

Cite site

We now have adapted our preprint archive, *Nature Precedings*, to host project descriptions, community standards papers and funder policies. Citable project descriptions provide a guide to the resources available and create a mechanism to give data producers citation credit.

The US National Institutes of Health and National Science Foundation both now require their grantees to produce a project description document explaining their aims and the resources they hope to generate. We created an archive where these documents (often referred to as ‘marker papers’) can be publicly deposited and, if necessary, regularly updated. Project descriptions can function as guides to projects that are complex and that deposit disparate resources in several databases. They can also help data users keep track of differing and evolving conditions for data use (if any exist). Because the archive is citable, it can help data users and journal editors handle competing publications more fairly. Citations to project descriptions may also be tracked to ensure effective use of resources and to make recommendations for best research practice.

Funding body policies and the ideals of the participants in the public Human Genome Project have resulted in genomic sequence data being deposited in the public domain as soon as the data producers could complete their declared basic quality control measures. Although this early data release has generated both research results and goodwill, the advance publicity the sequences generated for the data producers’ own papers has been hard to quantify because no citation convention was established between the databases and the journals. Consequently, researchers have been somewhat reluctant to release other kinds of data prior to publication of the research articles that give them citation credit. In effect, to release data early in the absence of quantitative citation is to short-circuit the economy of knowledge production.

We have previously discussed the argument in favor of quantitative citation of data accessions in an Editorial (“Data producers deserve citation credit,” *Nat. Genet.* 41, 1045 (2009)). We have also frequently made the case against imposing use conditions on data that have been released into the public domain prior to publication. However, the justification that others have put forth for such a practice includes arguments that a stated moratorium on competing publications may protect the interests of resource generators and that, because moratoria encourage

simultaneous publication, even the data users themselves may benefit from this practice, especially where a large proportion of the field is supported by a single funding agency. In contrast, as an international journal, we are keen to encourage all users to avail themselves of public data. Consequently, we support all forms of data release so long as the data are truly available. However, we can no longer argue that use restrictions are intrinsically confusing, as they can be explicitly laid out in the project description according to a funder template. In citing the project description, data users are declaring that they have read the use conditions, if any exist.

We would like to thank the researchers and administrators of the US National Human Genome Research Institute’s Human Microbiome Project (HMP), who have been our partners in the experiment of adapting the preprint archive to the needs of data producers and their funding bodies. We also thank the staff of the National Center for Biotechnology Information’s dbGaP and Sequence Read Archive databases for advice and bidirectional linking. We hope that the simple format we have adopted will be flexible enough to serve all data producers worldwide and that it will permit researchers to update their project descriptions as often as their needs require.

Previously, it was impossible for journals and even peer referees to track data resources and data use, let alone make consistent recommendations on secondary use of data. Researchers who produce and post a citable project description can now expect to be treated respectfully under the accepted academic practice of accurate citation. As a precondition for peer review by the journal, we will now require users of all unpublished deposited data to cite the accession number and database of the data they used, together with the identifier of the project description. We think that *Nature Precedings* DOIs will be suitable for use by a range of journals, and we invite other journals to adopt our policy. In return, we are also happy to adopt solutions offered by other repositories, provided there is a current unique identifier for each project description and a unique accession for the matching dataset. ■