DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
Food and Drug Administration

Data Management Plan

Form and Instructions
Intramural Research Programs

A data management plan (DMP) describes how scientists will handle digital data both during research and after a research project is completed. Preparing a DMP before digital data are collected ensures that digital data are correctly formatted, annotated, and organized from a research project's outset.

DMPs also enable supervisors to ensure that their staff have the required resources to comply with digital data management and sharing policies. In addition, the DMP allows supervisors to understand the scope, format, and structure of the digital data, should others need to access the data.

Researchers are expected to conduct research pursuant to an approved DMP. However, it is a living document and may evolve as the research project evolves. A DMP should be reviewed for possible revision by the researcher and his or her supervisor whenever a data management procedure is changed, or a formal status update, progress report, or reporting of results is provided to management or to the FDA office providing intramural grant funding. Such reports should include a statement of compliance with the approved DMP or a description of, and reasons for, any departures from the approved one.

Supervisors reviewing requests to conduct research or requests for funds will review DMPs on their merits in making decisions to approve and fund research. Supervisors will consider the following in evaluating proposed DMPs:

- restrictions on the disclosure of research data based on FDA regulations, statute, privacy concerns, proprietary interests, requirements imposed by an IRB, or otherwise;
- data preservation or Federal records retention requirements;
- available Center or Agency resources—monetary, physical, human, technological, or otherwise;
- the value of the long-term preservation of research data—both to the public and to the specific FDA strategic priorities or research program to which the research proposal relates; and
- the cost of long-term preservation of research data—to FDA, the Center, and the research program to which the research proposal relates.

Supervisors should weigh the value of the long-term preservation of, and access to, data against the associated cost and administrative burden. Additionally, for research proposals submitted for approval, supervisors must ensure that the researcher has the resources and funds necessary to comply with the DMP, as approved.

Researcher Information

Lead Researcher	Organization
Date	This is an ☐ initial DMP or ☐ a revised DMP

Title of Research Proposal/Project

Have all of the researchers involved in the project successfully completed the Responsible Conduct of Research training within the last four years?

☐ Yes ☐ No

PSC Publishing Services (301) 443-6740 EF

Digital Data

1. Will digital data be collected, produced, or recorded during this proposed research? *(If NO, then a DMP is not required.)*

   ☐ Yes ☐ No

2. Will the proposed research use a contract, grant, or assistance agreement *(see instructions)* to acquire or collect digital data? *(If YES, please see instructions)*

   ☐ Yes ☐ No

3. Are the metadata schema (attributes of the data) defined *(see instructions for definition)*? *(If NO, please see instructions)*

   ☐ Yes ☐ No

4. What type(s) of digital data will be collected *[check all that apply]*

   ☐ Instrument-generated digital data output (including images)

   ☐ Researcher-generated databases, tables, and/or spreadsheets

   ☐ Other: please add any additional information/attachments to the email

5. Will all digital data supporting a peer-reviewed research article be in (or be converted into) an open, non-proprietary, format(s)? *(If NO, please see instructions)*

   ☐ Yes ☐ No

6. Will the proposed research use existing digital data (i.e., for meta-analysis or new analyses)? *(If YES, please see instructions)*

   ☐ Yes ☐ No

Data Storage, Retention, and Preservation

7. Estimate the total anticipated digital data that will be generated during this proposed study. Will there likely be more than 1 terabyte of digital data collected? *(If YES, please see instructions)*

   ☐ Yes ☐ No

8. Will the digital data related to a peer-reviewed article be stored on FDA digital storage resources indefinitely? *(If NO, please see instructions)*

   ☐ Yes ☐ No

9. Will back up and other redundant storage strategies be used to ensure the data's security and integrity? *(If NO, please see instructions)*

   ☐ Yes ☐ No

Data Sharing and Public Access

10. Does the researcher commit to making all metadata (structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage the digital data) relating to a peer-reviewed article freely available in one or more open, non-proprietary formats to the public upon publication? *(If NO, please see instructions)*

    ☐ Yes ☐ No

11. Does the researcher commit to making all final digital data supporting a peer-reviewed article freely available in one or more open, non-proprietary formats to the public upon publication? *(If YES, part of the data, or NO, please see instructions)*

    ☐ Yes ☐ No

12. Will the digital data contain any private, privileged, or otherwise confidential information that must be redacted before public release? *(If YES, please see instructions)*

☐ Yes ☐ No

13. Will there likely be an unusually large amount (e.g., more than 1 terabyte (1,000 gigabytes)) of digital data generated that will require storage for public access? *(If YES, please see instructions)*

☐ Yes ☐ No

14. Where/how will the data be made publicly available? *[check all that apply]*

- ☐ The data will not be publicly available
- ☐ Data will be made available upon individual request
- ☐ Figshare
- ☐ Dryad Digital Repository
- ☐ Dataverse
- ☐ Data repositories maintained by National Library of Medicine
- ☐ Supplemental materials to the publication
- ☐ FDA web page (e.g., NARMS)
- ☐ Other digital dataset repository: please identify
- ☐ To be determined

Explanations and attached information *(please add any additional information/attachments to the email)*

INSTRUCTIONS

Researcher Information

1. Lead researcher's name & organization

   The lead researcher is responsible for submitting the data management plan and ensuring the digital data is collected and managed in accordance to this DMP. For the organization, please list the Center/Office and any relevant divisions and branches (e.g. FDA/OC/OCS/OSI)

2. This is an initial DMP or a revised DMP

   DMPs are living documents. As the research progresses and changes are made to it, the plan may be revised and resubmitted.

3. Have all of the FDA researchers involved in the project successfully completed the Responsible Conduct of Research training within the last four years? Supervisors should ensure that staff have taken the Responsible Conduct of Research training within the past four years before approving this DMP.

Digital Data

Digital data are the digitally recorded factual material that would be commonly accepted in the scientific community as necessary to validate published, peer-reviewed scientific articles.

The following are expressly *excluded* from the definition of digital data

- preliminary materials underlying the data or factual information, including lab notebooks, preliminary analyses, drafts, plans for future research, peer-review reports, communications with colleagues, or physical objects, such as lab specimens;
- data shared with FDA but owned by other organizations (e.g., aggregate electronic health care data from other parties used by FDA in product safety monitoring pursuant to FDA's Sentinel program);
- data FDA receives as part of an application for market authorization or application for exemption from marketing restrictions for investigational use;
- data obtained under licensing or data use agreements, or cooperative research and development agreements that include terms restricting the release and/or sharing of the data;
- data or information not available for disclosure pursuant to statute or regulation; and
- technical (i.e., designs and drawings) and administrative (i.e., personnel records) data.

1. Will digital data be collected, produced, or recorded during this proposed research? If NO, then a DMP is not required.

2. Will the proposed research use a contract, grant, or assistance agreement to acquire or collect digital data?

   NOTE: This is specifically for extramural research. Do not indicate "yes" if a contract, grant, or assistance agreement is used to fund research that is primarily conducted at FDA (e.g., ORISE fellows or outsourced research services such as DNA sequencing).

   - If YES and an FDA employee is the lead or principal investigator, please add any additional information/ attachments to the email about the digital data to be collected by the recipient of FDA funds to this DMP.
   - If YES and an FDA employee is the Program Official, ensure that the extramural recipient of FDA funds has submitted a DMP (effective 10/01/2016). The Program Official does not need to complete one.

   If Federal funds are used to acquire or collect digital data, then public access to resulting publications and the digital data are required.

3. Are the metadata schema (attributes of the data) defined?

   - If NO, please attach an explanation for why not and the plan to define the metadata.

   Metadata are often defined as data about data. It is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. A good example of metadata is the cataloging system found in libraries, which records the author, title, subject, and location on the shelf of a specific book.

   Metadata usually falls into three categories:

   - **Descriptive metadata** describes an information resource for identification and retrieval through elements such as title, author, and abstract.
   - **Structural metadata** documents relationships within and among objects through elements such as links to other components (e.g., an index or Table of Contents).
   - For purposes of file management, rights management and preservation, **administrative metadata** helps to manage information resources through elements such as version number, archiving date, and other technical information.

   Metadata schemes (also called schema) are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. Metadata schemes generally specify names of elements and their semantics. They may specify content rules for how content must be formulated (e.g., how to identify the main title, filename, and date format), representation rules for content (e.g., capitalization rules), and allowable content values (e.g., terms must be used from a specified controlled vocabulary).

   An example of a metadata scheme is the use of annotated study forms (each unique study form is annotated with its corresponding table name(s), variable names, and formats. Input fields are populated with controlled terms where applicable). For example, case report forms (CRFs) are commonly used in clinical and nonclinical research; annotated CRFs may be included as an integrated component of a DMP to precisely demonstrate data definitions, relationships, and organization structure.

   A metadata scheme specifies how the digital data will be labeled and aims to facilitate access to the data and its interpretation. This includes the consistent and clear labeling of files, spreadsheet data, figures, and other digital data. Different digital data repositories may have different guidelines on metadata formatting and organization. In order to facilitate the later submission of the digital dataset it is highly recommended that the metadata scheme fulfill the requirements of the digital repository where the data are expected to be deposited.

4. What type(s) of raw digital data will be collected [check all that apply]

   - Instrument-generated digital data output (including images)
   - Researcher-generated databases, data transformations, data linkages, tables, and/or spreadsheets
   - Other: please describe

   Please select the types of digital data that will be collected. If the data do not fall into one of the broad classifications, please add a description of the type(s) of data to the email. Researcher generated digital data are typically described as “data entry” that complies with open data standards (i.e., file types that facilitate data sharing). If the data will be instrument-generated, it would be beneficial to determine in advance whether the data can conform to open data standards (see next question).

5. Will all digital data supporting a peer-reviewed research article be in (or be converted into) an open, non-proprietary, format(s)?

   - If NO, please add an explanation for why not to the email. Also identify the proprietary (closed) file formats, including the software and software version that will generate the data file.

   An open file format is a published specification for storing digital data, usually maintained by a standards organization, which anyone can use and implement. Examples of open formats are American Standard Code for Information Interchange (ASCII) text (*.txt), comma separated values (*.csv), or Joint Photographic Experts Group (JPEG) (*.jpg).

   A closed file format contains data that are ordered and stored according to a particular encoding scheme that a company or organization designed to be secret unless decoded and interpreted with particular software or hardware that a company or organization has developed. Examples of proprietary formats are Microsoft Word (*.doc), Microsoft Excel (*.xls), and Microsoft PowerPoint (*.ppt).

   The key to long-term archiving of digital data is using open formats that may be accessed long after acquisition software has been rendered obsolete. Digital data should be organized into an open file format unless closed file formats cannot be converted or exported to an open file format. If open format isn't possible, please provide detailed information about the software, software version, and any other pertinent information (such as dedicated hardware) that would allow long-term access to the digital data.

6. Will the proposed research use existing digital data (i.e., for meta-analysis or new analyses)?

   - If YES, please add an explanation of the source, including ownership, and any restrictions on use or publication of these data to the email.

   Existing digital data may be subject to restrictions on sharing, publication, and long-term storage. There may be additional security concerns surrounding the use and storage of the data. It is important to understand those restrictions before the data are used for analysis.

Data Storage, Retention, and Preservation

7. Estimate the total anticipated digital data that will be generated during this proposed study. Will there likely be more than 1 terabyte (1,000 gigabytes) of digital data collected?

   - If YES, are additional resources (including personnel) required to support collection, live storage, regular back up, or analysis of these digital data?
     - If YES, please add an explanation and provide the plan and cost for the additional IT resources (including personnel) to the email.

   These costs should be part of the research proposal and the evolving research plan. The researcher should ensure that IT resources (including imposed quota limitations) are sufficient to store and archive the digital data as they are collected, analyzed, and backed up. If large amounts of digital data are to be collected, the researcher should ensure that the current infrastructure can accommodate the data. Consideration should also be given to changes to infrastructure and how that may affect the ability to access digital data.

   In addition to sufficient storage space, consideration should be given to additional computational power required for collection, processing, and analysis (e.g., high performance computers). If personnel are needed (i.e., informatics experts), please describe the needed requirements.

8. Will the digital data be stored on FDA digital storage resources indefinitely?

   - If NO, please add an explanation and indicate how long the digital data will be retained and how they will be decommissioned to the email.

   Ideally, digital data will be stored indefinitely. However, if resource limitations do not allow for indefinite storage, please outline the plan (including the relevant record schedule) for final disposition of the digital data.

9. Will back up and other redundant storage strategies be used to ensure the data's security and integrity?

   - If NO, please add an explanation for why not to the email.

   It is critical to take every possible measure to prevent the loss of digital data due to hardware or software failure. Should a loss occur, a back-up copy can be used to restore the original.

Data Sharing and Public Access

10. Does the researcher commit to making all metadata (structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage the digital data) relating to a peer-reviewed article freely available in one or more open, non-proprietary formats to the public upon publication?

    - If NO, please add an explanation for why not to the email.

    Every effort should be made to include all dataset metadata in the final published article.

11. Does the researcher commit to making all or a subset of final digital data supporting a peer-reviewed article freely available in one or more open, non-proprietary formats to the public upon publication?

    - If YES, part of the data, then add to the email (a) descriptions of the digital data that will, and will not, be freely available, and (b) a justification for not making some of these data freely available.
    - If NO, please add an explanation for why not to the email. If applicable, also explain why data redaction cannot be practically applied.

    A researcher must provide access to the digital data supporting the published research upon publication of a peer-reviewed article. If not, the researcher must justify why the data should not be made publicly available.

12. Will the digital data contain any private, privileged, or otherwise confidential information that must be redacted before public release?

    - If YES,

      a) generally describe the types of confidential information requiring redaction,

      b) generally describe any strategy employed to de-identify personally identifying data, and

      c) describe the security measure that will be in place to ensure limited access to confidential information.

    The expectation is that the digital data will be made public after proper redaction and de-identification, which may require additional training and, possibly, interaction and coordination with the FOIA staff. Planning ahead and careful data and metadata structuring will make future redaction easier.

13. Will there likely be an unusually large amount (e.g., more than 1 terabyte (1,000 gigabytes)) of digital data generated that will require storage for public access?

    - If YES, will there be a need to procure additional IT resources (including personnel) to store, retain, preserve, and provide access to digital data? If so, please add an estimate to the email.
    - If YES, please add an explanation and provide an estimate for the additional IT resources (including personnel) to the email.

    Note that long-term storage and storage for public access are typically not the same for security purposes. Also, the project may collect a large amount of data, but only a fraction will be made publicly available.

    If the data will not be directly accessible by the public, consideration should be given to personnel needed to provide the data upon request by a member of the public.

14. Where will the data be made publicly available?

    Early identification of a dataset repository will enable the researcher to structure the data at the outset to facilitate subsequent uploading and sharing. Please see the list of URLs at the end of the document for more data repository resources, including discipline-specific repositories.