

GigaScience launches: overseeing the transition from papers to “executable research objects”

Scott Edmunds 

Published July 12, 2012

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/9p2mj-m6k12>

Keywords

Open Access, GigaDB, GigaScience, Launch

Copyright

Copyright © Scott Edmunds 2012. 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.

Research papers have been the predominant form of scholarly communication for the past few centuries, and despite moves towards online publication and open access, the process and structure of publication has not fundamentally changed in that time. With biological and biomedical research becoming increasingly data-driven, and the amount of information, computational tools, and code supporting a publication in areas such as [genomics](#) and [imaging](#) growing at exponential rates, the lack of access to the resources that the paper is built upon is leading to a growing "reproducibility gap". [Recent scandals](#) relating to falsified data that went long undetected (including [this](#) particularly egregious recent example of an author fabricating the data supporting at least 172 of their publications) further highlight the need to make data easily accessible for purposes of validation and to maintain public trust in science.

As research has shifted to work within, and handle, this data-rich environment and to utilize advances such as cloud computing and automated workflow systems, publishing needs to be able to follow in a similar direction. For a number of years [much talk](#) has been made about the potential for executable papers: aiding review and re-use of data by having all of the tools and data associated with a publication accessible in a reproducible and standardized environment. An important first step in the path has been reached today with the first articles published in [GigaScience](#).

Aiming to become a home to research from the growing number of biological and biomedical fields handling "big-data", [GigaScience](#) is a new type of open access, open data journal that provides standard scientific publishing linked directly to a database that hosts its relevant data. Our associated [GigaDB](#) database, [launched](#) last year, provides a home to all of the supporting data and tools associated with research, thus overcoming one of the biggest challenges holding back reproducible research. Through [GigaDB](#), we assign [DataCite DOIs](#) to these accompanying datasets to provide additional credit to the authors who make their data publicly available and to boost data discoverability and tracking of data. Data citation is important in incentivizing the effort needed to present data to the public in a usable form, and recent successes by others and us in promoting its use were highlighted in [this](#) recent commentary. Our commitment to data DOIs also means that we are well placed to take advantage of the upcoming data citation index [announced](#) recently by Thomson Reuters.

Using the data hosting capabilities and expertise in data handling and [cloud computing](#) of our partner, [BGI](#), we are able to host a much larger and broader range of datasets than journal supplementary files are usually able to handle, outside the capacities of most other journals and repositories. We have been testing and building up our new informatics platform by releasing a number of datasets from BGI to demonstrate new mechanisms of pre-publication data-release. For example, the deadly 2011 outbreak *E. coli* dataset was the [first we released](#), and this led to the crowdsourcing of its analysis, which was [cited](#) in the recent Royal Society "Science as an Open Enterprise" report as an example of "The power of intelligently open data".

Now hosting datasets up to 14TB in size (such as [this](#) enormous resource of 88 tumor-normal paired genomes) and containing several datasets from articles currently in press (including [this](#) cancer single-cell genome data that will be published shortly in [GigaScience](#)), we are providing

examples of how data and papers can be combined in our launch issue today. Exemplifying *GigaScience* and *GigaDB*'s innovative approach is a research [article](#) from Stephan Beck's [group](#) at UCL focusing on ways to conduct whole-genome analyses of DNA methylation. In addition to having the raw data available in [NCBI](#), all of the supporting data and software tools needed to recreate the experiments — a total of 84 GB — are freely available for [download](#) and reuse under the most open [CC0](#) public domain waiver from *GigaDB*. We are also maximizing data-interoparability and re-use by encouraging and rewarding authors for providing much richer metadata. For this dataset, metadata are available in ISA-tab format – the first time a journal has taken ISA-compliant submissions (for more, see the [recent](#) ISA-commons community paper in *Nature Genetics* we contributed to).

The [GigaDB](#) website is continuing to evolve and the next version will be released in a few months with a more extensive search interface. As the final step in attaining fully executable and reproducible papers, we will be working with authors to make the computational tools and data processing pipelines described in their papers available and, where possible, executable on an informatics platform we are developing with [collaborators](#) at the Chinese University of Hong Kong. We hope that by making both the data and processes involved in their analysis freely accessible, this novel form of publication will help articles published in our journal to have a much higher impact in the scientific literature and to maximize their reuse within the community. Check out our two editorials for more on our goals with the [database](#) and [journal](#).

As well as this innovative, big-data-driven publication format, the journal also provides reviews and commentaries that address the many hurdles that still need to be surmounted to improve future big-data handling. Many of these are part of our first [thematic series \(GSC and beyond\)](#) covering the best practices in genomics research, published [in concert](#) with the [Genomic Standards Consortium](#). In addition to commentaries discussing data-sharing issues in [neuroimaging](#) and [genomics](#), our launch issue has a more detailed [review](#) tackling data-compression from Guy Cochrane and Ewan Birney at the EBI, a white-paper on DNA collection for large vertebrate genome projects, and two papers on using genomics data to characterize ecosystem and disease outbreaks. We also have a software/technical note paper detailing a novel data format that facilitates the interoperability of bioinformatics tools and an associated commentary from Jonathan Eisen on 'badomics' terminology and the explosion of "omes" (good and [bad](#)), a topic that readers of his [blog](#) will well know.

We hope you enjoy our first set of articles, and please follow our progress here in our blog, our various social media outlets, and on the journal [homepage](#). We would like to thank the BGI and our collaborators at the Chinese University of Hong Kong, British Library, DataCite, and ISA-Tab for their help and support building the data platform. We'd also like to thank our authors, reviewers and fantastic Editorial Board, as well as [BioMed Central](#) for their part in getting this issue out. BGI is generously covering the open access [article processing charges](#) for the journal's first year, so please contact us at editorial@gigasciencejournal.com if you have related work you would like to submit to this series or journal; alternatively submit a manuscript [here](#). This week we will be at the [ISMB](#) meeting in Long Beach, so come and get hold of us at the BMC booth (#36) if you'd like to meet the editors.

GigaBlog

The post [GigaScience launches: overseeing the transition from papers to "executable research objects"](#) appeared first on [GigaBlog](#).