

Why data-sharing policies matter

Alan E. Guttmacher^{a,1}, Elizabeth G. Nabel^b, and Francis S. Collins^c

^aNational Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; ^bNational Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892; and ^cNational Institutes of Health, Bethesda, MD 20892

Data from biomedical research are more broadly available to the research community today than in the past. Technical developments, such as web-based databases, have played a role in this transition, but so has a fundamental shift in the view of who “owns” research data. The model of the investigator owning data has been increasingly replaced by one in which society owns data. Scientific and cultural forces have converged in the past decade to foster this new model. Numerous examples of broad data sharing, ranging from the Human Genome Project, to the Framingham Heart Study, to the myriad genome-wide association studies deposited in the dbGaP database of the National Institutes of Health (NIH) (see www.ncbi.nlm.nih.gov/sites/entrez?Db=gap), offer compelling testimony to how broad access accelerates and empowers scientific investigation to benefit society.

However, for both ethical reasons and the purely practical concern of making broad data access workable, it is vital to recognize and protect both participants’ and investigators’ interests (1). Participant protections include a number of measures, including appropriate consent processes, Institutional Review Board review, technical and statistical database safeguards, and requiring researchers who access data to agree not to attempt to identify participants whose data are included.

The interests of the investigator who places data in an accessible database also require protection. The major available protection is the guarantee of a period of exclusivity in submission of abstracts and publications for a number of months (usually 6 to 12). This exclusive period is assured by allowing data access only to end users who agree to abide by it. The investigator also frequently profits both from the value added to the data in its deposition in a community database (for instance, the genotypic data added in dbGaP) and from the collaborators that the wider data availability attracts. However, the period of exclusivity is the key benefit to the investigator; breaching this guarantee threatens the very existence of broad access to data in biomedical research.

With these principles in mind and after considerable public input, the NIH implemented a “Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)” (see <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html>). This policy provides guidance for researchers who are interested in accessing data from the National Center for Biotechnology Information (NCBI) database dbGaP, requiring recipient investigators and their institutional officials to sign an agreement (the Data Use Certification) by which they will comply with the terms of data ac-

cess, including a 12-month period of exclusivity.

A recent breach by a recipient investigator of the Data Use Certification led to the on-line publication by PNAS of a manuscript that should never even have been submitted (2). While both PNAS and the NIH will deal with this specific breach, it is the wider research community that must police itself and prevent inappropriate publication in the future. This will require that recipient users of community data resources be fully aware of data use limitations to which they agree and be scrupulous in honoring them. It will require that reviewers question whether data access terms have been followed in submitted manuscripts. It will require that publishers ensure that authors observe the same level of ethical behavior for data access as for conflict of interest or research misconduct. It will require that the NIH design effective strategies for alerting the research community to this issue and implement steps that make breaches difficult to commit and easy to discover.

Wide access to data benefits the research community and society. We must all play an active role in protecting the rights of both research participants and principal investigators if this important practice is to flourish.

¹To whom correspondence should be addressed: Email: guttmach@mail.nih.gov.

breach of NIH embargo policy on a PNAS paper. *Proc Natl Acad Sci USA*, 10.1073/pnas.0910317106.

1. Toronto International Data Release Workshop Authors (2009) Prepublication data sharing. *Nature*

461:168–170.

2. Schekman R (2009) PNAS takes action regarding