Arcus
Data Contribution Guide

Table of Contents

Introduction

In this document, you will learn about contributing data to Arcus. Arcus archivists can preserve your data according to archival standards while facilitating data sharing. CHOP Researchers will learn about your work and explore collaborations with you in a controlled environment that protects your research and patient privacy.

By contributing your research to Arcus, it will become part of the Arcus Archives, the canonical home for the research data of the Research Institute. Data in the Archives powers the discovery tools of Arcus and your research will, as appropriate, be discoverable in the Arcus Cohort Discovery Tool, the Arcus Data Repository, and the Arcus Data Catalog. Data deposited in the Arcus Archives can be further analyzed in an Arcus Computational Lab.

Contributing your data to Arcus also means that Arcus will steward the data as part of its long-term preservation. The responsibilities involved will be documented in a Data Contribution Agreement (DCA) between Arcus and the PI. The DCA will also define access rights to the data, ensuring the contributor continued access and defining the parameters by which other CHOP researchers may access the data.

Getting Started

To get started, provide the Arcus Library Science team with key information about the research and the data by completing the Arcus Data Contribution Intake form. Filling out this form allows us to set a timeframe for ingesting your data based on the complexity and organization of your project. Your request will be reviewed, and the Library Science team will reply with next steps.

The Library Science team will work to understand your research and requirements and determine how best to preserve your data and ensure reusability. If necessary and depending on the nature of the research and the data, the Library Science team will include technical, privacy, and subject matter experts as part of the process.

If you have any questions, please reach out to the Library Science team at: arcus-support@chop.edu.

What about data organization and management? Is my data ready for Arcus?

We are always happy to meet with researchers at any phase of their research effort to offer a data management consultation. We can help set up a project template file directory structure for storing data, recommend metadata best practices, and provide data dictionary examples. The Arcus Education team supports various learning modalities, including 1:1 consulting, group workshops, office hours, asynchronous self-guided education, and user group meetings. Their website portal is a great resource on how to incorporate reproducible techniques into your existing research.

Research Data Management(RDM) resources can be found here.

Contributing Data to Arcus

Why would I contribute data to Arcus?   It will support new research at CHOP.

One of the primary functions of Arcus is to link otherwise disparate data sources to enable new opportunities for CHOP researchers and to accelerate the path to discovery. Contributing data to Arcus allows us to achieve this goal in that research data sets can be discovered along with and linked to other research data sets, clinical data sources, and computational tools.

Your data will be preserved long-term.

Arcus 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. Your data will be replicated across geographically disparate, secure, monitored storage environments to prevent loss in the case of a catastrophic event.

Your data will be reproducible, reusable, and repurposable.

Digital preservation is more than file transfers and secure storage, it's about making sure that what's preserved is understandable and reusable in the future. To that end, Arcus archives more than just the data, it preserves the software, tools, code, and metadata that were central to the research. The Arcus Library Science team will work with you to make sure your data is structured, organized, and described so that whether it's you looking to revisit your data, or other researchers looking to answer new questions, it will be easy to work with.

Can I contribute research that is still in progress?

Yes, we are eager and willing to archive fledgling, ongoing, and concluded research efforts. We have designed the data contribution process to be iterative and to complement the diverse needs of the research institute's user community. Some Arcus projects that are hosted on or dependent on other Arcus or DBHi services (registry or cohort creation, for example) also have a data contribution component that is part and parcel of this research effort.

I haven't started my research yet, but I want to contribute.

How can I be proactive about archiving my research effort?

We are always happy to meet with researchers at any phase of their research effort to offer a data management consultation. We can help set up a project template file directory structure for storing data and recommend metadata and organization best practices. The Arcus Education team supports various learning modalities, including 1:1 consulting, group workshops, office hours, asynchronous self-guided education, and user group meetings. Their website portal is a great resource on how to incorporate reproducible techniques into your existing research. Structuring a research effort around these principles leads to greater success in the pathway to publication and to the public dissemination of data.

Privacy & Regulatory Considerations

All contributions will be assessed for potential privacy and regulatory issues prior to ingestion. If needed, the Arcus Privacy Analyst will review consent forms, IRB protocols, and other relevant documents.

Can I still contribute to Arcus if my data contains identifiers?

Yes. Arcus can accept a full range of Data Contributions, from those that are completely de-identified to those that are fully identified and including ones that must remain coded. For coded data sets, Arcus offers honest broker services so that data retains its linkages without violating the privacy concerns of patients and researchers. In these cases, we strongly suggest contributed data includes a participant list and other underlying crosswalk information so that privacy can be retained while still allowing for requisite linking across multiple data sources.

Do I need to modify my IRB protocol or exemption to contribute to Arcus?

You do not need to modify your protocol for us to see your data; Arcus is overseen by its own CHOP IRB protocol which gives us permission to work with your data and archive it as a Contribution. Anyone listed as Arcus study personnel can steward identified data sets.

However, if your protocol explicitly mentions a data management system (e.g., REDCap) and you now will be storing some or all of your data in Arcus in order to fulfill the operational requirements of a Scientific Project, in addition to or in place of the other system, the IRB will require an Amendment with an explanation of your relationship to Arcus.

My data has special restrictions or requirements that must be observed; how will you take note of them?

Data requiring additional restrictions or conditions is customized through each researcher's Data Contribution Agreement. A Data Contribution Agreement is an agreement between an Arcus Contributor and Arcus library science staff that outlines a Contributor's role in contributing Archival Data to Arcus and allowing for the sharing and use of such Archival Data.

Data Eligibility

Data generated in the course of CHOP research will be archived in approved formats, including, but not limited to:

- Genomic data
- Clinical data specifically designated for research use
- Research datasets (raw, processed, analysis, reporting)
- Audiovisual files (images, video)
- Representative innovative technology
- Representative research datasets produced using data provided by Arcus
- Data resulting from animal studies

In addition, data and files that provide valuable context and support reusability will also be archived. This includes, but is not limited to:

- Project documentation (IRB data, descriptive metadata, administrative information, grant reports, manuals, patient questionnaires)
- Databases, spreadsheets, crosswalks
- Software, scripts, code, logs
- Taxonomies and ontologies
- Cohort definitions
- Readme files and file directory ownership
- Papers/publications

Data Access

Who can access my data? What if I want to share my data, but only in limited ways?

Arcus is a controlled environment for discovering and accessing data and all requests to work with data will be vetted according to the nature of the request and the limitations of the data in question. As a contributor you will be able to select from a set of access conditions for your data, including embargos and requests for collaboration prior to access. Your choices will be recorded in your Arcus Data Contribution Agreement, which will also include requirements for academic attribution of shared data. Arcus is focused on the professional needs of scientists for publishing, authorship, and attribution and data requestors will be required to abide by those access and attribution conditions. However, contribution to Arcus does require a commitment to data sharing, whether in the future or as part of collaborations. Absolute embargoes and restrictions to access are not available.

Will I be able to access and download the data I contribute?

As an Arcus contributor you and your approved study staff will be able to access and download contributed data at any time. You'll also be able to download de-identified data to support publishing or mandated data sharing.

Contact Arcus

If you would like to contribute data to Arcus, please contact us at arcus-support@chop.edu.

If you have any questions, please reach out to the Library Science team at arcus-support@chop.edu.