# DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on [sharing.nih.gov.](https://sharing.nih.gov/) The Plan is recommended not to exceed two pages.

Text in italics should be deleted. There is no “form page” for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format*

shown below.

# Element 1: Data Type

Briefly describe the scientific data to be managed, preserved, and shared.

**Types and amount of scientific data expected to be generated in the project: Summarize the types and estimated amount of scientific data expected to be generated in the project.**

NIH defines scientific data as the recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens

**VIVLI SAMPLE TEXT: For this project we anticipate data collection for 500 participants for a Phase II trial for 4 timepoints over two years. We will collect demographic, genomics, biological specimen, QOL and outcomes information.**

**Data will be converted to R and csv formats for analysis and sharing.**

DMP TOOL SAMPLE TEXT: This project will produce \_\_\_\_\_\_\_\_\_ [Data type, e.g., imaging, sequencing, experimental measurements] data generated/obtained from \_\_\_\_\_\_\_\_\_\_ [e.g., instrument, method, survey, experiment, data repository].  Data will be collected from  \_\_\_ [number] of research participants/specimens/experiments, generating \_\_\_ [number] datasets totaling approximately \_\_\_ [amount of data] in size. The following data files will be used or produced in the course of the project: \_\_\_\_\_\_ [list input data files, intermediate files, and final, post-processed files]. Raw data will be transformed by \_\_\_\_ [analysis, method] and the subsequent processed dataset used for statistical analysis. To protect research participant identities, \_\_\_\_\_\_\_\_\_\_\_ [e.g., individual, aggregated, summarized] data will be made available for sharing.

# Scientific data that will be preserved and shared, and the rationale for doing so:

Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

*Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision. NIH expects that researchers will take steps to maximize scientific data sharing, but recognizes that certain factors (i.e., ethical, legal, or technical) may necessitate limiting sharing to some extent. Foreseeable limitations should be described.*

1. **Metadata, other relevant data, and associated documentation: Briefly list metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data. Indicate if none.**

**VIVLI SAMPLE TEXT: In addition to the patient-level data, the metadata, data dictionary, statistical analysis plan, and final protocol will be shared. The sharing of the data dictionary, statistical analysis plan and final protocol with amendments will enable researchers to understand how the data was collected and to correctly interpret the data for future secondary analysis. The final patient level dataset will include demographics, primary and secondary outcomes as well as …**

# Element 2: Related Tools, Software and/or Code:

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

# Element 3: Standards:

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

**VIVLI Notes: Vivli does not currently mandate any data standard, we suggest that files contributed should be provided in a format that can be used by the standard statistical packages (R, SAS, STATA, Excel, csv) used for analysis. Data sets are not required to be standardized; however, we recommend that data sets be made available in CDISC SDTM (Standard Data Tabulation Model) format to support for the most efficient data aggregation, re-use, and sharing. CDISC SDTM is now becoming the consensus standard for those therapeutic areas that have a CDISC standard for data collection** [**https://www.cdisc.org/standards/therapeutic-areas**](https://www.cdisc.org/standards/therapeutic-areas)**.**

# Element 4: Data Preservation, Access, and Associated Timelines

1. **Repository where scientific data and metadata will be archived:**

Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see [Selecting a Data Repository](https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/selecting-a-data-repository)).

**VIVLI SAMPLE TEXT: We plan to share datasets resulting from the xxx study in Vivli, a non-profit institution that supports the Vivli repository). The Vivli Platform is a sustainable platform that provides metadata that are indexed via Datacite and provides DOIs for all objects. The data are backed up to the Microsoft Azure cloud. All data is shared publicly under a data use agreement (DUA) through a managed access process. Long-term access is freely provided to researchers.**

# When and how long the scientific data will be made available:

Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

**Vivli Sample text: The data shared will be archived and available on the platform for request by researchers for a minimum of 10 years after contribution.**

**Data will be made accessible no later than the time of our associated publication or the end of the grant period (whichever comes first). On an ongoing basis, Vivli evaluates its data holdings with regard to maintaining access and reserves the right to discontinue the distribution of a data collections when deemed appropriate. When materials are deaccessioned, the data are no longer publicly accessible at Vivli, they may still be preserved in Vivli’s storage vault. Because digital files are assigned a persistent digital object identifier (DOI), the study description is still available to view, but is not searchable through Vivli.**

# Element 5: Access, Distribution, or Reuse Considerations

1. **Factors affecting subsequent access, distribution, or reuse of scientific data:**

NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See [Frequently](https://sharing.nih.gov/faqs%23/data-management-and-sharing-policy.htm) [Asked Questions](https://sharing.nih.gov/faqs%23/data-management-and-sharing-policy.htm) for examples of justifiable reasons for limiting sharing of data.

**Vivli Sample text: We will make the data available via the Vivli platform (**[**http://vivli.org**](https://vivli.org/)**/). Vivli is a non-profit clinical research data sharing platform that has been created to meet the needs of researchers who use and produce clinical research data worldwide. Using the Vivli platform, researchers can share or access anonymized data using a managed access process. In order to access anonymized individual patient-level data (IPD) arising from this project, users complete the Vivli data request form and sign the Vivli Data Use Agreement, which limits subsequent use to the terms of the approved proposal and requires that users maintain data security, and refrain from any attempts to re-identify research participants or engage in unauthorized uses of the data. Vivli will then make the data available via secure download. Researchers have a requirement to publish their findings as part of the Data Use Agreement and once the project is complete, must confirm that the original data and documents have been destroyed.**

# Whether access to scientific data will be controlled:

State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

**Vivli Sample text: The data being shared is human data from clinical trials and therefore a higher level of protection is required. Access to this data will be controlled by a managed access process whereby access is provided only after approval.**

# Protections for privacy, rights, and confidentiality of human research participants:

If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

**Vivli Sample text: The Vivli platform has multiple levels of protection for managing human participant data. There is a requirement that all data contributors attest that the data is anonymized prior to contribution. There is a managed access process in place for data requests and data are securely archived in Microsoft Azure. Controls and audit procedures are in place. Each user must sign a Data Use Agreement and there is an established policy for managing violations. In addition, this study will include the appropriate informed consent, based upon NIH** [**Informed Consent for Secondary Research with Data and Biospecimens**](https://osp.od.nih.gov/wp-content/uploads/Informed-Consent-Resource-for-Secondary-Research-with-Data-and-Biospecimens.pdf) **to allow for the data sharing outlined in this plan.**

# Element 6: Oversight of Data Management and Sharing:

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

**Vivli Sample text: Oversight of the Data Management and Sharing of this study, is the responsibility of the Principal Investigator(s) as well as the institution official (title and role goes here). Progress will be reported as part of the Research Performance Progress Report (RPPR)—Annual, Interim, and Final. As evidence of the delivery of this Data Management and Sharing Plan, will be met by sharing the Digital Object Identifiers (DOI) provided by Vivli that link directly to the metadata associated with this study on an annual basis. The Data Manager / PI in charge of the program will review if any trials are eligible for sharing each quarter per their DMSP as papers are published. OSP has created a DSMP compliance system that will be monitored by the PI of each program. This compliance program is (describe here …)**