

Bill & Melinda Gates Foundation's Data Access Principles

Frequently Asked Questions

This FAQ document is designed to provide guidelines for response to the anticipated frequently asked questions regarding the launch of the Bill & Melinda Gates Foundation's Data Access Program in 2011. This is considered a pilot year during which foundation staff and grantees should endeavor to apply the principles to grant making and management, seek guidance on the challenges that arise, and document these challenges for future program enhancements. As such, there are no "wrong answers" regarding the specific ways in which the principles will be applied in 2011, but good faith efforts should be made. The goal is to implement the principles broadly and to learn from experiences to refine the process. We appreciate your cooperation and patience.

Principles and rationale

- [What are the Bill & Melinda Gates Foundation's Global Health Data Access Principles?](#)
- [What is the difference between these principles and related elements of the Global Access Policy?](#)
- [Why has the foundation developed these principles?](#)
- [What are the benefits of providing access to data?](#)

Scope of the principles

- [To which activities and types of data do these principles apply?](#)
- [What is meant by final data?](#)
- [For which grants must a grantee prepare a Data Access Plan?](#)
- [Do these principles apply to existing grants?](#)
- [Do these principles apply to qualitative data?](#)

Data Access Plans

- [What should a Data Access Plan include?](#)
- [What is the timing for completing a Data Access Plan?](#)
- [I plan to publish a paper with the findings of this project. Do I need to provide access to the data?](#)
- [I am concerned that others will publish papers based on my data before I publish my paper. How should I proceed?](#)
- [Will the foundation provide funding for data access activities?](#)
- [How quickly does the foundation expect data to be released?](#)
- [What methods are recommended for data access?](#)

Privacy, informed consent, and intellectual property

- [What are the risks of data access?](#)
- [How can the principles be balanced against the need to safeguard privacy rights of individuals?](#)
- [How can the principles be balanced against the need to protect intellectual property?](#)
- [This project involves proprietary data; do the data still need to be shared?](#)

Principles and rationale

What are the Bill & Melinda Gates Foundation's Global Health Data Access Principles? The foundation's Global Health Data Access Principles provide the rationale for and basic tenets of our commitment to the broadest possible access to data in the interests of enhancing innovation, collaboration, efficiency, accountability, and capacity building in the public interest. Explicit in these principles is the expectation that our partners will make data generated by foundation-supported activities broadly available as quickly as possible.

What is the difference between these principles and related elements of the Global Access Policy?

One of the goals of the foundation's Global Access Policy is to facilitate the broad and prompt dissemination of data and information to the scientific community. The Data Access Principles discussed here are an elaboration of Global Access Policy; the foundation is increasing the emphasis it places on ensuring access not only to the publications, but also to the data itself.

Why has the foundation developed these principles? The foundation is committed to global access to the outputs of the activities it funds to ensure that these generate the fullest possible public health benefits. These outputs include data, which, if more accessible, are more likely to lead to innovation, analysis, and insight above and beyond that of the individuals or institutions initially funded to gather it – thereby enhancing our overall prospects for saving lives and improving health.

What are the benefits of providing access to data?

Data access is intended to promote:

- *Innovation*, by encouraging diversity of analysis and scientific opinion, facilitating evaluation of alternative hypotheses, permitting meta-analysis, facilitating the synthesis of results from individual projects into a larger whole, and enabling exploration of topics not envisioned by the initial investigators
- *Collaboration*, between scientific teams and among diverse disciplines, resulting in greater productivity
- *Efficiency*, by avoiding unnecessary duplication of efforts, enabling secondary analyses, permitting the creation of new datasets by combining data from multiple sources, and re-directing resources to the most effective endeavors and promising approaches
- *Accountability*, by encouraging independent verification and analysis and permitting replication, thereby potentially improving data quality
- *Capacity Strengthening*, by facilitating the education of new researchers and enabling broader access to data for secondary analysis, which is of particular importance to investigators in developing countries

Scope of the principles

To which activities and types of data do these principles apply? These principles apply to final data, quantitative and qualitative, generated by foundation funding via grants. The ways in which the principles should be applied will vary based on data type and other considerations, and as negotiated with the grantee via the required Data Access Plan. Data that may be perceived as “negative” or unfavorable should not preclude access, as these data may be equally important in furthering the foundation's mission.

What is meant by final data? These principles apply to final, annotated datasets and associated documentation (protocols, questionnaires, codebooks, data dictionaries, etc.) Final data do not include laboratory notebooks, partial datasets, preliminary analyses, drafts of scientific papers, plans for future research, communications with colleagues, or physical objects, such as laboratory specimens. Some of these items may however be covered under the broader global access principles.

For which grants must a grantee prepare a Data Access Plan? A Data Access Plan must be prepared for all Global Health grants over \$500,000.

Do these principles apply to existing grants? The principles outlined here apply to new grants only that are made through the foundation's Global Health Program. For existing grants through the foundation's Global Health Program it is not necessary to develop a Data Access Plan; however, partners should be encouraged to provide access to the data they generate as a result of foundation support. These principles do not apply to grants made through the foundation's Global Development or United States Programs.

Do these principles apply to qualitative data? Yes. However, ensuring privacy and confidentiality prior to sharing qualitative data is likely to be more complex and costly than is the case with quantitative data. Related concerns should be articulated by the foundation partner in the Data Access Plan and decisions made on a case-by-case basis.

Data Access Plans

What should a Data Access Plan include? A Data Access Plan should at a minimum address the nature and scope of data and information to be disseminated, the timing of disclosure, the manner in which the data and information is stored and disseminated, and who will have access to the data and under what conditions.

What is the timing for preparing a Data Access Plan? Development of a Data Access Plan should begin early in the process of preparing a grant proposal. For grants, an initial Data Access Plan will be included within the proposal submission. A final Data Access Plan will usually be required prior to grant approval. There are circumstances in which this may not be practical. In these circumstances, the Program Officer can use his or her best judgment about making agreement on a Data Access Plan a grant milestone.

I plan to publish a paper with the findings of this project. Do I need to provide access to the data? Yes. While publication may certainly be one aspect of a Data Access Plan, the intent behind these principles is to push the envelope with respect to providing access to the underlying data.

I am concerned that others will publish papers based on my data before I publish their paper.

How should I proceed? Typically, this situation has been addressed in one of the following ways: 1) The grantee has exclusive use of the data for a specified timeframe (e.g., 12-18 months). After that timeframe, the data is made available for public access. 2) The data is released immediately but with an embargo. A user requesting access to the data must enter into a data use agreement, which specifies that the user may publish based on that data - with the exception of publications on to the grantee's subject of focus.

Will the foundation provide funding for data access activities? The costs involved with making global health related data widely available may be included in the proposed budget and will be subject to review and approval.

How quickly does the foundation expect data to be released? The value of data often depends on their timeliness; thus, data should be made widely available as soon as possible while safeguarding the rights and privacy of individual subjects. This time point will be influenced by the nature of the data collected.

What methods are recommended for data access?

There are many ways to provide access to data. These include:

- Deposit in a data repository (often referred to as an “archive”)
- Deposit in a data enclave
- Direct sharing by investigators or their institutions
- Mixed mode sharing

The method for providing data access that a partner selects is likely to depend on several factors, including the volume, sensitivity, and complexity of the dataset, and the volume of requests anticipated. When they are available, the preferred means of providing access to data is to deposit them in a data repository/archive. This facilitates wide availability and offers the benefit of professional curator and maintenance services.

Some data archives include data enclaves, which provide a controlled, secure environment in which eligible researchers can perform analyses using restricted data resources. This can be particularly useful when participant confidentiality concerns or third-party licensing or use agreements prohibit wider sharing or distribution.

Partners providing access under their own auspices may simply post information about data availability on their website and/or in published work, and then mail a CD containing the data to requestors. Alternatively, they may post the data on their institutional or personal website, and include the URL in published work. Partners sharing under their own auspices should consider requiring a data-use agreement to impose appropriate limitations on users. Such an agreement helps ensure that the data will be used for legitimate purposes that are in the public interest, and can also incorporate privacy and confidentiality standards to ensure data security at the recipient site.

Partners may also wish to develop a "mixed mode" for data sharing that allows for more than one version of the dataset and provides variable levels of access depending on the version. For example, a redacted dataset could be made available for general use, but stricter controls through a data enclave would be applied if access to more sensitive data were required.

Privacy, informed consent, and intellectual property

What are the risks of data access?

- Re-identification of subjects – there is a small chance of residual identification of individuals or groups when de-identified data sets are linked with other datasets or identifiable variables; the risk of re-identification can be mitigated through technical safeguards, and disclosure control. In

the consent process, research participants should be adequately informed that their data will be accessible amongst researchers, and care will be taken to protect their privacy

- Risk of misinterpretation or misapplication of data –can be minimized by providing well annotated codebooks and data documentation. All data users are bound by the principles of data access that specify good stewardship of data; responsibility for appropriate use of data rests with all investigators that access data.
- Risk of being ‘scooped’ is minimized through period of exclusivity to publish main findings (a “data embargo”). Proprietary data is not expected to be accessible.

How can the principles be balanced against the need to safeguard privacy rights of individuals? It is the responsibility of the investigators, their IRB, and their institution to protect the rights of participants and the confidentiality of their data. Data should be redacted to strip all individual identifiers, and strategies should be adopted to minimize risk of disclosing a participant's identity. Options to protect privacy include: withholding part of the data, statistically altering the data in ways that will not compromise secondary analyses, requiring researchers who seek data to commit to protect privacy and confidentiality, and providing data access in a controlled site. Some investigators use hybrid methods, releasing a redacted dataset for general use but providing access to more sensitive data through a user contract or data enclave. In most instances, sharing data is possible without compromising participant confidentiality and privacy.

How can the principles be balanced against the need to protect intellectual property? The foundation recognizes the value of intellectual property and the benefits of first and continuing use of data, but not from prolonged exclusive use. The foundation also understands that an institution's desire to exercise its intellectual property rights may justify a need to delay disclosure of findings; a delay of 30 to 60 days is generally viewed as a reasonable period for such activity. The timing of data release is at the discretion of the individual Program Officer.

This project involves proprietary data. Do the data still need to be shared? There may be circumstances where a cofunder or private-sector organization places restrictions on data sharing. These restrictions should be identified in the Data Access Plan and discussed with the Program Officer. The foundation will continue to monitor these cases over the course of the pilot year to develop a refined guideline if needed.