

Ain't No Party like a Bring Your Own Data Party!

Scott Edmunds 

Published August 22, 2014

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/55qav-7ty83>

Keywords

Technology, Bioinformatics, Conferences, Data Publishing, GigaDB
Feature Image

Copyright

Copyright © Scott Edmunds 2014. 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.

Data Club is Gonna Show You How

As science is supposed to be about "standing on the shoulders of giants", we all know sharing scientific data should be a good thing, but there are obviously large technical and cultural challenges holding things back. Things are a long way from the [Jimmy Wales](#) "Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge" utopian dream, but some research fields (e.g. genomics) have done a better job making data available than others. Unfortunately sharing complicated scientific data usually isn't as easy as just dumping it in a dropbox folder, and to be reused scientific data needs to be properly structured, curated and described. This process of providing sufficient metadata and instructions for reuse can take considerable a lot of time, effort and expense, so sufficient incentives are needed to make this effort.

While a few specific fields and journals (including ourselves and [now PLOS](#)) have strict policies and mandates enforcing data deposition, more carrots are also required. While we written about positive incentives and crediting good practice through prizes like the [BMC Open Data Award](#) (of which some of our datasets have now won [two years running](#)), there is scope for further schemes to promote the liberation of the huge amounts of very useful research datasets out there. Data publication (such as the [Data Note](#) articles we've been publishing) is supposed to be one mechanism of incentivizing these efforts, but it can still be difficult for authors to organize their data, as well as know exactly what information is required with this still new and unconventional article type. There may be a need to for data producers to meet data curators and standards experts and go through in person some of their example datasets to lower the barriers of entry and make this process more understandable.

Don't Stop Curatin'

To try to address this issue, this June, supported by our [BBSRC UK-China partnering award](#), we and the [ISA Team](#) at the Oxford e-Research Centre organized our first "data hackathon" at our [newly refurbished BGI Hong-Kong](#) offices. Metabolomics seemed an ideal area to support, as the field is starting to produce larger and larger scale datasets, and there are now several public repositories linked through the new [MetaboleXchange portal](#) to take this data, but they are all in need of more data and users to prove their utility. The largest of these (with [49 public datasets currently available](#)) is the EBI Metabolights database. Participants at our first data get together were a number of young scientists and omics data producers from local universities (including Hong Kong Baptist University and BGI), as well as some of the UK metabolomics standards community including the EBI and [Birmingham Metabolomics Centre](#).

Timed just before the [Metabolomics Society meeting](#) that we subsequently attended in Tsuroaka in Japan, the goal was to establish common standards and curation practices for omics data as well implement new [ISA software](#) functionalities to facilitate deposition to the [EBI Metabolights repository](#) and support feature requests from journals using ISA formats, such as ourselves. A further important part was the 'bring your own data' track, allowing data producers to interact with the more curatorial groups and learn how to best report and structure their work for publication and deposition. On top of curators, editors were also on

hand to provide feedback on the writing up as Data Note articles, one of the main incentives for early data release.

Hackathons are always intense but productive affairs, and the fruitful interactions over the duration of the meeting resulted in the delivery of a new ISA-Tab viewing component for web browsers (see [this DOI](#) for the code and [this sourceforge page](#) for examples), and the conversion of Metabolights ISA-Tab content to RDF and 5 experiments, accounting for nearly 750 samples worth of data being generated. On top of boosting the data in Metabolights and [GigaDB](#), some of these outputs are currently being written up as Data Note articles and peer reviewed by us (see [guidelines](#)), so watch this space for the results. The work also led to sharing and refining curation guidelines, and we presented many of the outputs the following week at a workshop in Tsuruoka (see Rob and Scott's [slides here](#), and the Philippe from the ISA-teams slides [here](#)). Finally, efforts to deliver an API supporting the programmatic creation of ISA-Tab documents are well under way.

Say Goodbye

On top of the busy and productive work schedule there was also a useful social and networking side to the occasion, as a further incentive for participation was the chance to mix, learn from and interact with a diverse group of researcher and data experts, as well as experience some of the amazing culinary and touristic sights of Hong Kong (see picture).

After completing this hackathon, we discovered that in the same week the Dutch node of ELIXIR and the [Dutch Techcentre for Life Sciences \(DTL\)](#) organized and hosted an almost identical event. Termed by them a "[Bring Your Own Data](#)" (BYOD) party, this first event brought together data producers with experts in semantic web technologies to make their data available in a Findable, Accessible, Interoperable and Reusable, or "[FAIR](#)", manner. They found their experiences equally productive and fruitful, and ELIXIR-NL and the DTL are keen to promote future events. We wholeheartedly endorse this as well, and for our future events will likely co-opt their BYOD title. Owing to the success of these first hackathons and BYOD events, the same teams will be organizing follow-up meetings. We will keep you posted and, if you are interested in joining, do get in touch with us and the ISA group (isatools@googlegroups.com).

A shorter version of this article was highlighted in the [August MetaboNews newsletter](#).

The post [Ain't No Party like a Bring Your Own Data Party!](#) appeared first on [GigaBlog](#).