Amanda Courtright-Lim
courtright-lima@cardiff.ac.uk

Identifiability: Qualitative Analysis of the Ethical Issue of Data Sharing

The completion of the Human Genome Project marked a point in history that has set the field of biotechnology off at an exponential rate. Up until this point, sequencing genomic material was time consuming and complex. After the genome was assembled and the sequencing technology began to be utilized across the field of science, an untold amount of sequencing proceeded to be done. This led to a vast amount of sequencing data becoming the center point of research and a continued push to improve technology to produce even more data. Although the bench side of exploring the human genome was advancing, other areas lagged behind. These include ethics, policy, an understanding of the human genome, and the bioinformatics tools used to look at the data. While there have been advances in the latter two in recent years, progress has been slow to catch up in the areas of ethics and policy.

As the area of ethics attempts to catch up to this generation of vast amounts of sequencing data, one area that has emerged as needing additional research is in the identifiability of human derived microbial samples and RNA material. Identifiability is when genomic data could be linked to someone even when the sample has been de-identified, removing all links to the samples source. Advances in bioinformatics tools bring into question if RNA or microbial data derived from human samples can be used to identify the donor of the sample. Given the current push towards public data sharing of sequencing data, the issue of identifiability arises. With the strong possibility that people can be identified through different forms of data not initially seen as identifiable, how to safely share the data while protecting the privacy of the study participant becomes a key ethical consideration. To determine what this looks like, informed consent practices need to be considered to identify how data sharing databases are presented to the study participant. To contextualize this, a leading genomics research institute in Arizona was evaluated to determine where they stand on the topic of identifiability.

This project used a qualitative, multi-method approach to determine what steps needed to be taken to mitigate this ethical issue. The work done includes a literature review to evaluate the scientific communities' current viewpoints on this topic including existing policy, a platform analysis to determine the current practices for sharing this data, informal interviews to determine scientists' opinions on this topic, and qualitative assessment of the consent forms used at the institute. The data collected determined that the evaluated institute has stayed up to date with data sharing policy, the investigators are cognizant of the risk of sharing their data, and that the institute has a relatively standardized approach to consenting for data sharing. The data was used to evaluate where the institute stands on the problem of identifiability and determine that the institute wants to continue protecting research participants' data when data sharing.