TYPES OF DATA PRODUCED

Types of data, samples, physical collections, software, curriculum materials, and any other materials to be produced in the course of the project.

The research project will generate ___________________________ (types of data, such as images, textual data, demographic, or biometric, or qualitative data). We will be measuring ___________________________ (describe what you’ll be measuring and the means you’ll use to measure it—e.g. serum creatinine in mg/dl from blood samples, and then describe any other lab values obtained from blood samples, or 3D MRI images and the x machine used to generate them). We will also be collecting data from _______________________ (list any other data collection methods and the instruments used, e.g., “the Clinical Global Impression Screening instrument and the Montgomery Asberg Depression Rating Scale, all of which use numerical scales”). For each type of data you’ll be collecting, describe the general variables you’re collecting and the instruments—if you collect demographic data, survey data, and image data, you’ll need to enumerate what you’ll be collecting for each overarching data type and how you’ll collect it. All data will be captured initially on/in _______________________ (paper forms or electronic data capture systems such as REDCap), and then _______________________ (describe any subsequent changes to data entry, such as entering data from paper forms into an electronic system, and what that electronic system is). _______________________ will be configured with value limits in order to ensure data quality, and standardized date formats will be employed (describe any data quality assurance measures you’ll put in place). The data will be processed using _______________________ (data analysis packages such as SAS, SPSS, R), and analysis will be completed by _______________________ (biostatistician, PI, etc).

DATA AND METADATA STANDARDS

Standards to be used for data and metadata format and content. Where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies.

Data will be described using common data formats, including _________________________ (e.g. NINDS data elements for laboratory specimens, demographics and protocol-related experiences such as dropouts or missed visits, ISO standards for date or date/time elements, PhenX Medication History data elements, and Darwin Core data for overall description of the research project). Note: if you would like assistance identifying appropriate common data elements, please contact a librarian.

A ReadMe file will be provided, containing information about the methods of data collection, including the names of instruments used to assess ________________ (e.g. mood, pain, etc.), a full list of the laboratory values collected from ________________ (urine, blood, skin, tissue) specimens and an explanation of how files are related. We will also provide a data dictionary with all the names, descriptions, and acceptable values of variables.

Data will be collected in _________________________ (e.g. an Excel spreadsheet and in SAS). Final datasets will be saved in ________________ (e.g. csv, tiff) format for software interoperability.

POLICIES FOR ACCESS AND SHARING

Policies for access and sharing; Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

Curated data will be available through the _________________________ (trusted repository for data, such as the National Database for Clinical Trials Related to Mental Illness, or FlyBase). Subjects’ _________________________ (variables you’ll share, e.g. lab values) will be available freely, along with _________________________ (any other variables you’ll share, e.g. some demographic variables). List the variables that will be made available for public use, as appropriate. All subjects _________________________ (for human subjects, describe
any process of protections, e.g. “were anonymized at the start of the project, so no personally identifiable information will need to be removed. However, the inclusion of all demographic variables collected could facilitate reidentification of participants, so only the minimum number of demographic variables necessary for research will be released as part of the dataset.” No data from ___________ (e.g. clinical notes, dates of visits, or medication lists) will be provided. We will consult with university personnel from the Research, Economic Development, and Engagement office, along with a statistician, to determine the appropriate demographic data to be released with the ___________ data. A ___________ (e.g. co-investigator or librarian) will assist with the process of preparing the cleaned data for submission to the ___________ repository.

Data will be released ___________ (time of data sharing projected, e.g. once the trial has been completed, at the time when the manuscript from the trial is published). Copies of data withheld from the publicly available dataset will be retained by the PI and East Carolina University, stored on the university's secure server, and a secure drive retained by the researcher.

Please note: if it is not appropriate to share any data, due to concerns about human subjects protection or patents, please provide an explanation.

POLICIES FOR RE-USE, REDISTRIBUTION

Policies and provisions for re-use, re-distribution, and the production of derivatives.

Cleaned data will be freely available. Researchers requesting access to the full dataset may request access by contacting ___________ (e.g. the PI). ___________ (anything not in the publically available data, e.g. Clinical notes and the participant key) will not be made available to other researchers. For ___________ (any scales or instruments, such as mood assessment tools) not in the public domain, requestors will need to obtain permissions from the copyright holder for those tools prior to accessing the full instruments. All requests to access the data will undergo review by ___________ (e.g. IRB, HIPAA, ITCS), and other appropriate university personnel involved in assurance of research and data integrity and security prior to approval. This data is intended for use by other ___________ (e.g. researchers, students, and clinicians). This data may be used for ___________ (e.g. secondary analysis, meta-analysis, or educational purposes). For derivative products, we request that the dataset be cited. If you already have a preferred citation for the dataset, you can provide that here, along with information about the type of license you’ll assign (e.g. CC-BY-NC); talk to a librarian if you’d like help with this.

PLANS FOR ARCHIVING AND PRESERVATION

Plans for archiving data, samples, and other research products, and for preservation of access to them.

A secure copy of the data will be stored in the departmental East Carolina University Pirate Drive. Pirate Drive provides long-term, encrypted storage of data and regular backups. For purposes of long-term preservation, upon request, East Carolina University's Department of Information Technology and Computing Services can, upon request, assist with file format migration over time. Interoperable file formats will be used to minimize frequency of required format migration over time and increase the usability of the dataset. ___________ (e.g. copies of the instruments used, the protocol, the metadata, the data dictionary, statistical analyses, and any/all published manuscripts) from this study will be deposited in the Pirate Drive as well in order to preserve all products of the research process and ensure meaningful reuse of research.

Note: keeping all data-related products of research is extremely important for reproducibility purposes, including statistical software versions if possible.