

Open Data For The Win!

Scott Edmunds 

Published October 2, 2013

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/ybs8m-2p298>

Keywords

Open Access, Epigenomics, GigaDB, Methylation, Nematode
Feature Image

Copyright

Copyright © Scott Edmunds 2013. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Depositing data in GigaDB helps authors win BMC Open Data Award by boosting confidence in unexpected research findings

Last night at the [Beyond the Genome](#) conference in San Francisco, researchers were presented with this year's [BioMed Central Open Data Award](#) for their work demonstrating that DNA methylation occurs in the parasitic worm *Trichinella spiralis*, a human pathogen also known as "pork worm" due to it being found in undercooked pork products. One of the challenges in getting researchers to put in the time and effort to make their data available in a curated and usable form is a perceived lack of incentives. Journal and funder policies to data release only cover a limited number of data types and are [poorly enforced](#), and carrot approaches may be a more constructive way to encourage data producers to make more of their data available. On top of [scientometric](#) work demonstrating directly to authors the increase of citations and credit they will receive upon making their supporting data available (a great example being Heather Piwowar and Todd Vision's latest paper on "[Data reuse and the open data citation advantage](#)"), competitions and prizes are another way of positively reinforcing and incentivizing good behaviour, so the BMC Open Data award is a very welcome asset in this area.

Presented by [BMC](#) and sponsor [LabArchives](#) to recognize authors who have demonstrated leadership in the sharing, standardization, publication, or re-use of biomedical research data, this is the 4th time the award has been presented. [The article](#), published in our sister BMC-journal *Genome Biology*, has shaken up the field of epigenetics by shattering the assumption that DNA methylation is absent in nematodes, the tiny worm species that serves as an important model organism for cell and developmental biology research. As a novel and potentially controversial finding, the huge amounts of supporting data were [deposited](#) in our [GigaDB database](#) to assist others to follow on and reproduce the results.

The judges and editors of *Genome Biology* were impressed by the numerous extra steps taken by the authors in optimizing the openness and easy accessibility of this data, and were keen to emphasize that the value of open data for such breakthrough science lies not only in providing a resource, but also in conferring transparency to unexpected conclusions that others will naturally wish to challenge. In addition to depositing the raw sequencing and transcriptomic data in the [GEO](#) and [SRA](#) databases, the authors went beyond standard research practices by making publicly available all additional supporting data in as usable a form as possible and under a [CC0](#) public domain waiver. The authors worked closely with the curators of our [GigaDB](#) repository to host associated data types that do not have well-established repositories. We also worked closely with the [ISA-team](#) to present this work in the interoperable ISA-Tab format to maximize its reusability (for more insight see our more detailed blog posting see the [ISA-commons paper](#) we contributed to).

[GigaDB](#) is our next generation data repository set up to host data and tools associated with our articles, and also provides our host institute [BGI](#), the world's largest genomics organization, a way to release their data as rapidly as possible and in a citable format. Partly supported by funding from [China National Genebank](#) (CNGB), [GigaDB recently relaunched](#) with a new look, and helps meet our, CNGB and BGI's aims to provide better support for information sharing and exchange of scientific research. Through an association and [associate membership](#) with the

[DataCite](#) consortium, datasets are assigned digital object identifiers to allow them to be independently cited. If you have similar large-scale and heterogeneous datasets contact us about [submitting](#) a [Data Note](#) to the journal. All submissions received before the end of the year are free, and all data hosting and curation in gigaDB is currently included. While there are growing number of data publishers launching and following our lead, our unique selling point is integrated data hosting, important to take into account in this "big-data" era, when cloud computing providers and data hosting repositories can charge up to \$10,000 a terabyte.

The benefits to an author of making their data open

We've previously published author Q&As in this blog (see here for [Xin Zhou on the "squishome"](#), and [Keith Bradnam on Assemblathon2](#)), so to further highlight the advantages making data available in this way, we asked lead author of the study Dr Fei Gao a few questions on how open data has helped his research and career.

How does it feel to get the BMC Open Data award?

As a young researcher, I certainly feel excited about such positive feedback about my work from the scientific community, as this award was evaluated by a committee of scientists. This excitement will certainly inspire or push me to go further in my research field.

Has there been any positive feedback or developments since the paper has been published?

Yes, indeed. For instance, the chief editor of *BioEssays* invited me and my coauthor to write a manuscript for this journal right after he read our paper, and the staff from [WormBase](#) also contacted me, saying that WormBase would like to curate our paper and data in their database. Many scientists in the community have also written emails to us, showing their interests and willingness to collaborate in the future.

Do you think how you released all of the data has made your work more visible and used?

I think that's positive, at least in the field of nematode research, as [WormBase](#) has collected our paper and data, which is the most used nematode information resource in the world. It has also attracted interests in the epigenetics field, since as *Trichinella* is the first nematode found to have a DNA methylation system it could potentially be used as a new model organism.

As it is a controversial and new finding, do you think releasing all of the data associated with the paper has made people more confident about it?

I guess people will tend to be more confident if they really see the data. If they aren't, they can then test the data by themselves, I guess that's also one merit of sharing data with the scientific research community.

Further Reading

1. Fei Gao, Xiaolei Liu, Xiuping Wu, Xuelin Wang, Desheng Gong, Hanlin Lu, Yudong Xia, Yanxia Song, Junwen Wang, Jing Du, Siyang Liu, Xu Han, Yizhi Tang, Huanming Yang, Qi Jin, Xiuqing Zhang and Mingyuan Liu (2012) Differential DNA methylation in discrete developmental stages

of the parasitic nematode *Trichinella spiralis*. *Genome Biology*, **13**:R100 doi:[10.1186/gb-2012-13-10-r100](https://doi.org/10.1186/gb-2012-13-10-r100)

2. Gao, F; Wang, J; Ji, G (2012): Bisulfite-PCR combined with cloning Sanger sequencing data for validating DNA methylation level in *Trichinella spiralis*. GigaScience. <http://dx.doi.org/10.5524/100043>

The post [Open Data For The Win!](#) appeared first on [GigaBlog](#).