

GigaScience Journal Part of Global Data-Sharing Effort: New Standards Allow Disparate Data Sets to Integrate

Scott Edmunds 

Published January 28, 2012

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/m4prx-0vz10>

Keywords

Open Access, GigaDB, Interoperability, ISA-tab, Metadata
Feature Image

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.

ISA-commons logoLead by researchers at the [University of Oxford](#), a group of more than 30 scientific organizations around the globe, have worked to produce a common standard that will make possible the consistent description of enormous and radically different databases compiled in fields ranging from genetics to stem cell science, to environmental studies. One of the contributors playing a role in the project is [GigaScience](#), as we feel it potentially very useful to aid in the handling of the wide-variety of data-types covered by our scope our [scope](#).

The new standard provides a way for scientists in widely disparate fields to co-ordinate each other's findings by allowing behind-the-scenes combination of the mountains of data produced by modern, technology driven science.

This standard-compliant data sharing effort and the establishment of its on-line presence, the ISA Commons – www.isacommons.org, is described in a Commentary (and highlighted in the [editorial](#)) published [today](#) in the journal *Nature Genetics*.

"We are now working together to provide the means to manage enormous quantities of otherwise incompatible data, ranging from the biomedical to the environmental," says [Susanna-Assunta Sansone](#), Team Leader of the project at the [Oxford e-Research Centre](#), and founder of the [BioSharing Network](#) (of which BMC and *GigaScience* are both [members](#)).

"An example of how this works at the Harvard Stem Cell Institute is that we can now find a relationship between experiments involving normal blood stem cells in fish and cancers in children", says [Winston Hide](#), Professor of Bioinformatics at the Harvard School of Public Health (for more see this related [publication](#)).

It was necessary to establish common data standards, say the commentary's authors, because of the tsunami of data and technologies washing over the sciences. "There are hundreds of new technologies coming along but also many ways to describe the information produced" said Sansone, noting that "we can take a jigsaw puzzle of different sciences and now fit the many pieces together to form a complete picture".

"One of the things that I find most empowering about this effort is that now small research groups can begin to store laboratory data using this framework, complying to community standards, without their own dedicated bioinformatic support. It is a bit like facebook allowing everyone to create their own website pages – suddenly you don't need to be an expert in computing to get your data out to the rest of the world", says Dr. [Jules Griffin](#), of the University of Cambridge.

"What we like about it is its unifying nature across different bioscience fields and institutions", says [Dr. Christoph Steinbeck](#), The European Bioinformatics Institute.

And "it also has the potential to work for large centers too", says Scott Edmunds, of the [BGI](#) and [GigaScience](#). As *GigaScience* aims to take as many types of "large-data" as possible, the need to handle as many formats as possible was essential, and the large number of data-types supported by ISA-commons and ability to create new configurations potentially addresses this very important issue. This has lead to [GigaScience](#) being the first journal to offer authors the

GigaBlog

option to submit data in ISA-commons format, and these resources have also been made available to the BGI (the worlds [largest](#) Genomics institute) to release their enormous quantities of data quicker the wider research community through the associated [GigaDB](#) database.

For more on the aims and goals of [GigaScience](#) please see [this](#) previous BMC Blog posting, and for news and updates follow [GigaBlog](#) and the [@GigaScience](#) twitter feed. The journal is now taking [submissions](#) for "big-data" associated research, tools and software for handling large-scale data, and reviews and commentary on issues dealing with data-handling and standards.

References:

1. ISA Commons: [isacommons.org](#)
2. It's not about the data. **Nature Genetics** **44**, 2 (2012).
3. Sansone, S-A. *etal.* Toward interoperable bioscience data. **Nature Genetics** **44**, 2 (2012).
4. Ho Sui SJ *et al.* The Stem Cell Discovery Engine: an integrated repository and analysis system for cancer stem cell comparisons. **Nucleic Acids Res**. 1;40(D1):D984-D991.(2012).

The post [GigaScience Journal Part of Global Data-Sharing Effort: New Standards Allow Disparate Data Sets to Integrate](#) appeared first on [GigaBlog](#).