

The Latest Weapon in Publishing Data: the Polar Bear

Scott Edmunds 

Published May 14, 2014

Citation

Edmunds, S. (n.d.). In *GigaBlog*. GigaBlog. <https://doi.org/10.59350/855r6-3zq41>

Keywords

Biology, Publishing, Data Citation, Data Publishing, DOI
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.

[Polarbearonice](#) Being the largest land predator, the fearsome and enigmatic Polar Bear is seen by many as a powerful symbol to highlight of the threats to the environment through global warming. With a [new publication](#) on the Polar Bear genome out last week in *Cell*, they surprisingly are also an impressive example of how far data publication and citation has come in the last few years, and help debunk many of the negative arguments about the early release of datasets in this manner.

Providing a comparison of the genomes of polar bears and brown bears reveals that the polar bear is a much younger species than previously believed, having diverged from brown bears less than 500,000 years ago. This has surprised many, as the unique adaptations polar bears have to the arctic environment must have evolved in this very short amount of time, including not only a change from brown to white fur and development of a sleeker body, but big physiological and metabolic changes to subsist on a blubber-rich diet of marine mammals.

Being an intriguing finding, more confidence can be given to these unexpected results by having all of the supporting data [available](#) for review and replication, and this also maximizes the potential for others to build upon these finding and advance science. This study is a good example of that, especially as the data has been publicly available for nearly three years before publication. While genomics leads many areas of biology through having mandated policies for data producers to release their raw genome sequence and assemblies in the [INSDC](#) databases, in practice this is only enforced upon publication, and only by the more thorough journals. There is little incentive for researchers to release unpublished data earlier than this, as well as make available intermediate and processed data, as well as computer scripts, workflows and detailed protocols that would allow others to properly reproduce complicated scientific work. This may be problematic enough, without there being perceived disincentives for early release of scientific information. The current "publish or perish" scientific culture understandably makes many very protective of their data, with much paranoia and fear of being scooped. These fears also not helped by some of the more archaic policies of the more conservative publishers that discourage any communication about a project before its eventual publication (see the [Ingelfinger rule](#)).

Data Publication to the Rescue

The concept of Data Publication aims to modify these cultural practices and incentive systems to more strongly credit data production and release. If people believe data generated in the course of research are as valuable to academic discourse as papers, then it should be treated in the same manner and cited in the references (see the [DCC guidelines](#) for more). Regular readers of this blog will well know [our efforts](#) carrying this out over the last few years with [DataCite](#) and the [British Library](#), and the number of data publishers and data platforms that have joined us in these efforts are growing all the time. Releasing datasets under [CC0 public domain waivers](#), but making them citable with [DOIs](#), this gives others mechanisms to credit the data producers.

Despite this move, there are still fears from many that releasing datasets without restrictions could lead to others scooping the subsequent analysis papers, or lead some of the more traditional publishers to invoke "[ingelfinger](#)" like policies that would see the publication of a

dataset with a DOI as 'prior publication' (in a similar manner to [pre-print servers](#)) that would preclude subsequent publications. As an experiment in openness, with the launch of the *GigaScience* database (now called [GigaDB](#)) in July 2011 we released a number of unpublished genomic datasets from our hosts at [BGI](#), the largest genomics organization in the world (see the announcement [here](#)). This included the genomes of species of [Macaque's](#), [Penguins](#), [the Pigeon](#) and [Polar Bear](#). Subsequently most now have had genome papers published without difficulties in journals such as [Nature Biotechnology](#) and [Science](#), but until recently the Polar Bear and Penguin genomes had still not been formally published. In this time the Polar Bear in particular has been provided an excellent example of data-reuse, as using the DOIs to [track subsequent citations](#) at least five other groups have published important comparative and population genomics studies using this data.

This included the following studies:

Hailer, F et al., Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science*. 2012 Apr 20;336(6079):344-7. [doi:10.1126/science.1216424](#).

Cahill, JA et al., Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *PLoS Genet*. 2013;9(3):e1003345. [doi:10.1371/journal.pgen.1003345](#).

Morgan, CC et al., Heterogeneous models place the root of the placental mammal phylogeny. *Mol Biol Evol*. 2013 Sep;30(9):2145-56. [doi:10.1093/molbev/mst117](#).

Cronin, MA et al., Molecular Phylogeny and SNP Variation of Polar Bears (*Ursus maritimus*), Brown Bears (*U. arctos*), and Black Bears (*U. americanus*) Derived from Genome Sequences. *J Hered*. 2014; 105(3):312-23. [doi:10.1093/jhered/est133](#).

Bidon, T et al., Brown and Polar Bear Y Chromosomes Reveal Extensive Male-Biased Gene Flow within Brother Lineages. *Mol Biol Evol*. 2014 Apr 4. [doi:10.1093/molbev/msu109](#)

Data Sharing, with Teeth

The [publication](#) of the first assembled genome of the polar bear featured [on the cover of the 8th May issue](#) of *Cell* provides a very positive example that the early release of data can assist others to produce and publish useful work, while at the same time not lead to scooping or raise any issues for a journal as prestigious as *Cell*. The genomics community is different to many others in that they have the [Fort Lauderdale rules](#) that asks for data producers to be given first right to publish full scale analyses, although (as with the practice of citation) this is more through etiquette rather than any legally binding means. It is reassuring to see that in this example these practices are still holding, and improper uncredited use of others data is still seen as bad practice and scientific misconduct.

The fact that this paper was published in *Cell* makes an even better example of how far data publishing has come, as Cell Press in 2011 were the only major biology publisher to [state in a survey](#) carried out by our fellow data publishers *F1000Research* that they would see the publication of data with a DOI as potential prior publication. With the example of the Polar Bear showing that this is no longer true, there is now one less reason for researchers not to release

GigaBlog

their data. Yesterday we released our 100th dataset in GigaDB (the as yet unpublished [genome of the Red Throated Loon](#)), and are currently doing a big push on our [Data Note](#) articles. We will have some important announcements and examples coming out in the next few months, so watch this space for more. [Submissions](#) (included unlimited data hosting in our [GigaDB](#) server) are currently free until the end of the year, so talk to us if you have interesting datasets you are interested in publishing.

References

1. Li, B; Zhang, G; Willersleve, E; Wang, J; Wang, J (2011): Genomic data from the polar bear (*Ursus maritimus*). GigaScience. <http://dx.doi.org/10.5524/100008>
2. Liu, S et al., Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell*. 2014; 157(4): 785-794

The post [The Latest Weapon in Publishing Data: the Polar Bear](#) appeared first on [GigaBlog](#).