Workshop on Establishing a Central Resource of Data from Genome Sequencing Projects

Crosscutting ELSI Issues for the Data Access Working Groups

Wylie Burke, Lisa Brooks, Nicole Lockhart, Pearl O'Rourke, Laura Lyman Rodriguez

This document is intended for consideration by the groups working on the data access models.

The four data access models are:

1. Releasing data publicly.
2. Allowing access through some form of controlled access (e.g., dbGaP and EBI).
3. Allowing access to a central database through a researcher certification process.
4. Central analysis server groups providing summary data and analyses but not underlying data.

General goals to achieve the scientific "ideal" for broad analysis:

- Allow broad use of the data to address many research questions, rather than for specific diseases.
- Allow a broad set of users, such as researchers at academic, government, commercial, and insurance organizations, rather than only the initial study group and collaborators.

Baseline assumptions:

- All submitted data will be derived from procedures for which participants provided informed consent for research.
- An institutional IRB has reviewed consent and other study procedures and determined that the data are appropriate to submit to a repository utilizing the data access model under consideration (for public access (model 1) or for a form of controlled access (models 2 or 3); an IRB may need to approve the information that a server provides (model 4), without each data set having to have IRB approval for release that way).

Elements to consider regarding ethical and legal questions for each model:

- What ELSI issues are raised by each of the four models?
- What policy implications or challenges must be addressed to implement each of these models in a scalable way?
- Specific issues:

1. *Issues of informed consent provided by participants*

    1. Would expectations for consent processes differ among the access models, and if so, how?
    2. Existing research consent: What elements and processes of consent should be addressed to allow the data to be accessed through the model being considered?
    3. New consent will be obtained: What elements or language should be in new consent forms and processes to allow data to be accessed through the model being considered?
    4. Would data access be possible if the consent form mentioned samples but not data?

2. *Return of research results*

    1. If return of individual results is not a possibility in this model, how is non-return of results meeting NHLBI consensus criteria (Fabsitz *et al.*) justified?
    2. If return of results is a possibility, what procedures and criteria apply?
    3. Might researchers in the original studies have different responsibilities to return results than researchers who later use the data from the repository?

3. *IP issues*

    What IP issues are raised by the various access models, and how should they be addressed? For example, controlled access models allow agencies to expect that users of the data do not patent findings in a way that prevents others from accessing the data. What issues are most relevant for choosing among access models?

4. *Scalability, feasibility and acceptability*

    What issues are most relevant for choosing among access models?

5. *Issues for particular data access models*

    What would be the expectations and consequences for researchers using each access model?

    1. Releasing data publicly

        How would public data release affect the willingness of participants to be in a study?

    2. Allowing access through some form of controlled access model

        What challenges exist with current examples (e.g., dbGaP and EBI) and what are suggested improvements to enhance them?

    3. Allowing access to a central database through a researcher certification process

What elements are needed for a trustworthy certification process? Should there be two bins (any use of data, and conditions on use of data such as disease use restrictions), or should access be to the entire data set, with researchers being responsible for adhering to any use conditions? What, if any, considerations for enforcement or tracking of certified user behavior are incumbent in this model?

4. Central analysis server groups providing summary data and analyses but not underlying data

   What types of data could be released publicly? What types of data would require agreements by researchers to conditions, and what should those conditions be (e.g., controlled access, respecting disease data use restrictions)? What is gained (in an ELSI sense) by not providing individual-level data?