

Editorial: Data Repositories, Meeting Proceedings, Open Access/Online Publications, and First Reporting Issues

What constitutes the first reporting of a scientific finding or data? The two *Letters to the Editor* that follow this editorial illustrate a growing problem for many of us that rely on standard literature searches for information that we use to design our experiments and to report our results. As a result of an ever increasing number of data repositories, online publication of meeting abstracts and proceedings, and number of Open Access (OA) (primarily) online journals, it may be difficult through standard literature searches to identify all previous reports of specific applications or experimental results. In this case, Bellido et al. (2014) indicate that “to the best of our knowledge” their article was the first to report the application of the Richardson-Lucy algorithm to resolve plasmonic resonances in electron energy loss spectra (EELS) obtained with a monochromated electron beam. As many investigators would do, they relied on a search of the commercial online indices, Thomson Reuter’s *Web of Science* and Elsevier’s *Science Direct*, for identification of published materials concerning their topic of interest. Unfortunately, these search indices did not identify materials on this topic presented by Walther et al. (2012, 2013) in the form of meeting Proceedings. Without prior knowledge of the Walther publications this made location of the Proceedings reports difficult to discover, as I found when searching these sources based on the key words “Richardson-Lucy”, plasmonic resonance, and EELS. I strongly believe that in this case there was no intent to mislead or improperly claim the first report using the algorithm, which is why both *Letters to the Editor* have been published here.

In further studying the issue of what constitutes the first reporting of data I was quickly overwhelmed with information about options for placing data in data repositories, the variety of options for rapid publication of research results, copyright issues, and acceptable use of various forms of information that can be obtained online. By definition, a Data Repository is an online database used by institutions and organizations to capture, preserve, and provide access to the intellectual output of a scholarly community (<http://ucblibraries.colorado.edu/scholarlycommunications/oa/repositories.htm>) with the goal of providing access to a broad range of materials including data sets, articles and books, documents, technical reports, presentations, conference proceedings, creative activities, master’s theses, open-access dissertations, and more (<http://scholarcommons.sc.edu/>). A large number of universities such as The University of Colorado (<http://ucblibraries.colorado.edu/scholarlycommunications/oa/repositories.htm>), Purdue University (<https://purr.purdue.edu/>), University of Minnesota (<https://www.lib.umn.edu/datamanagement/datacenters>), and the University of South Carolina (<http://scholarcommons.sc.edu/>) host institutional data repositories designed to provide access to an array of materials.

In addition to institutional based repositories, other topical repositories such as Biosharing (<http://biosharing.org/>) for deposit of data related to biology, natural, and biomedical sciences, the MODERN repository (Modeling the Environmental and Human Health Effects of Nanomaterials) (<http://modern-fp7.biocent.cat/doc/MODERN%20D2.1.pdf>) for data applicable to nanomaterials, and arXiv (<http://arxiv.org/>) for material in mathematics, physics, astronomy, computer science, quantitative biology, statistics and quantitative finance are also available. Efforts to ameliorate the problem of searching so many data repositories for relevant materials include the creation of entities that list large numbers of repositories at a single site such as Databib (<http://databib.org>) and the Registry of Research Data Repositories (re3data), which indexes research data repositories and currently lists over 600 data repositories from around the world covering all academic disciplines (<http://www.re3data.org>). In addition, DataCite is a recent collaborative effort between Databib, re3data,

and others to further improve accessibility of the many repositories with the aim of the merger to reduce duplication of effort and to better serve the research community with a single, sustainable registry of research data repositories (<http://www.datacite.org/>). These large registries often have oversight boards, but nearly anything can be deposited with limited oversight.

Investigators should exercise caution when citing non-peer reviewed data and manuscripts available in many repositories, and when depositing their own data or early versions of manuscripts into repositories. As noted in the Wikipedia material for arXiv, posted material is not peer reviewed (<http://en.wikipedia.org/wiki/ArXiv>). This is also true for many, if not the great majority of other data repositories. While it is important to have rapid access to ongoing studies, it is also essential to understand the goal of repositories in providing very rapid access of data and early versions of manuscripts, and to use this information appropriately. We would all like to believe that intentional fraud does not occur in science, but a quick search of the Retraction Watch website (<http://retractionwatch.com/>) proves that even after the process of careful peer review all data must be considered carefully.

A recent effort to address the issue of dataset publication is that of *Nature and Scientific Data* (<http://www.nature.com/sdata/>), which "is an open-access, online-only publication for descriptions of scientifically valuable datasets, and exists to help researchers publish, discover and reuse research data. *Scientific Data's* main article-type is the Data Descriptor: peer-reviewed, scientific publications that provide an in-depth look at research datasets. Data Descriptors are a combination of traditional scientific publication content and structured information curated in-house, and are designed to maximize reuse and enable searching, linking and data mining. Each is peer-reviewed under the supervision of our Editorial Board." *Science Data* requires placement of datasets in a repository, but does not host the data. They do provide guidelines for data deposition and a list of approved data repositories (<http://www.nature.com/sdata/data-policies/repositories>).

Another effort addressing the rapid growth of digital publishing is FORCE11 (The Future of Research Communications and e-Scholarship), which is a community of scholars, librarians, archivists, publishers and funding agencies that has the goal of facilitating the effective use of information technology in modern scholarly communication (<https://www.force11.org/about>). FORCE11 has published a list of eight Data Citation Principles including Importance, Credit and Attribution, Evidence, Unique Identification, Access, Persistence, Specificity and Verifiability, and Interoperability and Flexibility. Further discussion of each of these topics can be found at <https://www.force11.org/datacitation>. An important point in the FORCE11 Preamble is accessibility of data, which again addresses a key component of this Editorial.

An additional consideration, which some authors have faced when publishing in *Microscopy and Microanalysis*, is the issue of copyright when material has previously been placed in repositories such as arXiv. For many society-owned journals, such as *Microscopy and Microanalysis* which is owned by the Microscopy Society of America, copyright transfer must be made to the Society upon publication. When posting early versions of manuscripts in repositories there are a variety of different copyright statuses that may be implied, some of which may preclude subsequent transfer of the copyright to the society. Prior to posting early versions of manuscripts in a repository, investigators should examine copyright implications if they intend to publish in a journal that requires transfer of copyright to that publisher, society, or publication. They must also be aware of copyright restrictions which prevent placing accepted or published manuscripts into a repository when copyright has been transferred from the author to the publisher or society. A full discussion of these restrictions for *Microscopy and Microanalysis* can be found on the Copyright and Restrictions page (<http://journals.cambridge.org/action/displaySpecialPage?pageId=4676#>) on the Cambridge University Press website.

While it is certainly an admirable goal for all data to be copyrighted for investigator protection and rapidly and readily available for researchers to study and utilize, it is apparent that in many cases our efforts to make our results available have overwhelmed our ability as researchers to identify and utilize all available resources for comparison and

reporting of our data with respect to the data of others. This brings us back to the original question posed above concerning “what constitutes the first reporting of a scientific finding or data?” Should datasets that have been submitted to institutional and/or other non-peer reviewed data repositories such as those discussed above, meeting proceedings published online, etc. be considered the first reporting of a finding? I believe to fully claim the first reporting of a finding that it must be published in a peer reviewed publication where ownership of copyright (whether society, publisher, or retained by the author) is clearly understood. Every effort should be made to search for these publications through relevant comprehensive search engines such as PubMed, Web of Science, Science Direct, or Google Scholar. Additional use of topic focused large repository registries such as DataCite, re3data, Databib, or Biosharing are also encouraged and should become standard practice. However, it is still conceivable that a report such as that by Dr. Walther that is published in a meeting Proceeding might be missed. Dr. Bellido’s group properly qualified their statement that this was the first report “to the best of our knowledge” which is the appropriate way to handle claims of this nature.

The use of digital media for the reporting of our data, regardless of our specific research discipline, will certainly continue to increase. This will require all of us to be diligent in identifying the best online resources for our specific research disciplines. Future editorials will assess advancements in the use of digital media in analyzing and reporting our data as developments occur.

Robert L. Price
Editor-in-Chief, Microscopy and Microanalysis
Department of Cell Biology and Anatomy School of Medicine
University of South Carolina

REFERENCES

- BELLIDO, E.P., ROSSOUW, D. & BOTTON, G.A. (2014). Toward 10 meV Electron Energy-Loss Spectroscopy Resolution for Plasmonics. Toward 10 meV Electron Energy-Loss Spectroscopy Resolution for Plasmonics. *Microscopy and Microanalysis* **20**(3), 767–778.
- WALTHER, R., MÜLLER, E., SCHNEIDER, R. & GERTHSEN, D. (2013). Accessing low-energy cavity modes in rectangular slits in a thin Au film by STEM-EELS in combination with the Richardson-Lucy algorithm. In *Proceedings of the MC 2013*, Rachel, R. (Ed.), pp. 174–175.
- WALTHER, R., SCHNEIDER, R., MÜLLER, E., CARMELI, I., RICHTER, S., MANIV, T., COHEN, H. & GERTHSEN, D. (2012). Surface plasmon polariton mediated collective excitations of cavity resonances in gold sub-wavelength slit-arrays. In *Proceedings of the EMC 2012*, PS1.1.