Converting a resource sharing plan into a DMS Plan

Final draft: 2022-08-31
OSF Link: https://doi.org/10.17605/OSF.IO/UADXR

DATA MANAGEMENT AND SHARING PLAN

Based on NOT-OD-21-014: Supplemental Information to the NIH Policy for Data Management and Sharing:
Elements of an NIH Data Management and Sharing Plan

In two pages or less, describe your proposed approach to data management and sharing. If your program or data type has specific data sharing expectations (e.g., repository selection), that should be reflected in the Plan. Consider how your Plan will be consistent with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles and reflective of practices within specific research communities.

This template can also be used when researchers work with Arcus on their research. For example, suppose investigators plan to contribute their data to the Arcus Archives or are working in an Arcus lab and will contribute data back. In that case, they can use this template wholesale. Researchers should substitute directions in brackets with specifics related to their research project.

Project Name: The Feasibility of the XYZ Method in Patients
PI Name and Affiliation: John Doe, Children's Hospital of Philadelphia
Date Finalized: 2023-01-05
DOI: XX.XXXX/XXXXXX

Section 1: Data types

- Describe all the data types and approximate amounts of each type to be collected during the study
- Indicate which data will be preserved and shared and address security measures (if applicable)
- Include descriptions of metadata and if/where study protocols will be accessible

1. Data types. [List the type of data you are using]. [Describe the file formats from data collection through to analysis and finalized data.]

Converting a resource sharing plan into a DMS Plan

Data will be deposited into the Arcus Archives and will be made available by request. The Arcus Archives is the canonical repository for the data of Children's Hospital of Philadelphia (CHOP) Research Institute's research efforts.

The Archives aims to store research projects holistically, archiving data, contextual files, tools, and metadata so that data will be reproducible, reusable, and repurposable.

Protocols and details regarding instrument settings, data transformation, and analysis will be made available in the accompanying README document. These details will also be documented in standard csv manifest files accounting for the participants, data, and methods used in the research.

Section 2: Tools, software and/or code

- Include the tools and computer software that will be used
- Indicate if proprietary file formats will also be saved by non-proprietary means or if special software will be necessary for other users to access and reuse the data
- Get help from the technicians who are running the data collection and processing

2. Tools, software and/or code. [*State how you will collect data.*] [*State how you will analyze data.*]

An institutional instance of Box will be used for active data storage, which is HIPAA-compliant. For the proteomics data, resulting spectra will be converted to Mascot generic format (MGF) files using Proteome Discoverer v2.1.0.81. For the metabolomics data, raw structural information about GC-MS features will be obtained through spectral matching with the NIST 14 spectral library. Mass Profiler Professional software (Agilent) for GC-MS data and Analyst (AB Sciex) for LC-MS/MS will be used to assign peaks to raw ion chromatograms.

[*If applicable:*] no specialized tools or software will be needed to access or reuse the shared datasets, which will be available via request from Arcus. [*If depositing data to*

Converting a resource sharing plan into a DMS Plan

other repositories, e.g., dbGAP, please note those here).

Section 3: Data standards

- Look up the data standards for each type of data output
- Consult the NIH maintained list of common data elements
- If there is no standard, note this information and build your own standard with a data dictionary

3. Data standards. A data dictionary will be provided for the clinical dataset that defines column headers, units of measurement, and other pertinent metadata as necessary to understand and reuse the dataset. [*State standards used for each data type in the research project.*] Both repositories selected for the omics data streams comply with the accepted standards for their field.

Section 4: Data preservation, access, and timelines

- Provide repository name(s) and indicate how data will be findable
- Make sure to discuss access and distribution for all of the types of data generated in the study Indicate what level of data will be shared (raw, aggregate, de-identified etc.) and when and how long it will be shared (not all data needs to be shared at the same level)
- Avoid hyperlinks in the DMSP, save these for the RPPR submission

4. Data preservation, access, and timelines. Both raw and analyzed data along with any accompanying metadata will be contributed to the Arcus Archives for storage. Scripts and coding workflows will also be made available via Arcus. Arcus Archives data is preserved according to international digital archiving standards and utilizes a custom ingestion and processing workflow designed by Digital Archivists, Application Research Developers, and DevOps Engineers.

Metadata about research data and participants in the Archives are available for browsing and mediated requests via the Arcus Cohort Discovery (ACD) tool and the Arcus Data Catalog. Arcus Archives data is retained indefinitely across geographically

Converting a resource sharing plan into a DMS Plan

disparate, secure, redundant, and monitored storage environments (Google Cloud Platform and Amazon Web Services) to prevent loss in the case of a catastrophic event. Data will be assigned a persistent unique identifier during the contribution process to the Arcus Archives.

[*State publicly available repositories holding data and their retention and access policies, if any. Note whether data is raw or analyzed, and whether it will be deidentified or not.*]

Section 5: Access, distribution, or reuse considerations

- Discuss limitations of the data sharing
- Address security concerns here, including how access will be controlled

5. Access, distribution, or reuse considerations. The Arcus platform was designed to maximize FAIR sharing of research data, including the data to be collected from this project. Contributors of data to the Arcus Archives agree to share their data under standard terms:

- Data shall be accessed in accordance with any governance terms (informed consent, protocols, grant agreements, data embargoes, etc.) under which the data was originally collected.
- Data shall be made available to authorized Arcus users without access restrictions.

While metadata about research data and participants in the Archives are available for browsing and mediated requests via the Arcus Cohort Discovery (ACD) tool and the Arcus Data Catalog, the data itself is available only upon request and review for compliance to existing governance terms for reuse. Review of requests for reuse allow delivery of identified, de-identified, limited, or scoped datasets from this project according to the reuse requester's research aims, protocols, etc.

[*Describe what data (raw, analyzed, clinical) will be made available through other repositories aside from Arcus. Note how or if that data will be deidentified and how others will access or request to access that data.*]

Converting a resource sharing plan into a DMS Plan

Section 6: Oversight

- Oversight will usually be the responsibility of the PI
- Oversight includes revising the DMSP and adhering to submission deadlines for sharing data
- All members of the team should have training on the DMSP

6. Oversight. The PI of the proposal will make the plan available to all personnel involved in the project. The PI will be responsible for ensuring faithful adherence to the DMS Plan and revising the plan annually, as the research project evolves.