NIH Immersion Workshop Breakout Rooms

# Breakout 1: Assessing Policy Readiness (~15 min)

Provide descriptions of tasks that have (1) been completed, (2) are incomplete, or (3) are necessary, but not feasible in the chart below. This is a reflective practice to map out capacity for **you and your institutions** and identify any action items to prepare for the upcoming policy. And although we do not have time to review this today, it will be helpful for you to go through the comprehensive [Policy Readiness Checklist for Librarians](https://osf.io/4e6wd) before the policy’s rollout.

## Instructions:

**Step 1 (1 minute)**: Add your name to one of the cells in column 2 (“Group Member(s) Name”). There can be more than one name per row, if needed.

**Step 2 (5 minutes)**: Jot down details across the row for your assigned question. Add details about completed work, work yet to be completed, and any limitations regarding the Policy and its requirements.

**Step 3 (9 minutes)**: **Your breakout room facilitator** will briefly review each question and ask for comments from the group members who contributed to that question, detailing completed work, work yet to be completed, and any limitations.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Group Member(s) Name | Completed | To Do  | Cannot Be Done / Limitations |
| What efforts have been made to become familiar with the details of the policy and its requirements?  |  |  |  |  |
| What synchronous education efforts have you or your institution made? |  |  |  |  |
| What asynchronous education efforts have you or your institution made? |  |  |  |  |
| Are there communication or outreach strategies in place to talk to faculty, researchers, and leadership? |  |  |  |  |
| Is there any infrastructure needed for addressing this new policy? (e.g., Institutional DMP platforms, adjusting of grant workflows, etc.) |  |  |  |  |
| Have institutional partners or necessary collaborators (e.g., OSR, IRB, etc.) been identified? What is the level of coordination among these groups? (See [Partners: People and Departments](https://osf.io/4e6wd)) |  |  |  |  |

# Breakout 2: Data Types in Repositories (~15 min)

## Instructions:

**Place your name next to one repository here.** There can be more than one name per repository, but make sure at least one person is listed for each.

1. [Data Sharing for Demographic Research](https://www.icpsr.umich.edu/web/pages/DSDR/index.html) —
2. [Vivli](https://vivli.org/) —
3. [dbGaP](https://www.ncbi.nlm.nih.gov/gap/) —

For your repository, try to figure out:

1. The data types that are accepted
2. Any standards required to upload data — standards for description and standards for data collection
3. Any other limitation on data submission

**After 7 minutes, your breakout room facilitator** will lead a discussion on each repository, asking for your input.

## Answer Key:

1. **Data Sharing for Demographic Research**
	1. Accepted Data types:
		1. Survey, demographic, census
	2. Standards required:
		1. No standards required, but detailed instructions on data description and organization. Those instructions do not seem to correspond to a specific standard
	3. Other limitations on data submission
		1. No
2. **Vivli**
	1. Accepted Data types
		1. Clinical trials data
	2. Standards required
		1. Use of CDISC data standard highly suggested. No standards strictly required, but outside of the CDIS suggestion, there are also detailed instructions on data description and organization
	3. Other limitations on data submission
		1. Fee for some submission; Review and approval of submissions required
3. **dbGaP**
	1. Accepted Data types
		1. Genomic data; Phenotypic data
	2. Standards required
		1. Detailed instructions and requirements for data description and organization, not tied to an external data standard (e.g., CDISC), but standardized across submission
	3. Limitations on data submission
		1. Review and approval of data submissions required;

# Breakout 3: DMP Assignments (~30 min)

## Instructions

Below is the text of an NIH Data Sharing Plan, which contains some, but not all, of the information needed for the new DMSP.

### Step 1 (15 minutes):

**Each attendee will:**

1. **Choose an element of the NIH DMSP**
2. **Review the sample NIH Data Sharing Plan below, and Identify all the text in the data sharing plan which would go into the element that you chose**
3. **Identify all the information needed for the NIH DMS Policy that is not in the sample data sharing plan.**

**Place your name in the table here to claim your element. There can be more than one name per section, but ensure that each section has at least one name.**

|  |  |
| --- | --- |
| **Group Member Name(s)** | **Section** |
|  | Element 1: Data Type |
|  | Element 2: Related Tools, Software, and/or Code |
|  | Element 3: Standards |
|  | Element 4: Data Preservation, Access, and Associated Timelines |
|  | Element 5: Access, Distribution, or Reuse Considerations |
|  | Element 6: Oversight of Data Management and Sharing |

* Below the plan is [a listing of DMSP elements](#_yu70ygeub873) for your reference.
* Feel free to use document comments for this work. Highlight the text you want to label and use the blue plus button to “add comment” (or right click the highlighted text to add comment). You can also respond to other people’s comments!

### **Step 2 (15 minutes):**

**Your breakout room facilitator** will review each section, asking for input from group members who worked on that section.

## Data Sharing Plan: Phonological aspects of word learning: Advantages of bilingualism and impacts of developmental language disorder: Data sharing plan

### What data will be shared:

We will share fully de-identified participant response data associated with the project at PDX Scholar, a digital open-data repository that meets the open-data requirements of federal research grants and is free for use by members of the Portland State University (PSU) research community. Data will be stored as binary files (eye tracking data), text files (choice response data), and transcriptions (spoken response data). Depending on obtainable permissions, and contingent on PSU IRB approval, we will release recordings of spoken responses (WAV files), demographic data (in Excel spreadsheets), and language and cognitive assessment scores (in Excel spreadsheets). We will include permissions for data sharing (with particular emphasis on WAV files) in IRB-approved consent forms. These file formats are the formats that will be originally used in the study. Some formats are automatically generated by software or hardware (e.g., binary files, text files, WAV files) and others are widely used (e.g., transcriptions, Excel spreadsheets). After study recruitment and data analyses are completed, we will further process some files (e.g., Excel spreadsheets) to make sure their organization is intuitive. We will consult with PSU Library staff who coordinate the PDX Scholar database to develop a metadata scheme, informed by community metadata standards. Data files for each study will be accompanied by text "readme" files describing assignment of participants to conditions. "Readme" files will also include additional documentation to enable secondary use of the data, including information on methods of data collection, definitions of column and variable names, units of measurement, data formats and file types, software programs used for data collection and processing, and instructions for how to cite and refer to the original study.

### Who will have access to the data:

I agree that data will be deposited and made available through PDX Scholar, which is an open-data repository for use by PSU-affiliated faculty, and that these data will be shared with investigators working under an institution with a Federal Wide Assurance (FWA) and could be used for secondary study purposes. Access to the data will not be restricted as long as the researcher meets this FWA criterion. A data-sharing agreement will not be required.

### Where will the data be available:

We will share participant response data associated with the project at PDX Scholar, an open-data repository. We have consulted with the coordinator of data-management services at PSU Library, Rick Mikulski, for recommendations on using PDX Scholar for data management, and he has encouraged us to continue consulting him on an ongoing basis.

### When will the data be available:

I agree to deposit outcome data into the PDX Scholar repository as soon as possible, but no later than within one year of the completion of the funded project period for the parent award, or upon acceptance of the data for publication, whichever is earlier.

### How will researchers locate and access the data:

I agree that I will identify where the data will be available and how to access the data in any publications and presentations that I author or co-author about these data, as well as acknowledge the repository and funding source in any publications and presentations. The PDX Scholar repository has policies and procedures in place that will provide data access to qualified researchers, fully consistent with NIH data sharing policies and applicable laws and regulations. The PSU Library staff who manage PDX Scholar include specialists in open-access resources, who I have already consulted about considerations for open-access publication and for disseminating research as widely as possible. Thus, I will continue to consult PSU librarians to ensure that persistent identifiers will be properly applied so that my data can be easily discovered online, and so that I can track citations and secondary uses of the data.

###

## Data Management and Sharing Plan Instructions

####

### Element 1: Data Type

##### A. 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.

#####  B. 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.

##### C. Metadata, other relevant data, and associated documentation:

Briefly list the 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.

### 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.*

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

##### A. 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)*).*

##### **B.** **How scientific data will be findable and identifiable:**

*Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.*

##### **C.** **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.*

### Element 5: Access, Distribution, or Reuse Considerations

##### A. 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 Asked Questions](https://sharing.nih.gov/faqs%23/data-management-and-sharing-policy.htm) for examples of justifiable reasons for limiting sharing of data.

##### B. 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).

##### C. 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).

### 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).

## Answer Key:

#### Element 1: Data Type

##### A. 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.*

* Data will be stored as binary files (eye tracking data), text files (choice response data), and transcriptions (spoken response data).
* These file formats are the formats that will be originally used in the study. Some formats are automatically generated by software or hardware (e.g., binary files, text files, WAV files) and others are widely used (e.g., transcriptions, Excel spreadsheets).
* After study recruitment and data analyses are completed, we will further process some files (e.g., Excel spreadsheets) to make sure their organization is intuitive.

#####  B. 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.*

* Depending on obtainable permissions, and contingent on PSU IRB approval, we will release recordings of spoken responses (WAV files), demographic data (in Excel spreadsheets), and language and cognitive assessment scores (in Excel spreadsheets).

##### C. Metadata, other relevant data, and associated documentation:

*Briefly list the 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.*

* Data files for each study will be accompanied by text "readme" files describing assignment of participants to conditions.
* We will consult with PSU Library staff who coordinate the PDX Scholar database to develop a metadata scheme, informed by community metadata standards.
* "Readme" files will also include additional documentation to enable secondary use of the data, including information on methods of data collection, definitions of column and variable names, units of measurement, data formats and file types, software programs used for data collection and processing, and instructions for how to cite and refer to the original study.

No estimates of amount of data

#### 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.*

* "Readme" files will also include additional documentation to enable secondary use of the data, including information on methods of data collection, definitions of column and variable names, units of measurement, data formats and file types, software programs used for data collection and processing, and instructions for how to cite and refer to the original study.

Mentions software, but gives no information is given about what that software is, and whether this refers to homegrown or licensed software.

#### 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.*

* We will consult with PSU Library staff who coordinate the PDX Scholar database to develop a metadata scheme, informed by community metadata standards.

No mention of which community metadata standards they will look to as they develop their metadata schema.

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

##### A. 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)*).*

* I agree that data will be deposited and made available through PDX Scholar, which is an open-data repository for use by PSU-affiliated faculty, and that these data will be shared with investigators working under an institution with a Federal Wide Assurance (FWA) and could be used for secondary study purposes.
* We will share participant response data associated with the project at PDX Scholar, an open-data repository.

##### **B.** **How scientific data will be findable and identifiable:**

*Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.*

* I agree that I will identify where the data will be available and how to access the data in any publications and presentations that I author or co-author about these data, as well as acknowledge the repository and funding source in any publications and presentations.
* The PSU Library staff who manage PDX Scholar include specialists in open-access resources, who I have already consulted about considerations for open-access publication and for disseminating research as widely as possible. Thus, I will continue to consult PSU librarians to ensure that persistent identifiers will be properly applied so that my data can be easily discovered online, and so that I can track citations and secondary uses of the data.

##### **C.** **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.*

* I agree to deposit outcome data into the PDX Scholar repository as soon as possible, but no later than within one year of the completion of the funded project period for the parent award, or upon acceptance of the data for publication, whichever is earlier.

Does not specify how long the data will be available.

#### Element 5: Access, Distribution, or Reuse Considerations

##### A. 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 Asked Questions*](https://sharing.nih.gov/faqs%23/data-management-and-sharing-policy.htm) *for examples of justifiable reasons for limiting sharing of data.*

* We will share fully de-identified participant response data associated with the project at PDX Scholar, a digital open-data repository that meets the open-data requirements of federal research grants and is free for use by members of the Portland State University (PSU) research community.
* Depending on obtainable permissions, and contingent on PSU IRB approval, we will release recordings of spoken responses (WAV files), demographic data (in Excel spreadsheets), and language and cognitive assessment scores (in Excel spreadsheets).

##### B. 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).*

* I agree that data will be deposited and made available through PDX Scholar, which is an open-data repository for use by PSU-affiliated faculty, and that these data will be shared with investigators working under an institution with a Federal Wide Assurance (FWA) and could be used for secondary study purposes. Access to the data will not be restricted as long as the researcher meets this FWA criterion. A data-sharing agreement will not be required.

##### C. 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).*

* We will include permissions for data sharing (with particular emphasis on WAV files) in IRB-approved consent forms.

#### 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).

* We have consulted with the coordinator of data-management services at PSU Library, Rick Mikulski, for recommendations on using PDX Scholar for data management, and he has encouraged us to continue consulting him on an ongoing basis.

This indicates who they will consult for data management, but not who will be responsible, how it will be managed and monitored, or the frequency of oversight.