Northwestern University Feinberg School of Medicine

Galter Health Sciences Library & Learning Center

New Journal Links Datasets and Code to Articles

On July 12, a new peer-reviewed open access journal, titled *GigaScience*, published its first set of articles. *GigaScience* represents a novel format for science journals: manuscripts in *GigaScience* are published with links to software tools used for analysis and to corresponding datasets in the journal's integrated database GigaDB.

The journal is a collaboration between BGI Shenzhen—the world's largest genomics center—and the open access publisher, BioMed Central. The first issue of *GigaScience* includes a manuscript demonstrating the novel format of the journal. Stephan Beck's group at the University College of London published an article on methods for whole genome analysis of DNA methylation. This article includes a section marked "Availability of supporting data". This section links to an entry in the article's references which displays a DOI link to the supporting data in GigaDB. The supporting data for this particular manuscript is 84 GB in size.

GigaDB will accept datasets up to 14 TB in size, so it is a valuable resource for authors who want to share their data or those who are required to do so by federal mandates such as the NIH Data Sharing Policy or the NSF Data Management Plan Requirements. The combination of manuscript, supporting data and links to software or code to analyze the data will improve reproducibility of computational results in experiments involving large datasets, and will also improve the likelihood of new discoveries from repurposed/shared data. The journal also falls in line with the White House Office of Science and Technology Policy goals on access to "Big Data".

The first issue of *GigaScience* also includes an editorial on GigaDB, a few commentaries, and an interesting review on the future of DNA sequence archiving, which had already been accessed over 2300 times within two weeks of its publication. This journal provides a valuable service, since many researchers may not have other means for online deposit of large datasets such as BAM files, FASTQ files and other large files produced by their experiments. Many other journals do provide access to supporting documents or supplemental files, but do not provide a means for storage of files of such a large size. It is my hope that the journal proves successful and sustainable, and that other journals consider duplicating this service.

For more information on *GigaScience*, see:

- The BioMed Central blog post by the editors of *GigaScience*
- The *GigaScience* journal blog, GigaBlog
- Follow GigaScience on Twitter: @GigaScience

Pamela Shaw
Biosciences & Bioinformatics Librarian
email Pamela

The Biosciences & Bioinformatics Blog highlights new tools and news items of interest to the biosciences research community at Northwestern University.