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 data management and sharing approach. If your program or data type has specific data-sharing expectations (e.g., repository selection), the Plan should reflect that. 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

---

Children's Hospital of Philadelphia®
Department of Biomedical & Health Informatics

Converting a Resource Sharing Plan into a DMS Plan

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

Data will be deposited into the Arcus Archives and made available upon request. The Arcus Archives is the canonical repository for the data of the Children's Hospital of Philadelphia (CHOP) Research Institute's research efforts. Deidentified clinical data with appropriate consent will be shared for public access in an Arcus organization within the Harvard Dataverse. Harvard Dataverse is a nonprofit, generalist data repository built with the open-source Dataverse Project software and affiliated with Harvard University's Institute for Quantitative Social Science and Harvard Library.

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 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.

[If applicable for tabular data]: PI will ensure that tabular data files are ingested properly by the Dataverse software. To best support archival preservation, Harvard Dataverse stores the raw data content extracted from successfully ingested tabular data files in plain text, TAB-delimited files. The metadata information that describes this content is stored separately in a relational database so that it can be accessed efficiently by the application. For archival preservation, it can be exported in plain text XML files using a standardized, open DDI Codebook format.

Children's Hospital of Philadelphia®

Department of Biomedical & Health Informatics

Converting a Resource Sharing Plan into a DMS Plan

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 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.

Children's Hospital of Philadelphia®
Department of Biomedical & Health Informatics

Converting a Resource Sharing Plan into a DMS Plan

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 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.*]

Children's Hospital of Philadelphia®
Department of Biomedical & Health Informatics

Converting a Resource Sharing Plan into a DMS Plan

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. [*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.*]

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 allows delivery of identified, de-identified, limited, or scoped datasets from this project according to the reuse requester's research aims, protocols, etc.

Deidentified clinical data with appropriate consent will be shared for public access in an Arcus organization within the Harvard Dataverse. Before sharing, data will be reviewed for identifiers by Arcus data and privacy analysts. Before downloading data, users must agree to a Data Use Agreement specifying custom terms of access and reuse and provide name, institution, and contact information.

[*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.*]

Children's Hospital of Philadelphia®
Department of Biomedical & Health Informatics

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 ensure faithful adherence to the DMS Plan and revise the plan annually as the research project evolves.