

More on our ISMB workshop: What Bioinformaticians need to know about digital publishing beyond the PDF

Scott Edmunds 

Published August 9, 2013

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/4sc41-9fz56>

Keywords

Open Access, Publishing, Berlin, Conferences, ECCB
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.

Big Data Publishing (credit [Jenny Cham](#), CC-BY)

As mentioned in our [previous posting](#), on top of the many great talks and sessions we attended at [ISMB](#) in Berlin last month, we were kept even busy helping to organize and present in a special [Beyond-the-PDF](#) inspired "[What Bioinformaticians need to know about digital publishing beyond the PDF](#)" workshop. Most of the heavy lifting has to be credited to the hard work of Marco Roos, but the other organisers included Oscar Corcho, Carole Goble, Barend Mons, Jun Zhao and Erik Schultes, and we had a ridiculously overqualified list of speakers, panelists and supporters that help make it a success.

You can see the blurb and line-up on the [ISMB website](#), but the main aim of the session was to inform participants of changes and new opportunities in scientific communication. While there have been a lot of recent developments spurred by related future of scholarly communication conferences and projects (see [Force11](#)), we wanted to take these sorts of discussions out of the usual publishing crowd, and present them to researchers who may potential use them. As mentioned in our [previous posting](#), ISMB and the computational biology community are a particularly receptive and positive audience for open science as the whole field has been built on data sharing and open-source tools, so we wanted to provide visual guidelines and present tools that take this open approach even further.

Phil Bourne brilliantly set the scene, giving an overview and his thoughts on what the issues are, and how people can participate to make a change (slides [here](#)). Giving some sense of the urgency that researchers need to address the issue of making their data available, as few are prepared for the implications of [new NIH open access policy](#) that starts being enforced next month (if their work doesn't end up in PMC, grants will not get renewed), and similar mandates for data will be here sooner than most are prepared for too. Second on the program was Rebecca Lawrence from [F1000](#), talking about data publishing and data peer-review initiatives. Showing some of the great examples [F1000Research](#) have done in this area, their post publication peer-review pipeline has managed to review papers and take them from submission to publication in as little as 34 hours. If you have seen our [recent postings](#) on the unusual peer-review of our [Assemblathon2 paper](#), like us, they have also had very positive experiences with open peer review. Also touching on data review, Rebecca has been very closely involved with the [JISC PREPARDE](#) (Peer REview for Publication & Accreditation of Research data in the Earth sciences) and some of this work was presented.

GigaScience meets ISA, RO and Nanopublications

Scott's [talk](#) on "Big Data Publishing" in the session covered work on deconstructing the paper to reward reproducibility, deposition and transparency of data, methods and analyses. Using our [SOAPdenovo2 paper](#) as an example to we show how we can issue separate DOIs resolving to all of the associated [data](#) as well as the [scripts, pipelines and workflows](#) we are hosting.

Scott Edmunds ISMB talk on Big Data Publishing from [GigaScience](#), [BGI Hong Kong](#)

We have already presented in the past on our [GigaDB](#) database and [GigaGalaxy](#) data platform, but tied with this workshop and related presentations at the meeting this was the first time we

presenting some initial findings from a case study we are carrying out with the [ISA-TAB](#), [Research Object \(RO\)](#) and [Nanopublication](#) communities to see how these three data models can support representation of scholarly artifacts. Studying how complimentary these models can be, how much value we can add to the publishing process, as well as encourage their use by demoing them as "digital instruction for authors", this work is very much still in progress, but we were keen to show the results so far to the workshop audience. We have been working with the ISA team for a while, using their interoperable metadata format in a number of our datasets (see the [methyated nematode genome](#) – of which there will be an announcement next month), and as we have been working with [Galaxy workflows](#) the workflow-centric focus of the [Research Object](#) model made it an obvious system to trial in this case study too. [Nanopublications](#) are a new area to explore for us, but following our "deconstructed paper" approach, being smallest the unit of publishable information (an assertion), the ability to attribute and cite these makes them attractive through their potential to provide incentives for researchers to make their data available.

While the work on the case study so far has shown we can represent the parts of the paper in these models, and that we can recreate results from the paper using our [GigaGalaxy](#) workflow, we are still struggling to recreate other parts of the results and represent them as nanopublications. It was great seeing our board member Carole Goble further elaborate on the costliness of reproducibility in her ISMB keynote the following day (see her [great slides](#)) and include some of this case study as examples to the wider ISMB audience. The RO and Nanopublication work was elaborated upon by Marco Roos in his Tech Track talk (see [abstract](#)), and Mark Thompson in his [poster](#), and for completeness sake we will try to post these when they are made available online.

The session ended with a packed panel chaired by Barend Mons, and including Niklas Blomberg (Director of [Elixir](#)), Carole Goble, Larry Hunter, Winston Hide, as well as the speakers and organizers summarising the topic together and fielding questions from the audience. Polling the audience at the end of the session, Marco got a very positive response when asking whether to do the workshop again [next year in Boston](#), so watch this space for updates on the ongoing case study, and if there will be a follow up workshop. We'd like to thank all of the case study participants, particularly Jun Zhao, Philippe Rocca-Sera and Alejandra Gonzalez-Beltran at Oxford, and Mark Thompson at LUMC, and Marco Roos in particular for doing most of the hard work setting up the workshop, as well as the ISMB organisers for selecting it and helping us make it happen. We'd also like to thank Jennifer Cham at the EBI for the [great sketch](#) featured above, and for making it CC-BY. You can see the other fantastic sketches she made at the conference from the her [Flickr page](#).

The post [More on our ISMB workshop: What Bioinformaticians need to know about digital publishing beyond the PDF](#) appeared first on [GigaBlog](#).