

What is a Data Repository?

- A data repository is a collection of information from many participants.
- Researchers can use data from a repository to answer research questions.
- Some repositories also contain biological samples (e.g. human tissue) for future use. These are called biorepositories or biobanks.

- Some examples of data repositories are the “All of Us” research project in the United States and the “UK Biobank” in the United Kingdom.
- These repositories have data and samples directly donated by thousands of participants with their consent.



How are data repositories created?

- Data repositories are created when data is obtained directly or indirectly:
- Direct: participants join a repository research study with the goal of collecting data for future use
- Indirect: after a clinical trial, saved data is anonymized then donated to a repository for reuse
- Participants consent to have their data stored for future use to help advance science

Is my data safe in a repository?



- Your data is saved securely.
- When researchers request to use data from a repository, they do not get any identifiable information, unless you have given permission.
- Data repositories store and share “de-identified” data.
- When data is de-identified, all obvious identifiers (i.e elements that could be used to identify you) are removed.

You have the right to ask questions about your data.

- What repository is my data saved in?
- Who has access to my data in the repository?
- What kinds of research will be done using this data?
- How will my data be shared with other researchers?
- Will researchers who request data from the repository know who I am?

