

## **Data Deposition as an Integral Part of the Publication Process**

**Sandra Orchard\***

EMBL Outstation - European Bioinformatics Institute,  
Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

\*Corresponding author: Sandra Orchard, EMBL Outstation - European Bioinformatics  
Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK,  
Fax: +44-(0)1223-494-468; E-mail: [orchard@ebi.ac.uk](mailto:orchard@ebi.ac.uk)

Received July 14, 2009; Accepted July 24, 2009; Published July 25, 2009

**Citation:** Orchard S (2009) Data Deposition as an Integral Part of the Publication Process. J Proteomics Bioinform 2: 334-335.  
doi:10.4172/jpb.1000092e

**Copyright:** © 2009 Orchard S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Keywords:** Data standardization; Human proteome organization; Proteomics standards initiative

Proteomics data has been available in published literature for approximately 15 years and, until very recently, its appearance in a journal manuscripts has been the only route for making the results of such work accessible to the research community. Whilst journal articles are undeniably the best medium for telling the “story” of an experiment, for summarising the methods, collating the results and describing the conclusions drawn from those results, they are not the most appropriate method for publishing the supporting, complex experimental data – the long lists of protein identifications generated by a proteomics experiment and the underlying spectra from which they are derived. Usually only a small subset of such data can be included in a journal article and the majority of the information will be added as Supplementary Materials, if at all, making it inaccessible to any search engine. This limits the value of the many thousands of experiments performed since the inception of the field of mass spectrometry based proteomics science, as the resulting data are not readily available for the community to search, compare to or cite as they perform related work. Assembling the proteome of a cell, or comparing a disease state against a previously published healthy state tissue protein list is a far from trivial exercise when the requisite information is scattered over multiple pdf files, Excel sheets or author-maintained websites.

An alternative to this state of affairs became available when a number of public domain databases were established to collect this data, repositories such as PRIDE (Jones et al., 2008) which collect both the spectra and

protein identifications attached to each publication. In order for these services to be established, common data formats were established to enable data from many different sources to be collected. This work was directed by the Proteomics Standards Initiative of the Human Proteome Organisation (HUPO-PSI, [www.psidev.info](http://www.psidev.info)) over the last 7 years and has involved a broad spectrum of the proteomics community in either preparing these standards or in consultation over their implementation. Community-based standards now exist in the areas of mass spectrometry, protein separation (both gel and column-based), proteomics informatics, protein modifications and molecular interactions. A series of publications describe these interchange formats and, of equal importance, the subsequent creation of an increasing number of tools and resources which are currently building on this work. The Minimum Information About a Proteomics Experiment (MIAPE) documents (Taylor et al., 2007), detailing the information which should be included in any publication ensuring that a reader can comprehend both how an experiment was performed and the conclusions drawn from it, have also been released. Submission tools for proteomics data repositories are now also in existence and are easing the path towards making data submission to a public domain repository an integral part of the publication process. There is currently a groundswell of opinion that the time is ripe for the journals publishing in this area of interest to actively support this process and encourage both conformity to the MIAPE guidelines, and also data deposition in the public domain as papers are being prepared for publication.

Recognising that establishing a consistent policy on data submission across journals would have obvious advantages for the potential author, a number of Senior Editors from relevant journals met with the HUPO-PSI and HUPO Publication committee, in Turku, Finland in April 2009 in an open meeting to discuss the way forward (Orchard et al., 2009 in preparation). Previous discussions had decided it was too premature to request data deposition into a public domain repository be part of the publication process, as there were not sufficient tools and resources to assist the small-scale laboratories to make this practicable. Now, however, it was decided that this hurdle has largely been overcome, with robust, open-source tools becoming increasingly available to assist in importing data generated by proteomics laboratories into public repositories. During this current round of discussions, it was agreed that the MIAPE reporting guidelines be further refined, in consultation with the domain-specific journals, to differentiate between the information required within a manuscript and the additional information which should be added to a database submission. Once this has been achieved, the way forward is clear for such journals to recommend data deposition as part of the submission process. It is now up to the scientific community to respond to this challenge, to make their data available for search and retrieval by researchers in the field and support the efforts of this journal, in encouraging data sharing and availability by ensuring that their published work is deposited in the public domain. The benefits of data availability to the individual scientist are many, including increased opportunities for collaboration and an increased citation rate as released data becomes the benchmark for

the next series of experiments. To the community as a whole, it is invaluable; with the proteome content of a particular cell under well-defined sets of experimental conditions being replicated and confirmed by multiple, independent experiments being performed in laboratories across the world and now becoming readily available for comparison.

The HUPO-PSI has prepared the interchange standards to make the deposition of data supporting a journal article both possible and practicable, the open-source community has responded with deposition tools to enable easy compliance to these standards. The domain-specific journals will increasingly be requesting that authors follow this route to publication and there is a clear advantage to the user community to having this data readily accessible. The individual scientist must now be the one who decides to respond to this drive towards data availability and to recognise that effectively restricting public access to valuable data is of benefit to no-one. Continuing to 'publish and vanish', and restrict public access to valuable data should no longer be considered an option by this community.

## References

1. Jones P, Cote RG, Cho SY, Klie S, Martens L, et al. (2008) PRIDE: new developments and new datasets. *Nucleic acids Research* 36: d878-883» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK Jr, et al. (2007) The minimum information about a proteomics experiment (MIAPE). *Nature Biotechnology* 25: 887-893» [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)