PUBLISHED: 3 MARCH 2015 | ARTICLE NUMBER: 15032 | DOI: 10.1038/NPLANTS.2015.32

editorial

Making the greatest impact

Plant research projects are increasingly producing large systematic collections of phenotype data. But how can it be stored so that others can easily use it and that proper credit goes to the creators of the data?

It is a sad reflection of the way that scientific work is judged that the most frequently asked question about *Nature Plants* is, what is its impact factor? For the record, as a newly launched journal we don't yet have an impact factor. Indeed Thomson ISI who calculate the journal impact factors will not be assigning us a number until 2017 at the earliest as there will not be sufficient data until then.

What underlies the question though is the fact that for many scientists their research funding and indeed their careers depends on publishing in journals with high impact factors despite the impact factor of a journal saying very little about the quality of any individual piece of work it publishes. This is a problem for all scientists but it is particularly acute for those working on projects that are not designed to produce neat, 'paper-sized' quanta of advance but instead are working towards broad-ranging practical applications. Or worse yet, the vast quantities of data that modern genotypic or phenotypic screens can produce.

Such projects aimed at producing direct practical applications are exactly what the Bill and Melinda Gates Foundation is funding (see News Feature by N. Gilbert). The foundation's deep pockets allow it to take a long view, funding projects for five or more years in order that bold experiments can be undertaken. Whether creating hybrid crops that maintain the benefits of heterosis into subsequent generations by reproducing clonally; re-engineering photosynthesis to increase its efficiency; or transplanting the ability to fix nitrogen into non-leguminous crops, the Gates Foundation will measure the return on its investment in terms of the potential improvements to the health and economic well-being of people in the developing world.

Elsewhere, in this issue Graham Moore of the John Innes Centre in Norwich, UK discusses his experiences with another initiative focussed on practical outcomes, the WISP collaboration (see Comment by G. Moore). WISP brings together academic and industrial scientists to identify and develop traits in wheat that will help maintain the increase in yields needed to satisfy our increasing global demand for the cereal. Genotyping and phenotyping of germplasm from landraces, wheat synthetics and wild relatives is providing a wealth of important data, but publications in journals are not the primary outcomes.

One of the results of WISP has been a genotyping data set for wheat representing more than 400 million data points leading to almost one million validated and characterized SNP-based markers. The WISP collaboration is large enough to be able to host this data and make it accessible to researchers, however smaller projects do not have the same luxury. Instead, there is an urgent need for central repositories for plant genotype and phenotype data, both to ensure their proper curation and also to help provide due credit to the researchers that have created them.

There is an urgent need for central repositories for plant genotype and phenotype data, both to ensure their proper curation and also to help provide due credit to the researchers that have created them.

Some initiatives already exist. One such is DivSeek, which was recently launched to encourage plant researchers, breeders and gene banks to share and use effectivly crop diversity information (http://www.divseek.org/white-paper). Authors publishing their genetic and genomic studies in *Nature Genetics* will be strongly encouraged to deposit their genomic and related phenotypic data with DivSeek to help other researchers to build on their work (see *Nature Genetics* **47**, 99; 2015).

DivSeek is not the only repository for this, or related, data. For example, the Jackson Lab has for many years maintained a QTL archive (http://qtlarchive.org). This was not designed with plant phenotype data in mind but there are plant data sets held there. Also Phenome Networks has a project called Unity (http://phenome-networks.com/unity) aiming at being a repository for plant phenotyping data for a variety of species. Concentrating a little more on the collection and analysis of phenotypic results, the Phytomorph project based at the University of Wisconsin (http://phytomorph.wisc.edu/index.php) is attempting to develop standards for both the documenting and analysis of systematic phenotyping experiments.

There is also the National Science Foundation funded iPlant Collaborative (http://www.iplantcollaborative.org) whose aim is to provide life science research with computational capacity for 'big data'. As the name suggests this initiative is particularly successful at handling plant data. It has also developed a suite of analysis tools to help interrogate the data that it holds. iPlant has recently announced that in collaboration with the Genome Analysis Centre in Norwich, UK it is creating a European node called iPlant UK, funded to the tune of £1.8 million by the Biotechnology and Biological Sciences Research Council.

As a way to achieve proper documentation of data resources, to make them easily citable in conventional research articles and to provide the authors of such data sets a degree of recognition for their efforts Nature Publishing Group last year began publishing a new kind of scientific article, the Data Descriptor, in our journal *Scientific Data* (http://www.nature.com/sdata). Data Descriptors create a degree of harmonization for data sets wherever they are held. They explain where the data have been deposited and how they were collected, together with the standards, definitions and ontologies employed.

The most easily quantified products of scientific research may be papers in journals, but the increases in understanding that they represent and the practical applications to which that knowledge can be put are the true goals. *Nature Plants* encourages any and all creators of large data resources to share those data through an appropriate repository and to consider publishing a Data Descriptor to help others build on their results. ❐