

SCIENTIFIC DATA

OPEN

Comment: A decade after the metabolomics standards initiative it's time for a revision

Rachel A. Spicer¹, Reza Salek¹ & Christoph Steinbeck^{1,2}

Received: 19 June 2017

Accepted: 22 August 2017

Published: 26 September 2017

A recent analysis of publicly available metabolomics data shows that the MSI guidelines are not well abided to in publicly shared metabolomics studies. We propose that the MSI guidelines should be revisited and revised, as has been done in other communities, to fit the current community needs.

In the year 2005, leading experts in metabolomics, with support from the Metabolomics Society, gathered to discuss metabolomics experimental standards and formed the metabolomics standards initiative (MSI)¹. Two years later, the MSI published a series of reporting standards for metabolomics^{2–8}. Another 5 years passed until a network of global repositories^{9–11} for metabolomics data arose, to which the MSI guidelines could be applied. Now, as of 2nd August 2017, there are more than 700 metabolomics datasets publicly available via MetabolomeXchange (<http://www.metabolomexchange.org/>), which, for the first time, allows for an assessment of metabolomics data sharing and compliance with the MSI guidelines.

In the decade since the first publication of the MSI reporting standards, the metabolomics community has matured. There is now greater understanding of genetic and environmental factors that can induce significant effects on the metabolome. The general purpose metabolomics data repositories MetaboLights⁹ and Metabolomics Workbench¹⁰, along with the smaller more specific repositories, MetaPhen¹² and MeRy-B¹³, were designed to adhere with the MSI guidelines for minimum metadata reporting. We have performed a study assessing the compliance of the datasets in the MetaboLights, Metabolomics Workbench, MetaPhen, and MeRy-B repositories. It was found that many of the MSI's minimal reporting standards are not complied with by many of the studies in these repositories (Fig. 1 shows examples from the MetaboLights repository). There were no MSI standards that were complied with in every publicly available study. There are, however, minimal reporting standards that are not applicable to every study e.g., treatment, which is only relevant to studies including a treatment. Being only applicable to a small subset of studies may result in the low percentage compliance to these standards (Fig. 1).

While all databases under investigation claim to demand compliance with the MSI guidelines, the true compliance is unexpectedly low. We have detailed our finding in Spicer *et al.*¹⁴ and will comment only on the main points below.

The current MSI reporting standards suffer from a variety of limitations. They are difficult to interpret and do not capture all metadata required for re- and further data analysis. Between the biological context metadata subgroups there is also a lack of consistency as to what constitutes a minimal and best practice reporting standard. Additionally there is unnecessary repetition of standards between the biological context metadata and the chemical analysis working group (CAWG) guidelines³.

Compliance with minimum information standards and related database submission procedures has always been a source for despair. Unless experiments and technological setup are carefully planned to take into account later data and metadata submission procedure, the experience of converting lab book entries into an ISA¹⁵ compliant dataset ready and acceptable for deposition is cumbersome, to say the least. Nevertheless, one would assume that the guidance of groups of world-leading experts in metabolomics in the form of minimum information (MI) standards would be enthusiastically welcomed and used. This is not the case. We believe that a major reason for this is the sheer inconvenience of

¹European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge CB10 1SD, UK. ²Friedrich-Schiller-University, Fürstengraben 1, Jena 07743, Germany. Correspondence and requests for materials should be addressed to R.A.S. (email: spicer@ebi.ac.uk) or to C.S. (email: christoph.steinbeck@uni-jena.de)

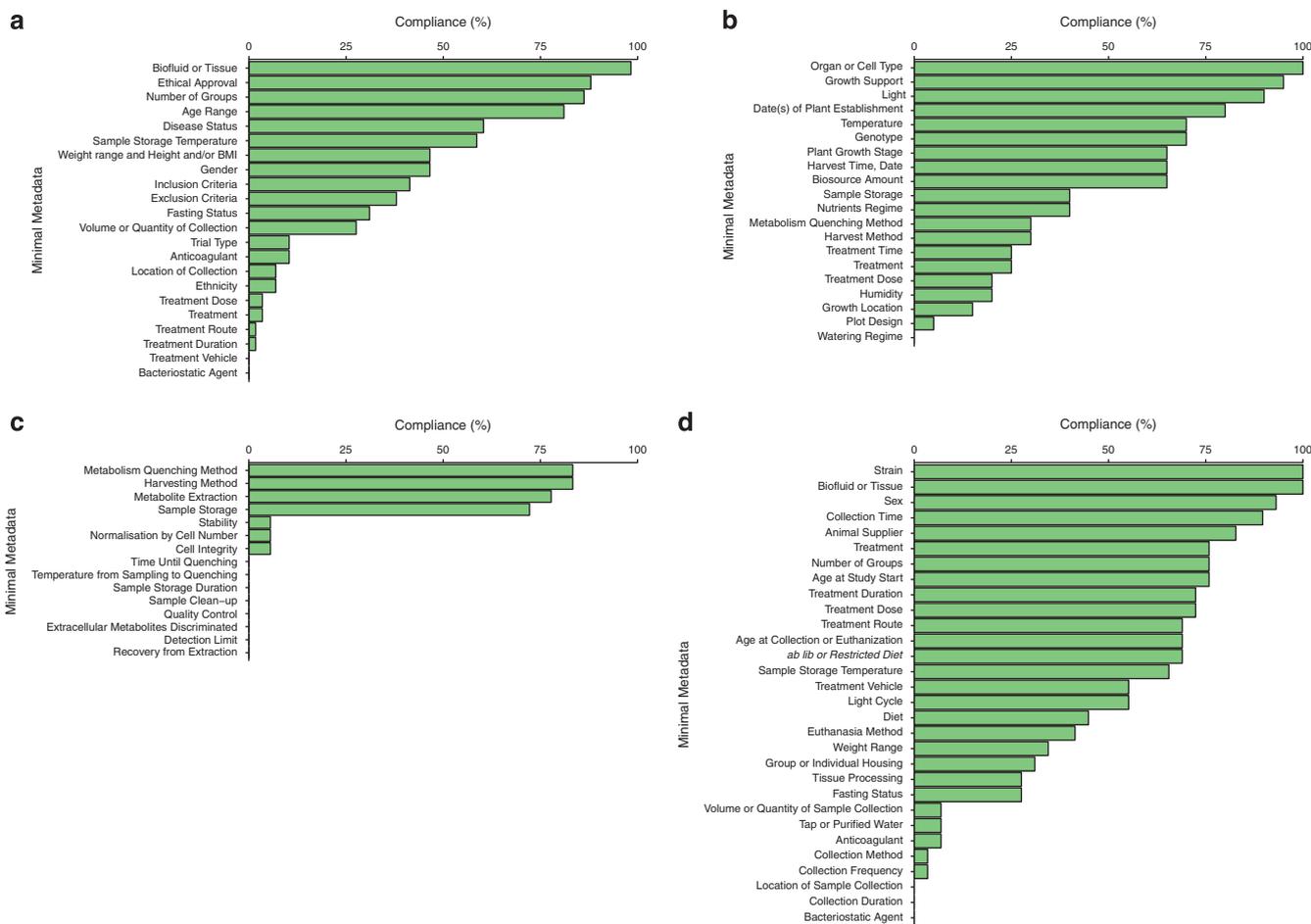


Figure 1. The percentage of studies in MetaboLights that comply with the MSI minimal reporting standards. (a) Clinical *Homo sapiens* studies compliance with the mammalian clinical trials and human studies minimal reporting standards. (b) *Arabidopsis thaliana* studies compliance with the plant minimal reporting standards. (c) *In vitro H. sapiens* studies using human cells and cell lines compliance with the microbial and *in vitro* minimal reporting standards. (d) Pre-clinical *Mus musculus* studies compliance with the pre-clinical minimal reporting standards. The percentage compliance with the MSI reporting standards is similar for all repositories (see Spicer *et al.*¹⁴).

recording rich meta-data, which might not seem relevant for one's own study at a given point in time. Furthermore, repositories do not agree on the set of metadata required and, in particular during their maturation stage, do not vigorously enforce MI standard compliance in fear of annoying submitters.

There have been efforts by the Metabolite Identification Task Group of the Metabolomics Society to reassess the reporting standards for metabolite identification¹⁶ and the community has suggested several improvements^{17–19}. The recently launched M**ET**abolomics standa**R**ds Initiative in Toxicology (MERIT) (<http://www.ecetoc.org/topics/standardisation-metabolomics-assays-regulatory-toxicology/>) will define best practice and minimal reporting standards for the application of metabolomics to regulatory toxicology. However, there have not been advances in the reporting standards for other areas of metabolomics. We believe that now, after putting the MSI biological context guidelines to test in ~400 publically available datasets, the MSI guidelines should be revisited and revised, and dataset submission procedures in MetabolomeXchange should be harmonized. The community, along with data curators, publishers and funders should be consulted. Consideration should be given to the minimum amount of metadata required to be able to a) repeat an experiment and b) re-analyse the data. The data that are now publically available in a number of open-access repositories around the globe will be a treasure trove for guiding the selection of minimal reporting standards.

References

1. Fiehn, O. *et al.* The metabolomics standards initiative (MSI). *Metabolomics* **3**, 175–178 (2007).
2. Goodacre, R. *et al.* Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **3**, 231–241 (2007).
3. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).

4. van der Werf, M. J. *et al.* Standard reporting requirements for biological samples in metabolomics experiments: microbial and in vitro biology experiments. *Metabolomics* **3**, 189–194 (2007).
5. Griffin, J. L. *et al.* Standard reporting requirements for biological samples in metabolomics experiments: mammalian/in vivo experiments. *Metabolomics* **3**, 179–188 (2007).
6. Morrison, N. *et al.* Standard reporting requirements for biological samples in metabolomics experiments: environmental context. *Metabolomics* **3**, 203–210 (2007).
7. Fiehn, O. *et al.* Minimum reporting standards for plant biology context information in metabolomic studies. *Metabolomics* **3**, 195–201 (2007).
8. Rubtsov, D. V. *et al.* Proposed reporting requirements for the description of NMR-based metabolomics experiments. *Metabolomics* **3**, 223–229 (2007).
9. Haug, K. *et al.* MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* **41**, D781–D786 (2013).
10. Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* **44**, D463–D470 (2016).
11. Salek, R. M. *et al.* COordination of Standards in MetabOmicS (COSMOS): facilitating integrated metabolomics data access. *Metabolomics* **11**, 1587–1597 (2015).
12. Carroll, A. J. *et al.* PhenoMeter: A Metabolome Database Search Tool Using Statistical Similarity Matching of Metabolic Phenotypes for High-Confidence Detection of Functional Links. *Front Bioeng Biotechnol* **3**, 106 (2015).
13. Ferry-Dumazet, H. *et al.* MeRy-B: a web knowledgebase for the storage, visualization, analysis and annotation of plant NMR metabolomic profiles. *BMC Plant Biol.* **11**, 104 (2011).
14. Spicer, R. A. *et al.* Compliance with minimum information guidelines in public metabolomics repositories. *Sci. Data* **4**, 170137 doi: 10.1038/sdata.2017.137 (2017).
15. Rocca-Serra, P. *et al.* ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics* **26**, 2354–2356 (2010).
16. Creek, D. J. *et al.* Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics* **10**, 350–353 (2014).
17. Schymanski, E. L. *et al.* Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* **48**, 2097–2098 (2014).
18. Sumner, L. W. *et al.* Proposed quantitative and alphanumeric metabolite identification metrics. *Metabolomics* **10**, 1047–1049 (2014).
19. Everett, J. R. A new paradigm for known metabolite identification in metabonomics/metabolomics: metabolite identification efficiency. *Comput. Struct. Biotechnol. J* **13**, 131–144 (2015).

Additional information

Competing interests: The authors declare no competing financial interests.

How to cite this article: Spicer, R. A. *et al.* A decade after the metabolomics standards initiative it's time for a revision. *Sci. Data* **4**:170138 doi: 10.1038/sdata.2017.138 (2017).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017