

Respectful re-use

The impact of the papers we publish depends increasingly on the data they describe. In insisting on data access for referees and readers, we prioritize scientific integrity above all and place the interests of research participants before impact.

The Nature research journals insist that materials, methods and data be made available and that authors detail any conditions for access where these exist. In our shared guide to authors (<http://www.nature.com/authors/policies/availability.html>), we state:

The preferred way to share large data sets is via public repositories. Some of these repositories offer authors the option to host data associated with a manuscript confidentially, and provide anonymous access to peer-reviewers before public release.

It is the practice of this journal to check all manuscripts for appropriate data access prior to peer review and, where necessary, to contact the authors to ensure that essential data for transcript expression (microarray and RNA sequencing), exome sequencing (*Nat. Genet.* 43, 921, 2011) and genome sequencing are available to the referees. Deposition of microarray genotyping data sets is strongly encouraged where consent conditions permit. We also check with the authors that an active accession code for the data is available at the time that they receive the proofs of their accepted manuscript.

The basic principles we apply to data access have a hierarchy that is generally used in editorial decision making, so it is good to make it explicit here: integrity over privacy over impact. All three are important considerations that underpin a publication's usefulness and success, but the re-use of data cannot take precedence over the protection of research participants. If privacy issues prevent the reader from accessing the data underlying a paper's conclusions, the authors must validate their results with data that are accessible to the reader.

Funder mandates for data deposition can help to boost the usefulness of data sets and the consequent impact of the papers that describe them. Corporations and privately funded researchers can hope to produce papers of equivalent impact by providing published details of the ways in which associated data sets can be accessed, specifying any special conditions that may apply. We are particularly keen to hear of innovations from biopharma companies for data access as well as public-private

partnerships that can offer secure data access. We also contact authors when readers have difficulty getting data or materials and will, when appropriate, detail difficulties with data access with an Editorial Note on the paper.

Authors should be aware of the differences between the two repositories with access control that we endorse for submission of genotypic data linked to human health-related phenotypes (for example, disease status). The database of Genotypes and Phenotypes (dbGAP; <http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.cgi>) requires authors to work with one of the data access committees established by the US National Institutes of Health (NIH). The European Genome-phenome Archive (EGA; <http://www.ebi.ac.uk/ega/submission>) explains how authors can establish an acceptable access committee for their own data. Both repositories require a declaration that the data deposition is in accordance with limits set on its use by research participants in the informed consent they provided and is in accordance with applicable laws and institutional policies. We require the same assurance. For publication, it is the responsibility of the corresponding author(s) to ensure that all data submitted to a repository are covered by the appropriate consent agreement with the research subject as well as institutional and regional regulations. We request a statement regarding consent for data deposition (as well as for the research itself) within the paper so that we can anticipate any exceptional circumstances and help to expedite data deposition.

While we recognize that unrestricted access to data boosts its re-use, the long-term success of research on human health depends crucially upon establishing and maintaining trust between researchers and research subjects that will encourage greater participation in research, future access and the progressive refinement of research conclusions that only follow-up allows. Central to this trust is use of the data in ways that fulfill the expectations of the research subjects when they consent to provide their materials and health data. In accordance with this view, sequence and genotype data sets with associated health-related phenotypes need to be submitted to accountable databases with controlled access. ■