# IRB data sharing guidance

When crafting an IRB proposal, it is important to include how data will be used once the data collection is over. Below, find information to include in IRB proposals, links to resources, and some template language for putting together consent forms, IRB proposals, and amendments.

## Protection of human subjects

In order to protect human subjects, ClinEpiDB takes a number of steps to minimize risk of study participant identification:

- Direct and indirect identifiers are removed for all studies
    - Names and GPS coordinates are removed
    - Dates are obfuscated
    - Locations are generalized
    - Additional indirect identifiers are removed or obfuscated as needed on a case-by-case basis
    - ID values are re-assigned to break the link to the institution's copy of the data
- Data access restrictions can be implemented as needed (see *Data Access and Use Policy*)
- Users are requested to acknowledge our *Responsible use agreement* which appears as a pop-up on the front page before accessing any of the studies for the first time. The agreement also contains a link to our *Data Access and Use Policy*.

## Informed consent

The onus is on the study team to provide information about the study to participants. When designing consent forms, it's important to include how confidentiality of subject records will be maintained without limiting researcher's ability to share data with the community.

- Avoid language that is inaccurate and overly restrictive, for example promising that only the study team will see the data, or that data will only ever be displayed in aggregate form
- Example 1 - We will make our best effort to protect your statements and answers, so that no one will be able to connect them with you. These records will remain confidential. Federal laws may require us to show information to university or government officials [or sponsors], who are responsible for monitoring the safety of this study. Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public.

- Example 2 - The information in this study will only be used in ways that will not reveal who you are. You will not be identified in any publication from this study or in any data files shared with other researchers. Your participation in this study is confidential. Federal laws may require us to show the information to university or government officials, who are responsible for monitoring the safety of this study.

## Transfer to ClinEpiDB

- Data is transferred to ClinEpiDB via Box, which provides end-to-end data encryption
- VEupathDB data is stored on secure, FISMA compliant servers, and websites are served over encrypted connections using secure hypertext transfer protocol (HTTPS) to protect against interception and manipulation of sensitive data

## Risk/benefit analysis

- Risk: due to the measures taken to protect participant confidentiality that are mentioned above, there is a minimal risk of identification of study participants
- Benefit: release of data may lead to further insights into X disease that lead to improvements in prevention and/or treatment, etc.
- Overall, the risk to benefit ratio is very small given the low likelihood of a participant being identified, while there is great potential benefit for increasing understanding of X disease

## Relevant NIH policies

- https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy.pdf - NIH genomic data sharing policy
- https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/human-data-sharing - NIH human data sharing policy
- https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/oversight/hds-sample_data_sharing_plan.pdf - NIH guide for developing human data sharing plans
- https://www.nih.gov/news-events/news-releases/nih-releases-strategic-plan-data-science - NIH strategic plan for data science