**Institute of Education Sciences (US Dept of Education)**

[An Overview or Summary is not necessary. If you want to add a short paragraph, place here.

IES DMPs can be up to 5 pages. This does not count against the narrative total. Remember that IES approval of the DMP is required prior to the commitment of funds for the grant.]

**I. Roles and responsibilities in the management and retention of data**:

[Enter content here and then remove the Guidance prior to submission]

**Guidance**:

Roles and responsibilities of project or institutional staff in the management and retention of research data.

Explain how the responsibilities regarding the management of your data will be delegated. This should include project management of technical aspects and contributions of non-project staff - individuals should be named where possible. Remember that those responsible for long-term decisions about your data will likely be the custodians of the repository/archive you choose to store your data. While the costs associated with your research (and the results of your research) must be specified in the Budget portion of the proposal, you may want to reiterate who will be responsible for funding the management of your data.

Consider the following:

* Who will be in charge of the project? Include PI and co-PI names and institutional affiliations if this is a collaborative project.
* Who will be responsible for data management and for monitoring the data management plan? Provide name(s) and explain their roles and responsibilities.
* How will adherence to this data management plan be checked or demonstrated?
* Do you have a data manager or data analyst? Provide name(s) and explain their roles and responsibilities.
* Who will be responsible for the creation of the data documentation? Provide name(s) and explain their roles and responsibilities.
* Who will manage the storage of the data during the project? Provide name(s) and explain their roles and responsibilities.
* Who will have responsibility over time for decisions about the data once the original personnel are no longer available? Provide name(s) and explain their roles and responsibilities.

IES says: “*The DMP should describe staff responsibilities in creating, maintaining, and documenting datasets and support access. Staff responsible for creation and oversight of data (as well as regularly updating) should be clearly identified and their duties described*.”

**Very Important**: Include a succession plan for everyone. Especially the PI and co-PIs. You don’t need to name individuals, but if you can, do so. Indicate that they have the necessary authority to handle the project. For example, you might say that “they will be replaced with staff members at a similar level of skill and experience” or “our CASTL group has several staff who can handle the data management requirements of this project”.

**II. Types of Data to be shared:**

[Enter content here and then remove the Guidance prior to submission]

**Guidance**:

The DMP should describe the types of data to be produced in the course of the project, and their respective formats.

Provide a description of the data you will collect or re-use, including the file types, dataset size, number of expected files or sets, and content. Data types could include text, spreadsheets, images, classroom observations, software, audio files, video files, reports, surveys, patient records, policy documents, demographic data, assessments, etc.

Consider the following:

* What data will be generated in the research?
* What data types will you be creating or capturing?
* How will you capture or create the data?
* If you will be using existing data, state this and include how you will obtain it.
* What is the relationship between the data you are collecting and any existing data?
* How will the data be processed?
* What quality assurance & quality control measures will you employ?

Be as thorough as possible. Break the different data types into paragraphs, or use a bulleted list. Provide a brief description of the data type, how you are collecting it (provided by XXX, notes collected in interviews, online survey, taping the classroom, children assessed by teachers, transmitted electronically by school district) and the format (.docx, .txt, .pdf, Dedoose file, .mp3, etc.). For interviews, include information such as “transcribed and stored as .docx files”. For data that are collected in paper formats, such as interview notes, paper surveys and observation notes, indicate if they are being entered into spreadsheets or databases, and if the paper docs will be kept or destroyed.

If you are keeping the confidential paper documents, you may have specific requirements in your IRB Protocol. Often 2 levels of security are required. This can be as simple as a locked file cabinet in a locked office. Be sure to indicate who has access to those records.

If you are destroying confidential data, indicate that you are doing so responsibly and permanently. ISPRO/Records Management can provide information, and UVa Recycling can handle the shredding of confidential records. Electronic records should be handled in accordance with UVa Policy IRM-004: <http://uvapolicy.virginia.edu/policy/IRM-004> UVa also has a Records Management policy IRM-007: <http://uvapolicy.virginia.edu/policy/IRM-017>

**III. Format of final dataset:**

[Enter content here and then remove the Guidance prior to submission]

**Guidance**:

IES defines final research data as the recorded factual materials commonly accepted in the scientific community as necessary to document and support research findings. This dataset should include the final, cleaned data and may include both original data and derived variables, which will be fully described in accompanying documentation. Final research data do not include laboratory notebooks, partial datasets, preliminary analyses, and drafts of scientific papers, plans for future research, peer review reports, or communications with colleagues.

This applies to both new data collections and to data obtained from transforming or linking extant datasets. There may be circumstances, such as when a state or district will not allow student data to be released, where investigators will not be able to share their complete data set. IES expects primary data collected by the project or extant data obtained from a private source to be shared.

This is the dataset that you will share per IES requirements. They will only accept electronic file(s) in most cases. They advise that the data be made available in several electronic file formats, one of which is non-proprietary to encourage data reuse. Documents should be .pdf or .text, spreadsheets should be .csv, etc. You will probably have several different data types and corresponding file formats. Be sure to indicate which format will be used for each file.

If you are working in a program such as Dedoose or NVivo, you can share the file in that format, but be sure to have a copy of the data in a non-proprietary format as well.

Your chosen data repository may have a preferred dataset type, and may provide the data to users in several formats.

Consider these questions:

* What files will be included in the final dataset?
* What file format will each file be in?
* Are any of the datasets from extant sources?
* Are their limitations imposed by those sources on data sharing? If so, be sure to identify them in section IX.

**IV. Dataset Documentation to be provided:**

[Enter content here and then remove the Guidance prior to submission]

**Guidance**:

Documentation that provides all the information necessary for other researchers to use the data must be prepared.

Describe the format of your data, and think about what details (metadata) someone else would need to be able to use these files. Metadata may entail descriptions of research details such as: experiments, apparatuses, computational codes, etc.

Consider these questions:

* Which file formats will you use for your data, and why?
* What form will the metadata describing/documenting your data take?
* How will you create or capture these details?
* Which metadata standards will you use and why have you chosen them? (E.g. accepted domain-local standards, widespread usage).
* What contextual details (metadata) are needed to make the data you capture or collect meaningful?
* What metadata and documentation will be submitted alongside the data when you deposit it in a repository at the end of the project?
* What metadata and documentation will need to be created to explain the differences between the original dataset(s) and the anonymized version that you deposit in a repository?

Metadata, or dataset documentation, provides the information necessary for others to find, understand, and use your data. There are metadata schemas for many disciplines, but there is no set of standards in education research. There are several types of documentation that are needed in order to support the use of data: files that support discoverability (i.e., the likelihood that researchers will find the data) and files that describe how to use the data. Researchers should document everything and strive to make notation as interpretable as possible.

Ideally, the information you provide in your data documentation will be sufficient for anyone with the appropriate knowledge and skills to replicate your research results using your data.

For each data file you should include the following information:

* Description of each study including the name, sampling, data collection procedures, date & time collected, description of missing data, description of the data collecting tools (including information on version number, subscales, publisher, etc.)
* Codebook which will include all variable names, descriptions, and the scale or coding of the variables.
* Data dictionary or a cross-walk for datasets that you de-identify for sharing. Be sure to explain what the differences are between the original dataset and the de-identified one.

Metadata files should either be included in the dataset itself, provided as a ReadMe.txt file, or as a supplemental file in a .pdf or text file format.

**V. Procedures for maintaining confidentiality:**

[Enter content here and then remove the Guidance prior to submission]

**Guidance**:

Confidentiality of human subjects and privacy issues are extremely important to IES. Data sharing must not compromise this commitment. IES notes that it may be difficult to provide access to the data due to institutional policies, IRB rules & protocols, state and federal laws and regulations concerning the rights and privacy of individual study participants. Investigators should therefore plan their study design and procedures to enable data access. They should seek to optimize the opportunity for data sharing while working with their IRB to protect the privacy rights of study participants and confidentiality of the data.

Consent forms and IRB approvals should reference future sharing of data and stipulate the conditions that will be put in place to protect the privacy of participants. They should be drafted with the idea of encouraging data sharing. The content of the informed consent can limit how that data can subsequently be used, including data sharing. ICPSR has some suggested language that can be included in your forms. <https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confientiality/conf-language.html>

Data that are to be shared should be free of identifiers that would allow linkages to individuals participating in the research as well as other elements that could lead to deductive disclosure of the individual study participants. In cases where data cannot be free of identifiers or when identifiers are important for linking datasets, then investigators should consider restrictions on data sharing, as provided by data archives or enclaves.

For the **Human Subjects** section of the application, discuss the potential risks to research participants posed by data sharing and steps taken to address those risks.

Consider these questions:

* Have you gained consent for data preservation and sharing?
* What have you done to comply with your obligations in your IRB Protocol?
* Are there ethical and privacy issues? If so, how will these be resolved?
* How will you protect the identity of participants?

For each study that you conduct provide the following information:

* participants
* type of data to be collected
* if you are collecting it
* if it is being provided by an existing school or other entity
* the form you will receive it in (raw or anonymized) if it is provided by another entity
* limitations imposed on you by that data source
* randomizing procedures
* types of variables you will be looking at
* timeline for the study
* data collection period

Be very careful of the risks of ‘deductive disclosures’ and indirect identifiers.

Be sure to include information about your consent form(s) for each study.

**VI. Expected Schedule for data access:**

[Enter content here and then remove the Guidance prior to submission]

**Guidance**:

Timely data sharing is important to the scientific process. IES anticipates that the data will be shared no later than the acceptance of the main findings from the final study dataset in a peer-reviewed scholarly publication and that data will be available for at least 10 years. Data may also be made available earlier as appropriate. Researchers are encouraged to share data that will inform the field more broadly than may be feasible only through published studies.

IES acknowledges that there may be issues associated with data sharing when the data collected are proprietary (e.g., when a published curriculum is being evaluated). These should be clearly identified in the DMP in section IX. If findings are published after the grant period has ended, grantees are still required to adhere to their Data Management Plan.

Consider these questions:

* When will you make the data available?
* What is the expected date range of availability?

**VII. Methods of data access:**

[Enter content here and then remove the Guidance prior to submission]

**Guidance**:

IES is looking for 2 different things in this section: how you are managing ‘active data’ - data being collected and used during the project, and how you intend to share your datasets per the IES data sharing requirements.

**Active data**:

Include information about how and where the data will be stored during the project. This should include information about servers, their security, who has access, back-ups (ideally off-site), how non-electronic data will be handled, who is managing them (department, PI, IT group), etc. Be very aware of the confidentiality and security concerns that IES has. If your IRB requires specific measures to be employed to protect the active data, be sure to include them here.

**Archived data:**

Identify where you will be archiving your data for sharing. IES encourages the practice of uploading data to a publicly accessible repository to facilitate the use of data and the provision of access after the grant has ended. You will probably need to de-identify your data to share it in publicly-accessible repositories. Remember that you will need to provide access to the dataset(s) for at least ten years after the end of the project. While IES does not explicitly state that you have to maintain that dataset in a useable format, it is strongly implied.

IES acknowledges that there are several methods to share data. These include:

* The investigator taking on the responsibility for data sharing, which may involve making data available to the requestor through a variety of means, including their institutional, departmental or personal website. Succession plans need to be in place if the PI leaves the project or institution.
* Use of a data repository, also known as a data archive. Repositories can be particularly attractive for investigators concerned about a large volume of requests, or vetting requests. They usually require extensive documentation for the dataset(s).
* Researchers can use a data enclave (restricted-access) when datasets cannot be distributed to the general public, for example, because of participant confidentiality concerns, third-party licensing, or use agreements that prohibit redistribution.
* Use of some combination of these methods. A mixed method for data sharing that allows for more than one version of the dataset and provides different levels of access depending on the version. An example would be an anonymized version in a public repository and a non-anonymized version in a restricted-access repository.

Consider the following:

* Will you share data via a repository, handle requests directly or use another mechanism?
* If your method of sharing is with an archive, which archive/repository/database have you identified as a place to deposit data?
* What procedures does your intended long-term data storage facility have in place for preservation and backup?
* What is the long-term strategy for maintaining, curating and archiving the data?
* What metadata and documentation will be submitted alongside the data or created on deposit or transformation in order to make the data reusable?
* What related information will be deposited? Will this include supplemental files?
* What costs will your selected sharing method charge (Be sure to include these in your budget)?

The major advantage of depositing the dataset(s) in a repository such as ICPSR or OpenICPSR is that they assume the responsibility to manage, curate, and transform the data over time. Sometimes a fee is involved, so identifying the intended location at this stage enables you to include the costs in your budget, which IES encourages. Both ICPSR and OpenICPSR allow you to restrict access to the dataset, enabling you to submit a non-anonymized dataset. Ideally you should place it in two locations, in original and de-identified formats (the mixed-method referenced above).

ICPSR requires a strong social science component, so not all education dataset(s) will be accepted. OpenICPSR accepts all data types. Because UVa is a member of ICPSR, costs are reduced. You can also consider locations such as Figshare, the Dataverse at Harvard’s IQSS, and the Odum Institute Dataverse at UNC-Chapel Hill, or a Secure Dataverse.

You should also deposit the de-identified dataset and data documentation in Libra Data, UVa’s IR. Be aware that it is a non-curated option. If you are already preparing the data for another publicly-accessible repository, it is easy to submit the same dataset to Libra. The following is a suggested statement to include in your DMP.

“As advised by University of Virginia Library Research Data Services group, I plan on also depositing my research data in the UVA institutional repository – Libra Data, UVa’s local instance of Dataverse. Libra Data assigns DOIs to each dataset, and uses standard data management best practices, including security and backup procedures, to provide a reasonable assurance that files will remain retrievable over time. I will submit the necessary metadata and other resources to make my data accessible for future users. In accordance with University of Virginia policy, the data will be preserved for a minimum of five years upon completion of the project. The Libra backup plan provides for data redundancy including off-site storage.”

**VIII. Data sharing agreement:**

[Enter content here and then remove the Guidance prior to submission]

Discuss whether or not a data sharing agreement that specifies conditions under which the data will be shared is required.

**Guidance**:

Data agreements are a way to impose limitations on users. They are not required for de-identified publicly available datasets. If your dataset is going to be shared in a non-anonymized format, then an agreement may be necessary. The agreement should provide rules governing access to the dataset(s). They usually ask for information about the researcher(s), their intended research, data storage formats and technologies, and approval from their IRB. After signing, the researcher would be allowed access to the data. The methods of the access are determined by the repository that holds the dataset(s). E.g.: ICPSR requires users of the restricted-access data to complete a data agreement form.

Consider the following:

* Will any permission restrictions need to be placed on the data?
* With whom will you share the data, and under what conditions?
* Will a data sharing agreement be required?

**IX. Circumstances that may prevent some data from being made accessible:**

[Enter content here and then remove the Guidance prior to submission]

### Discuss any circumstances that prevent all or some of the data from being shared. This includes data that may fall under multiple statutes and, hence, must meet the confidentiality requirements for each applicable statute (e.g., data covered by Common Rule for Protection of Human Subjects, FERPA and HIPAA).

**Guidance**:

There may be circumstances, such as when a state or district will not allow student data to be released, where investigators will not be able to share their complete data set. IES expects primary data collected by the project or extant data obtained from a private source to be shared. In cases where access to the data cannot be provided directly, researchers may be able to provide access to the information used to create the datasets so that other researchers can contact the original source of the data to obtain a similar dataset. Alternatively, the dataset may be given back to the source (e.g., school district or state) and researchers can contact the source to obtain access. Every effort should be made to identify a way to share as much of the final data as possible.

IES recognizes that data sharing may be complicated or limited by institutional policies, local Institutional Review Board (IRB) rules, as well as state and federal laws and regulations that address issues of the rights and privacy of human subjects.

Generally, datasets can be de-identified or anonymized sufficiently to protect the confidentiality and privacy of study participants. Circumstances may occur when an IRB Protocol imposes very strict requirements.

In cases where data cannot be free of identifiers or when identifiers are important for linking datasets, then investigators should consider restrictions on data sharing, as provided by data archives or enclaves described in section VII.

Any restrictions on data sharing, such as a delay of disclosing proprietary data, should be presented. If researchers believe that full data sharing is not possible, they must provide a written rationale in their DSP.

Some institutions have tried to restrict data sharing by relying on data ownership policies. IES reminds all parties that: “*Data used in federally funded grants are considered to be products of the grant*.” EDGAR 34 CFR.74.36(C) states that: "*The Federal Government has the right to: (1) Obtain, reproduce, publish or otherwise use the data first produced under an award; and (2) Authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes*."