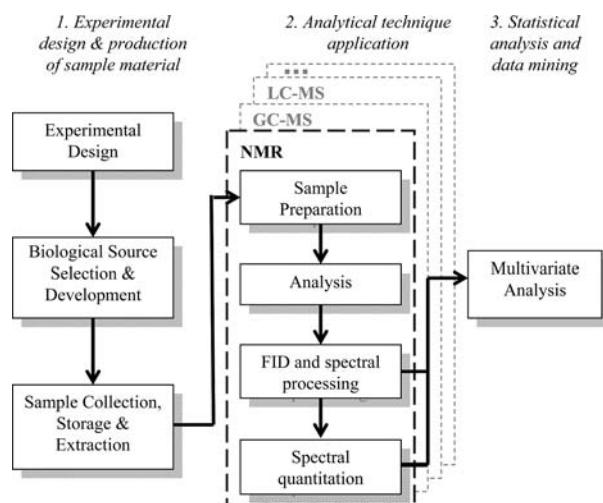


Figure 1. The work flow of a metabolomics experiment.

is used for chemical analysis, i.e. the choice of analytical technology will decide both how the extracted sample should be prepared for chemical analysis and the nature of the data sets that are produced and how they may be processed. The activities outside of the dotted box are not dependent on the analytical technology and may be performed in the same way regardless of the technology chosen for an experiment, i.e. an extracted sample may be divided and prepared separately for presentation to different analytical technologies and the algorithm underlying a data mining or statistical analysis technique will not change simply because the data to which it is applied is produced by a different instrument.

Working on the basis that this initiative would add to a number of pre-existing initiatives that aimed to provide data standards for metabolomics (Bino *et al.*, 2004; Jenkins *et al.*, 2004; Lindon *et al.*, 2005), and in anticipation that it would become part of the recent community-led initiative to provide data standards not only for a range of analytical technologies but also for the complete metabolomics work flow (<http://www.metabolomicsociety.org/mstandards.html>), our systems analysis of NMR-based metabolomics focused on the analytical technology dependent activities involved in the production of data sets. The proposed reporting requirements aim to describe these activities and specify the references to data on the other activities in the metabolomics work flow that are needed to provide a complete description of an NMR-based metabolomics experiment.

A further decision regarding the scope of the systems analysis was to address both one- and two-dimensional NMR experiments. While at present the field of NMR-based metabolomics is dominated by the use of one-dimensional NMR methods, a number of recent studies have highlighted the significant value of two-dimensional NMR in metabolomics (Viant, 2003; Wang *et al.*, 2003; Sandusky and Raftery, 2005). Therefore, by including two-dimensional experiments in the systems analysis we aim to increase the potential of the reporting requirements by enabling description of a new growth area in the field of NMR-based metabolomics. In theory, and because two-dimensional experiments and data are similar in structure to higher dimension experiments and data, we anticipate that these reporting requirements will also be appropriate for describing the third and higher dimensions of an NMR experiment; although higher dimensional experiments are not commonly performed in metabolomics at present.

### 3. Content of the proposed reporting requirements

Systems analysis of the activities involved in the production of data sets involves identification of (a) the structure and content of the output data; and (b) the set of meta-data items that describe how the output data

was produced. The aim is to produce a complete data set that describes an experiment and its results in such a way that its output may be correctly interpreted and used by third parties. In figure 2 we identify, for each activity involved in the production of data sets, the factors about which meta-data should be identified and the output data that must be analysed.

The following discussion of the content of the proposed reporting requirements will be structured according to the concepts provided in figure 2. The meta-data items contained in the proposed reporting requirements are presented in figure 3.

#### 3.1. Meta-data

##### 3.1.1. Sample description

The sample description contains details of the biological sample, together with details of chemicals added to the sample to facilitate its analysis by NMR: one or more solvents added to the sample, chemicals added to the solvent to modify its properties (e.g. a buffer to alter the pH of the sample), an optional chemical shift standard used as an internal reference point for aligning spectra, optional internal standards for metabolite quantification and a field frequency compound to lock the spectrometer frequency. The sample description also contains a reference to information external to the reporting requirements, which describes the history and provenance of the sample prior to its preparation for NMR analysis, i.e. a description of the activities involved in the production of sample material.

##### 3.1.2. Analysis description

For audit purposes, the reporting requirements specify that the date and time of data acquisition and contact details for the experimentalists responsible for the analysis should be recorded.

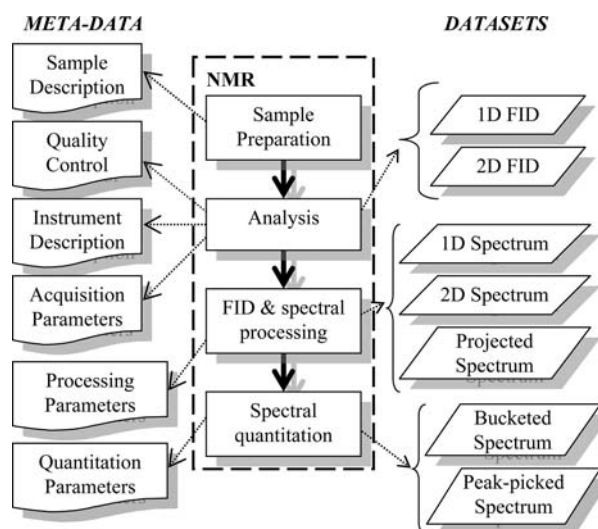


Figure 2. NMR data set production, meta-data and output data.

<p><b>Sample Description</b>  reference to biological sample provenance details  pH of biological sample (optional)  pH of sample after buffer has been added  <i>field frequency lock</i>  chemical name  <i>additional solute</i> (eg. buffer, chelating agent, optional)  chemical name, concentration in sample  <i>solvent</i> (one or more)  chemical name, concentration in sample  <i>chemical shift standard</i> (optional)  chemical name, concentration in sample  <i>concentration standard</i> (optional)  Internal/external, chemical name, concentration in sample</p> <p><b>Instrument Description</b>  geographical location of the instrument  <i>magnet</i>  serial no. (optional), manufacturer, model, field strength  <i>probe</i>  serial no. (optional), manufacturer, model, gradient strength  <i>console</i>  serial no. (optional), manufacturer, model  <i>acquisition computer</i>  serial no. (optional), manufacturer, model, operating system and version no., application software and version no.  <i>autosampler</i> (optional)  serial no. (optional), manufacturer, model, application software and version no.</p> <p><b>Acquisition Parameters</b>  <i>parameters recorded once for each sample</i>  acquisition parameters file reference  shaped pulse file reference (zero or more)  <i>sample details</i>  sample introduction method (tube, flow probe, rotor), size of tube/flow probe/rotor, sample temperature in autosampler (optional), sample temperature in magnet  <i>instrument operation details</i>  sample spinning rate, water suppression technique, pulse sequence name, pulse sequence file reference, pulse sequence literature reference  <i>data acquisition details</i>  number of steady state scans, number of scans, relaxation delay  <i>parameters recorded once for each NMR analysis dimension</i>  <i>instrument operation details</i>  irradiation frequency, acquisition nucleus, 90° pulse width of acquisition nucleus  <i>data acquisition details</i>  dwell time, number of data points acquired  <i>parameters recorded for higher dimensions</i>  the name of the encoding scheme  Hadamard frequency (zero or more)</p>	<p><b>Quality Control</b>  the identity of a signal within the output data  the linewidth (full width at half maximum) of the signal prior to window function processing.  a measurement of the width of the peak at 5% of its total height.</p> <p><b>FID and Spectral Processing Parameters</b>  <i>parameters recorded once when a raw data set is processed</i>  post acquisition water suppression method name (optional)  time-based to frequency-based data transformation method  name of chemical used to reference the spectrum  <i>processing software details</i> (one or more)  processing software name and version number  <i>spectral projection details</i> (zero or one)  method of projection, projection axis  <i>parameters recorded once for each NMR analysis dimension</i>  processing parameters file reference (optional)  number of data points in spectrum  zero order phase correction (optional)  first order phase correction (optional)  calibration reference shift  baseline correction method (optional)  spectral denoising method (optional)  <i>window function details</i>  window function name  function parameter and value (one or more)  <i>parameters recorded for the second dimension only</i>  <i>2D J-resolved processing details</i> (zero or one)  a flag to indicate if the data set has been tilted,  a flag to indicate if the data set as been symmetrised</p> <p><b>Spectral Quantitation Parameter Set</b>  quantitation type  quantitation algorithm  quantitation parameters (optional)  manual Quantitation description (optional)  <i>processing software details</i> (one or more)  processing software name and version number</p> <p><b>Analysis Description</b>  date and time of data acquisition  institution  operator  supervisor</p>
--	--

Figure 3. Meta-data items specified in the proposed reporting requirements (items in bold and italics represent groups and subgroups of data items respectively whilst the data items themselves are in normal font).

### 3.1.3. Instrument description

An NMR instrument typically constitutes a number of components, which may or may not have been constructed by the same manufacturer. As the type of instrument and, in particular, the software used for data acquisition can have an effect on the output data for a sample, the reporting requirements aim to capture this information in the instrument description.

### 3.1.4. Acquisition parameters

Insight into the configuration of an analytical instrument at the time that a sample is analysed is crucial to correctly interpret the output data that is produced. The complete set of instrument parameters can be quite large and will contain a range of values from those that rarely change and have little impact on the output data to those that are highly variable and have a direct impact on the output data. The reporting requirements specify a



small set of important parameters that should be recorded explicitly for each acquisition whilst at the same time requiring a reference to the acquisition parameters file produced by the acquisition software that contains the complete parameter set.

Of note within the acquisition parameters are values for the method of introducing the sample to the instrument and its size, e.g. 1 mm tube, 50  $\mu$ L flow probe. These parameters are included to provide enough information to enable a judgement to be made about the actual quantity and dilution of sample material that has been analysed.

### 3.1.5. Quality control

This type of information can be provided indirectly through description of the quality control procedures in place at the time of analysis of a sample, or directly *via* calculations performed on signals within the output data for the sample, e.g. a signal to noise ratio, the full width at half maximum (FWHM) of a reference peak or the width of a reference peak at 5% of its height. The reporting requirements specify that the latter two of these examples should be provided to enable third parties to make an assessment of the reliability of the data.

### 3.1.6. FID and spectral processing and spectral quantitation

All or part of FID and spectral processing is usually carried out under automation. As with acquisition parameters the full set of processing parameters will often be large and varied. Here again the reporting requirements specify a small set of important parameters that should be recorded explicitly and also require a reference to the processing parameters file produced by the processing software that contains the complete parameter set (where available).

In general methods for FID and spectral processing are better defined and standardised than those for spectral quantitation. The reporting requirements reflect this through specification of much looser descriptive data items for the description of spectral quantitation.

## 3.2. Data sets

The reporting requirements specify that at least one of the following should be provided for an analysis:

- A FID (or a reference to a file containing a FID)
- A spectrum that results from FID and spectral processing
- A spectrum that results from spectral quantitation

The reporting requirements specify the required content of FID and spectra based on the JCAMP-DX format for NMR (Davies and Lampen, 1993; Lampen *et al.*, 1999). This format was designed for spectral data transfer without loss of information. Use of the

JCAMP-DX format during the systems analysis means that JCAMP-DX files may be used to fulfil the data sets part of the reporting requirement in those situations in which JCAMP-DX files are easily exported from an analytical instrument. The decision not to specify the use of JCAMP-DX explicitly means that the reporting requirements can also be fulfilled by other file formats where JCAMP-DX is not an export option, and future evolution of the reporting requirements, for example to support experiments with more than two dimensions, is not dependent on future evolution of JCAMP-DX.

The JCAMP-DX style for spectral representation involves specification of the units of measurement for the axes of a spectrum, the number of data points, and starting and ending values for the  $x$ -axis. The data matrix for a 1D spectrum can then be composed of either  $y$ -values alone (where the complete spectrum is being provided and the  $x$ -axis values can be calculated from the starting and ending values and number of data points) or  $(x,y)$  pairs which allows for the provision of only selected regions of a spectrum, which is a likely requirement for the reporting of NMR datasets by industry. For a 2D spectrum the data matrix is composed of a series of 1D spectra each annotated with a value for the second dimension. JCAMP-DX specifies a similar format for the representation of “peak-picked” spectra. The other common output of NMR spectral quantitation, the “bucketed” spectrum, is not supported by JCAMP-DX. Therefore, our reporting requirements specify the content for “bucketed” spectra following the JCAMP-DX style (see figure 4).

## 4. Discussion

There are several pre-existing initiatives aimed at standardising reporting requirements for both metabolomic experiments and for NMR based experiments. During our systems analysis these initiatives have been considered to ensure compatibility. In addition, our model has benefited from discussions with the wider metabolomics community, including academia and industry, at “MetaboMeetings 1 and 2” in Cambridge, UK in 2005 and 2006 (<http://www.smrsgroup.sourceforge.net/metabomeeting.html>; [http://www.mpdg.org/metabomeeting2/MM2\\_Program.htm](http://www.mpdg.org/metabomeeting2/MM2_Program.htm)).

Considering those initiatives that deal specifically with NMR-based data, there appears to be only a small amount of overlap between the reporting requirements and the schemata for pre-existing NMR spectral databases. The NMRShiftDB spectral library for organic structures and the BMRB database of quantitative data from NMR spectroscopic investigations both aim to provide resources of spectral information for the community to enable users to further their knowledge of biological systems. As such, they support information

<b>Bucketed Spectra</b>	units of measurement on the x-axis, y-axis number of buckets in the spectrum. data matrix for the spectrum (the starting x-axis value for the bucket, the ending x-axis value for the bucket, the x-axis value at the centre of the bucket, the y-axis value)
<b>Peak-picked Spectra</b>	units of measurement on the x-axis, y-axis number of peaks in the spectrum. data matrix for the spectrum (the x-axis value for the peak, the y-axis value for the peak, the type of the peak)

Figure 4. The content for “bucketed” and “peak-picked” spectra items specified in the proposed reporting requirements (items in bold and italics represent groups and sub-groups of data items respectively whilst the data items themselves are in normal font).

that is outside of the scope of our reporting requirements such as molecular descriptions to annotate spectra, whilst at the same time requiring fewer experimental meta-data.

However, there is considerable overlap between our proposed reporting requirements and the information captured in the data model produced by the CCPN initiative. Although CCPN is a low level description of the meta-data associated with an NMR experiment, and was originally designed to describe structural NMR experiments in enough detail to fully define a protein structure, the current compatibility suggests that converters could be produced to extract information from CCPN based databases to populate a minimal description of an NMR based metabolomics experiment that is based on the reporting requirements described here.

There is also substantial common ground between our proposed requirements and the “Technique definition for NMR spectroscopy” described in AnIML. For example, most of the content of “Measurement Parameters” in AnIML can be mapped one-to-one to the items in our “Acquisition Parameters” section. The same applies to “Processing Parameters” and “Instrument” in AnIML, and “Post-Processing Parameters” and “Instrument Description” in our requirements. This implies mutual compatibility in terms of the information content and makes it possible to use AnIML as a format for exchange of data that complies with the reporting requirements without loss.

Similarly, and as mentioned above, JCAMP-DX for NMR may be used as a format for storing most of the data sets detailed in the reporting requirements. In addition, there is considerable overlap between the meta-data items that are listed in the reporting requirements and the JCAMP-DX header information and optional notes fields, making JCAMP-DX for NMR another format that may be used to exchange data sets that comply with parts of the requirements.

In terms of the current standardisation documents related specifically to metabolomics experiments, our reporting requirements comply with the SMRS requirements for sample handling, data acquisition and

instrument level data processing and FID and spectral processing. In this manner our project could be considered as being focused on a subset of the total SMRS description, which we have taken to a formal UML data model, as well as an XML-based implementation (see Current status and future development, below).

Considering previous initiatives within the plant metabolomics community, our reporting requirements also comply with the MIAMET recommendations that resulted from discussions at the International Plant Metabolomics Congress in April 2002 and April 2003 (<http://www.metabolomics-2003.mpg.de/>) and their organisation means that they may readily be implemented as sub-components of an ArMet core implementation, thereby placing them within the context of a complete metabolomics experiment.

Finally, as well as examining compatibility with previous initiatives focussed on describing metabolomic experiments or NMR spectroscopy data, it is important to consider how this work fits in with the wider functional genomic world. FuGE offers a model of the shared components in different functional genomics domains, such as experimental design, sample preparation, subject selection criteria, etc. FuGE has attracted substantial support in the standard development community as a possible common ground for integration of various functional genomics data standards. It has been adopted by MGED and PSI and is under consideration by the MSI. In this respect, we have developed our description so as to be compatible with the wider FuGE description and, following discussions with the FuGE team, they have created a proof-of-concept implementation of the reporting requirements using the current version of FuGE (Andy Jones, personal communication).

## 5. Current status and future development

Our main aim in this project was to generate a proposal for reporting requirements based on a systems analysis of an NMR-based metabolomics experiment. This has led to the design of a UML object model as a

proof-of-concept. The development of a data model was a natural next step in the systems analysis and enabled us to place the descriptions in a more formal context and identify potential problems with its implementation. It has allowed us to identify objects in the domain and specify relationships between them as well as set restrictions on their attributes. Further formalisation of this model has been done in order to implement the UML object model as an XML schema. The full reporting requirements and our example object model are available as Electronic Supplementary Material to the article.

Work on these reporting requirements was started following “MetaboMeeting 1”, Cambridge, 2005, and prior to the creation of the MSI. We anticipate that development of these requirements will be taken forward by the MSI; specifically by the Chemical Analysis Working Group on which the authors have representation. Terminology from the requirements has already been provided as an input to the Ontologies working group on which we also have representation. It is hoped that under the auspices of the MSI these reporting requirements may be refined and improved to produce a data standard that is of use to the metabolomics community as a whole.

### Acknowledgements

This project was partially funded by BBSRC Grant BB/B500107/1 and the BBSRC funded UK Centre for Plant and Microbial Metabolomics (Grant MET20483). Mark R. Viant thanks the Natural Environment Research Council, UK, for an Advanced Fellowship in Metabolomics (NER/J/S/2002/00618). John M. Easton thanks the EPSRC for a PhD studentship (supervised by Dr. Theodoros Arvanitis).

### References

- 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. and Sumner, L.W. (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* **9**(9), 418–425.
- Booch, G., Rumbaugh, J. and Jacobson, I. (1999) *The Unified Modeling Language User Guide*. Addison-Wesley, Reading, MA.
- Davies, A.N. and Lampen, P. (1993). JCAMP-DX for NMR. *Appl. Spectroscopy* **47**(8), 1093–1099.
- Fogh, R., Ionides, J., Ulrich, E., Boucher, W., Vranken, W., Linge, J.P., Habeck, M., Rieping, W., Bhat, T.N., Westbrook, J., et al (2002). *The CCPN project: an interim report on a data model for the NMR community*. *Nat. Struct. Biol.* **9**, 416–418.
- Griffin, J.L. and Shockcor, J.P. (2004). Metabolic profiles of cancer cells. *Nat. Rev. Cancer* **4**, 551–561.
- Griffin, J.L., Williams, H.J., Sang, E., Clarke, K., Rae, C. and Nicholson, J.K. (2001). Metabolic profiling of genetic disorders: a multitissue (1)H nuclear magnetic resonance spectroscopic and pattern recognition study into dystrophic tissue. *Anal. Biochem.* **293**, 16–21.
- Jenkins, H., Hardy, N., Beckmann, M., Draper, J., Smith, A.R., Taylor, J., Fiehn, O., Goodacre, R., Bino, R.J., Hall, R., et al (2004). *A proposed framework for the description of plant metabolomics experiments and their results*. *Nat. Biotechnol.* **22**, 1601–1606.
- Lampen, P., Lambert, J., Lancashire, R.J., McDonald, R.S., McIntyre, P.S., Rutledge, D.N., Fröhlich, T. and Davies, A.N. (1999). An extension to the JCAMP-DX standard file format, JCAMP-DX V.5.01. *Pure Appl. Chem.* **7**(8), 1549–1556.
- Lindon, J.C., Nicholson, J.K., Holmes, E., Antti, H., Bollard, M.E., Keun, H., Beckonert, O., Ebbers, T.M., Reily, M.D., Robertson, D., et al (2003). *Contemporary issues in toxicology the role of metabolomics in toxicology and its evaluation by the COMET project*. *Toxicol. Appl. Pharmacol.* **187**, 137–146.
- Lindon, J.C., Holmes, E. and Nicholson, J.K. (2004). Toxicological applications of magnetic resonance. *Progr. NMR Spectroscopy* **45**, 109–143.
- Lindon, J.C., Nicholson, J.K., Holmes, E., Keun, H.C., Craig, A., Pearce, J.T., Bruce, S.J., Hardy, N., Sansone, S.A., Antti, H., et al (2005). *Summary recommendations for standardization and reporting of metabolic analyses*. *Nat. Biotechnol.* **23**, 833–838.
- Nicholson, J.K., Connelly, J., Lindon, J.C. and Holmes, E. (2002). Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **1**, 153–161.
- Nicholson, J.K., Holmes, E. and Wilson, I.D. (2005). Gut microorganisms, mammalian metabolism and personalized health care. *Nat. Rev. Microbiol.* **3**, 431–438.
- Sandusky, P. and Raftery, D. (2005). Use of selective TOCSY NMR experiments for quantifying minor components in complex mixtures: application to the metabolomics of amino acids in honey. *Anal. Chem.* **77**, 2455–2463.
- Spellman, P.T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S., Bernhart, D., Sherlock, G., Ball, C. and Lepage, M., et al. (2002) Design and implementation of microarray gene expression markup language (MAGE-ML) *Genome Biol.* **3**(9), RESEARCH0046.
- Viant, M.R., Rosenblum, E.S. and Tjeerdema, R.S. (2003). NMR-based metabolomics: a powerful approach for characterizing the effects of environmental stressors on organism health. *Environ. Sci. Technol.* **37**, 4982–4989.
- Viant, M.R. (2003). Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.* **10**, 943–948.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696.
- Wang, Y.L., Bollard, M.E., Keun, H., Antti, H., Beckonert, O., Ebbers, T.M., Lindon, J.C., Holmes, E., Tang, H.R. and Nicholson, J.K. (2003). Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning H-1 nuclear magnetic resonance spectroscopy of liver tissues. *Anal. Biochem.* **323**, 26–32.