

## EDITORIAL

# Making Data Accessible: The Dryad Experience

Gary W. Miller<sup>1</sup>

Editor-in-Chief, Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322 and <sup>1</sup>For correspondence. E-mail: gary.miller@toxicology.org

*"There is no idea, however ancient or absurd, that is not capable of improving our knowledge."*

—Paul Feyerabend (2010)

The operative term in this quote from the late provocateur/philosopher of science is "capable." While all ideas and data are capable of improving knowledge, there is no guarantee that they will actually do so. From the beginning of time the crucial first step was being aware of the ideas. A century ago this was limited by the ability to gain access to books, journals, and colleagues. Overcoming these limitations was straightforward; gain access to the published work and interact with others in the field. Subsequently possessing the ability to overcome one's own biases and parochial worldviews to recognize that the ideas of others have value is an entirely different matter, but the point here is that we must first be able to access this information.

The data explosion that we are witnessing in science can be considered a triumph of human ingenuity or a tragedy of futility. The generation of new information "should" be viewed positively. Data derived from well-designed experiments has great potential to advance knowledge. However, if we do not aggressively work to build the framework to organize and exchange this information then it will be for naught. Indeed, it could well lead to the opposite of the desired effect; an avalanche of useless information that will just stifle the scientific enterprise. We can no longer rely on library access and scientific conferences to provide access to useful scientific information.

## THE STATUS QUO WILL CRIPPLE US

Historical means of archiving and storing data included file cabinets, supplemental data sections, and conference proceedings. In the past couple of decades investigators have resorted to setting up their own websites or servers. These piecemeal systems often evolve out of necessity, but they tend to be static or subject to misplacement, obsolescence, or evaporation. Within genomics and related off-shoots, there has been a robust and sophisticated data repository and sharing network, much of which had its genesis in the Human Genome Project (Green et al., 2015). This

situation has spurred some large research projects to establish centralized data repositories, but for the most part research in the toxicological domain has remained woefully inaccessible and bound in files not easily shared. This is a travesty given the importance of toxicological data in driving major regulatory and public health decision-making. Furthermore, accessibility to data is critical to reproducibility (Miller, 2014).

A previous editorial of mine addressed the importance of data sharing and dissemination of information to the toxicological and scientific communities (Miller, 2015). Data represent a continuum of complexity, from raw and original data that could have been found within lab notebooks, lab websites, or in supplemental data sections to more complex datasets that are generated from -omic technologies or code from software programming. Easy access to high quality and shareable files will require a considerable level of effort, time, and money for the proper set up and maintenance of large datasets resulting in the creation of a virtual and perpetual data repository.

## DRYAD DIGITAL REPOSITORY: A NEW OFFERING FOR TOXICOLOGICAL SCIENCES

We are pleased to announce that *Toxicological Sciences* now offers data deposition to our submitting authors using Dryad Digital Repository (Dryad; datadryad.org). Dryad is an established organization that already partners with several major publishing houses and journals. Datasets will be deposited on the Dryad server. Dryad then provides an initial level of curation, assignment of a Digital Object Identifier (DOI), and long-term data storage. The DOI transcends URLs, which can quickly become defunct from a change in server or domain, or from a lack of support and maintenance.

### Three key points about Dryad Data Repository

- Supported by major publishers and scientific organizations
- Mature system developed in collaboration with scientists and publishers

- Provides a virtual home for data sets that is readily accessible and citable

The data deposited with Dryad can be referenced in the present manuscript and others in the future. While the manuscript is under evaluation, reviewers will be given access to the Dryad server to examine the data. Reviewers will not specifically review the datasets, but rather be given unfettered access to the raw information. Data will not be publically available until the manuscript has been accepted for publication. If the manuscript is rejected, the dataset will be erased from the Dryad server with no further obligation to the authors or journal, or the submitting authors can leave the data there and pay the deposition fee. Authors will have the assurance that the dataset will be available for a minimum of 10 years with no additional cost (the initial pricing model, but the goal is that the timeline will be extended). This allows their valuable data to be made available to the scientific public in a manner that facilitates quality control and apportionment of credit.

It is difficult to predict what the future will hold, but there is no doubt that generation of data will only increase. It is likely that properly deposited and archived datasets will be able to be interrogated with various data-mining approaches. In 5–10 years data archiving and sharing may become an obligatory aspect of scientific publishing, but for now *Toxicological Sciences* will be offering this service on a voluntary basis with no cost to

the authors (costs will be absorbed by ToxSci). Other fields have embraced these approaches and the field of toxicology should not live in a continual state of catching up. The hope is that when the topics of data sharing, data accessibility, and archival sites like Dryad come up and you are asked “Are you experienced?” that you are able to say, “Yes!” Jimi and I would appreciate it.

## ACKNOWLEDGMENTS

The author would like to thank Dr. Patti J. Miller and Dr. David Crotty for their helpful comments and the late Jimi Hendrix for providing inspiration for the title.

## REFERENCES

- Feyerabend, P. (2010). *Against Method*, 4th ed. Verso Books, New York, NY (originally published in 1975).
- Green, E. D., Watson, J. D., and Collins, F. S. (2015). Human Genome Project: Twenty-five years of big biology. *Nature* **526**, 29–31.
- Miller, G. W. (2014). Improving Reproducibility in toxicology. *Toxicol. Sci.* **139**, 1–3.
- Miller, G. W. (2015). Data sharing in toxicology: beyond show and tell. *Toxicol. Sci.* **143**, 3–5.