

## Frequently Asked Questions (FAQs): Managing and Sharing PEER Data and Publications

**Question:** Will someone else use my data? If so, might that person or organization misunderstand or mis-use my data?

**Answer:** Yes, one of the goals of research data management is to facilitate future re-use. There are several steps you can take to ensure that others will understand the data and use it properly. Above all, you should thoroughly document your data and your procedures:

- Provide **metadata** that explains your research data.
  - Create codebooks for your data.
  - Get informed consent to collect data (where required).
  - Document the methodologies used to collect your data.
  - Include source code and analysis workflow! (if applicable to your project).
- 

**Question:** How do I ensure that I properly curate and store my data and yet avoid getting scooped?

**Answer:** PEER's Development Data program recognizes that researchers need to begin the curation and long-term management process before they are ready to publish the data to a wide audience. USAID's Development Data Policy and Public Access Plan, in line with other Federal Agency funding policies, provide the opportunity to apply for and set an **embargo** for your data submissions,

You can work with your chosen repository to implement the embargo and to set important sharing and publication procedures. **Data licenses** and data user agreements ensure you get credit for your work.

---

**Question:** Sharing data is a lot of work. What is the benefit for my work?

**Answer:** The benefits to you include:

- Getting credit for your hard work (e.g, citations for your data!)
- Potential for increased probability of journal publication acceptance (see article JAMA. 2018;319(4):410. doi:10.1001/jama.2017.20650)
- Demonstrating a concrete commitment to high-quality research results and procedures and to scientific reproducibility (i.e., a trusted researcher with transparent, trusted results!)
- Increase the potential for future collaborations. If you demonstrate a commitment to data curation and data publication, then other researchers will become familiar with your high-quality data. You can more easily engage others around the world in meaningful, productive collaborations.
- Embracing new best practices at the cutting edge of the digital research world

The benefits to future researchers (including YOU as a future researcher) include:

- Available, accessible data to verify past results
  - Available, accessible data to jump-start a new project (e.g., build off past results quickly!)
  - Available, accessible data for even larger-scale, interdisciplinary or comparative research
  - Available, accessible data to pioneer new, innovative ideas at the edge of scientific research (e.g., training data for artificial intelligence [neural network] initiatives, growing geospatial and real time traffic data for automated vehicle research)
- 

**Question:** How will I ensure that I get credit and that my data are properly shared and published?

**Answer:**

1. Get a data citation! Good data repositories will have requirements for data citation.
  2. Ensure that the citation uses a persistent identifier (e.g., a DOI).
- 

**Question:** How much does it cost to upload data to a repository?

**Answer:** Many of the most frequently used repositories allow researchers to upload data for free. Some repositories charge a fee, but they often waive it for researchers in low and middle income countries. You can usually find out about price in the repository's terms of service or data submission directions. USAID's [Development Data Library](#) (DDL) is always free.

---

**Question:** What do you mean by Personally Identifiable Information (PII)?

**Answer:** Personally Identifiable Information is information that could be used to identify an individual or item by either direct or indirect means. Direct PII immediately identifies something, and indirect PII is information that can be combined or linked with other information that could identify something. Researchers with these types of data must ensure the privacy of every individual in order to prevent embarrassment and in some cases harm to an individual, artifact of interest, or endangered animal.

Examples of *direct* PII include:

- Full name
- Home address
- Email address
- Social security number/national identification number
- Passport number
- Driver's license number
- Credit card numbers

- Date of birth
- Telephone number
- Log in details
- Latitude and longitude of house, field/plot, business etc.

Examples of *indirect* or *linkable* PII include:

- First or last name (if common)
  - County, province, village, city, postcode or similar official administrative unit
  - Gender
  - Race
  - Non-specific age (e.g. 30-40 instead of 30)
  - Job position and workplace, employment status
  - Number of people residing in the household (if unusually large)
  - Housing materials (e.g. thatch, tile, shingle roof, adobe, mud & wattle)
  - Unique things about the household that can be traced back to it (5 bicycles, 10 sheep, 20 rabbits etc.)
- 

**Question:** I work with sensitive information and/or vulnerable populations or resources. I am concerned about the safety and privacy of the people I work with/the security of the resources I work with. How can I ensure that my research data are protected?

**Answer:**

1. PLAN! From the beginning of your research project, you should assess the risk that your research information and data can present to research subjects/environment/resources.

- plan for data collection with protections in mind
- plan for any necessary informed consent
- plan for proper short-term data storage, processing, and analysis
- plan to submit proper documentation to a repository and (if necessary) plan to work with the repository to submit protected copies of data

2. Assess the risks and benefits during the course of your project

- MAKE SURE TO CONSIDER YOUR COUNTRY GOVERNMENT'S LEGAL/ETHICAL REQUIREMENTS
- document the risks that collected/generated data may present

3. Use best practice risk mitigation techniques and procedures

- protect sensitive data and information - use passwords, use access control, use strong encryption to store and transmit digital data. For physical records use access controls and a secure location.
- use statistical disclosure limitation techniques to create de-identified datasets
- create pseudonyms and codes for participants
- redact or use statistics to de-identify

4. Provide a repository with the best information to ensure data protection

- a risk assessment (or disclosure analysis plan) that identifies risks, documents how you mitigated risks, and makes recommendations about protecting and sharing data.
  - Include informed consent documents
  - You may opt to provide two data sets: 1. a dataset of micro-data for restricted access. 2. a data set with mitigations, such as redacted data, and data protections that can be shared more widely.
- 

**Question:** What happens if I don't publish by the end of my award? Do I still upload data? How will I avoid being scooped?

AND

Is there a time lag requirement between data collection and data submission?

**Answer:** The USAID Open Data Policy says that Principal Investigators need to submit datasets produced with USAID funding that are used (or of sufficient quality) to produce an Intellectual Work. This does not have to be a publication but that is the most common scenario. Essentially researchers need to put the data on USAID's Development Data Library (DDL) or on a sector-specific repository linked to the the DDL. This should be done at the same time as the intellectual work is accepted for publication; if there is no publication, researchers have 3 months after the PEER award ends to put any data of sufficient quality for intellectual work on the DDL. The researcher can embargo the data, or store privately, for up to 1 year after the PEER award ends to get more time to complete a publication or other intellectual work. This also ensure you don't get scooped as you finish your publication.

Here is the legal language:

Structured Data: Structured datasets resulting from USAID-funded research must be registered and/or submitted to the DDL no later than 90 calendar days after award completion. For research datasets underlying peer-reviewed publications, the structured datasets must be registered and/or submitted to the DDL no later than 15 working days after the date printed on the publication acceptance notification. The peer-reviewed publication and the underlying dataset will be made publicly available after the stipulated embargo period. USAID may embargo, or temporarily withhold from public release a dataset resulting from federally funded research while the dataset is the subject of a pending publication or pending patent application for a period not to exceed 12 months after the award completion date.

---

**Question:** What websites accept data paper uploads?

**Answer:** Data papers are basically long-form descriptions of metadata, and sometimes they are peer-reviewed. Websites that host data papers include F1000Research, PLOS One, Data, and Scientific Data. These are often sector-specific so it is good to make sure your data

matches the most popular sectors on these websites. The USAID DDL is also a good, free place to store a data paper.

---

**Question:** Does PEER need to review a manuscript before submission to a journal by an investigator?

**Answer:** No, you do not need to get your manuscripts approved by PEER.

---

**Question:** What kind of data shall be shared? Metadata, raw data?

**Answer:** Raw data that is an input for a journal article or intellectual work should be shared. The raw data should be accompanied by a metadata file that explains the raw data.

---

**Question:** Is there a standard format for data repository?

**Answer:** The format of both your raw data and your metadata may depend on the repository you choose to upload to. In some fields there may be customs and conventions for data formatting that may be different than others, as well. Check the data submission guidelines before you upload your data and metadata.

For example, many disciplines have recommended metadata standards (a **metadata standard** is a format or structure for organizing metadata that is widely used by a research community):

Selected Recommendations:

- Social, Behavioral, and Economic Sciences
  - Consider using the [DDI](#) standard
  - The tool [Nesstar Publisher](#) (Windows only) can help manage DDI metadata
- Earth, Environmental, and Ecological Sciences
  - Consider using the [EML](#) standard
  - The tool [Morpho](#) can help manage EML metadata
- Geospatial data
  - You will almost surely use the [ISO 19115](#) standard
  - Whatever tool you are currently using to manage your data will probably already use this standard (ArgGIS, QGIS, etc.)
- Other
  - Check [this list](#) to get ideas for an appropriate metadata standard
  - Ask your USG partner!
  - Ask me

---

**Question:** Can university repositories be used as well?

**Answer:** University repositories may be used depending on how they are formatted and maintained. There is no official USAID guidance on repository requirements, but a good repository:

1. Has a long term data management plan
2. Has a commitment to preserving your data or preservation plan (in mission statement or terms of service)
3. Provides a persistent identifier (ideally DOI) and a URL to the data
4. Has data citation guidelines to ensure you get credit for your work.
5. Makes data widely available and free for public use with clear licensing and use terms, [CC-0](#) or [CC-BY](#) licenses recommended
6. Allows wide sharing of metadata

You can go to a site like the Registry of Research Data Repositories (<https://www.re3data.org/>) to check these details. For further guidance, see the [Development Data Overview](#).

---

**Question:** Just want to know if any field notes are raw data? And should we share those data as well?

**Answer:** Field notes are data, usually unstructured data. They should be shared if they are important inputs for an intellectual work such as a journal article. Not all repositories can accept unstructured data, however. For this reason, you may include this type of data as an attachment or supplement to a journal article. You can also store it on [USAID's Development Experience Clearinghouse \(DEC\)](#).

---

**Question:** Is IPR (intellectual property right) also applicable for the data shared as required by the Public Access Plan?

**Answer:**

Yes! The researcher always owns their output, but as a funder, USAID has a license to it. To exercise this license, the data need to be shared with USAID via the DDL, but the data do not need to be posted publicly. For more, see the "Restrictions on Public Disclosure" section of the Public Access Plan, which states:

Although submission of the research data to the DDL or an approved third party repository will be required, the public release of research data will be governed by the six principled exceptions outlined in OMB Bulletin 12-01 "Guidance on

Collection of U.S. Foreign Assistance Data” outlined below, which are also reflected in the February 2013 OSTP memo:

- A. When public disclosure threatens national security interests;
- B. When public disclosure is likely to jeopardize the personal safety of U.S. personnel or recipients of U.S. resources;
- C. When public disclosure would interfere with the Agency's ability to effectively discharge its ongoing responsibilities in foreign assistance activities;
- D. When there are legal constraints on the disclosure of business or proprietary information of non-governmental organizations, contractors, or private sector clients;
- E. When the laws or regulations of a recipient country apply to a bilateral agreement and restrict access to information; or
- F. When data reveal private information about individuals that must be kept confidential consistent with ethical guidelines and federal regulations.

Exceptions also include the following, laid out in 2 CFR 200.315:

- I. “Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and
- II. Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.”

USAID recognizes that some information must be handled carefully to avoid negative consequences for individuals. As such, the procedures here apply to other Agency processes aimed at protecting individuals from release of potentially harmful information. This effort includes incorporating a review process, noted above, to screen information before release to the public, as well as requirements to prevent certain types of information from being submitted in the first place. Specifically, as outlined in ADS 579 and ADS 540, financially sensitive information or personally identifiable information (PII) must be removed prior to submission.