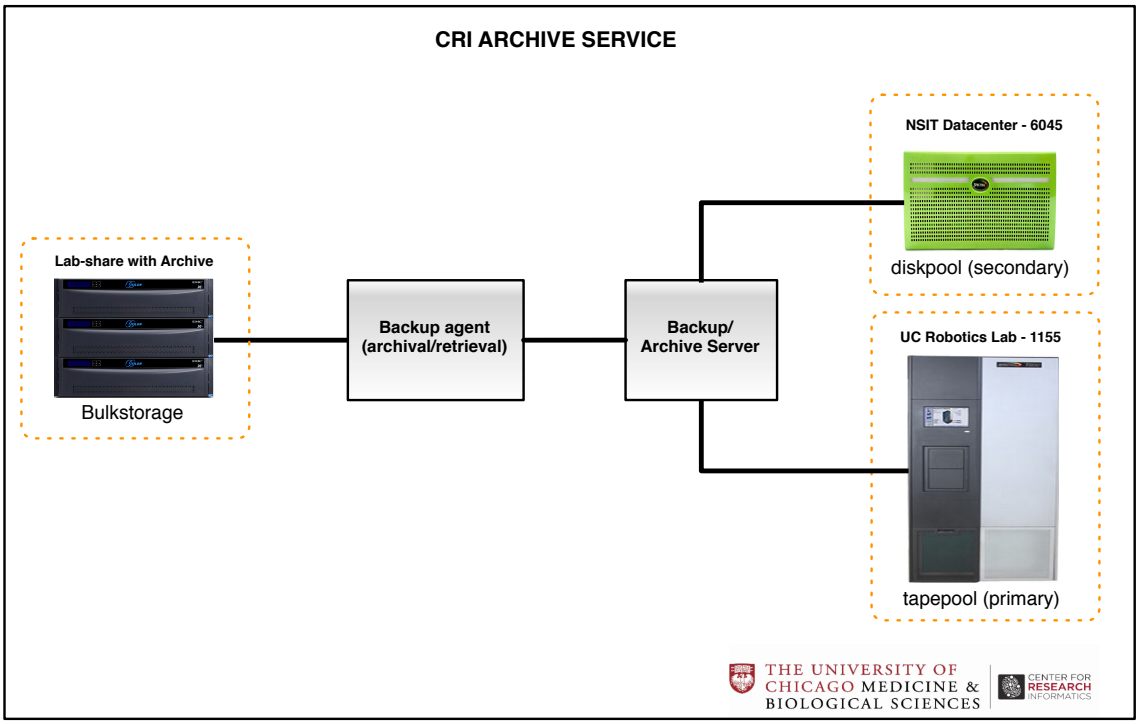
THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES
Center for Research Informatics

CRI Archival Data Storage
USER GUIDE

OVERVIEW

The CRI offers a storage service to BSD labs, called *Archive Storage*, designed for long-term storage of research datasets. This service is available only to users with existing labshares and must be requested separately. Data that users wish to archive is placed in an existing directory under **/group/<lab-share>/archive/**. The data in this archive directory will then be moved once a month by a CRI administrator to our archive storage, which will store the data in two copies on different media types located in different data centers. Once the dataset is successfully transferred to the archive storage system, all files are removed from the labshare, leaving only the directory structure (for ease of retrieval).



WHAT IS DATA ARCHIVING?

Data archiving is an important component of *Data Management*. Data archiving is the process of identifying and transferring data sets out of the primary production storage systems and into long-term archival storage systems. Specialized archival systems are able to store information more effectively in terms of cost and resources.

WHO NEEDS THIS SERVICE?

Most grant agencies require some type of data management plan to assure the grant agency that the data produced using the grant funds will be sufficiently protected. Data use agreements may also require data management plans to ensure that datasets and derived data products are well-protected. Other use cases include storing data that has been used for publications or that for any other reason will need to be kept around for a long time.

Data archiving helps to save on expensive storage by moving data to a tape backup solution at predefined intervals. This is an add-on service for users of the CRI's data storage solution (bulk storage labshares) and is set up within your share. Your frequently accessed data remains accessible and only archived data gets moved to tape. When your data is archived, only an administrator can retrieve it when requested. Frequently used or accessed data should not be archived.

THE ARCHIVE PROCESS

Data is archived using Tivoli Storage Manager Archiving Solution. The Archive Process selects data from a source and copies that data to a destination at predefined intervals. For your convenience, the archive directory is set up within your existing CRI labshare when you request this service; you simply move data to the archive directory. When the data is archived, it is removed from the archive directory, leaving the directory structure in place for reference. You will be provided with instructions on how to organize data in the archive directory to ensure your data can be retrieved quickly.

WHERE IS ARCHIVE DATA STORED?

Archived data is stored in two separate data centers. One copy is stored on tape in a secure and encrypted high-performance tape library and the other on a disk-based storage system optimized for archiving purposes. Both systems are managed by the Center for Research Informatics IT Infrastructure and Operations Team.

In order to comply with regulatory requirements for data retention, such as ensuring your data is securely discarded after a specified period of time, data will be stored in predefined retention directories. The data will be purged when it reaches the end of the retention period.

HOW MUCH DOES IT COST?

The first invoice is processed based on an initial archive size estimate. The following invoices are based on actual usage. The cost is $0.125/GB per year after the first 1 TB, which is free.

Size	GB/Year Cost	Annual Cost
Archive Storage (1 TB)	$0.125	$125

REQUESTING THIS SERVICE

To request CRI Data Archival Storage, fill out and submit the Request Research Data Archiving form. The form will include the specifications for the type of data to be archived, whether data contains sensitive information such as PHI, frequency of archives (one-time or continuous), data retention period, and existing share information. We will send you an invoice and when payment has been processed, the archive service will be set up and you will be provided with instructions on how to organize your data for archiving.

REQUESTING ARCHIVED DATA

To retrieve archived data, users must submit a Data Retrieval Request to the CRI. Fill out and submit the Request Research Data Retrieval form. If you need to retrieve a specific file or directory, you will need to know where it was stored. The organized archive directory structure will help you determine where the data was originally stored. (Refer to "Organizing Data to be Archived," below.) Requested data will be restored to **/group/<lab-share>/retrieve/<path-to-data-requested>/.** Retrieved data should be moved out of the retrieve directory to a different location on the share to prevent it from being overwritten during future retrieve operations. Depending on the size of the dataset to be retrieved, retrieval will take 1-3 business days.

ORGANIZING DATA TO BE ARCHIVED

Depending on your archive data retention requirements, directories will be created as below:

/group/<lab-share>/archive/5YR/
/group/<lab-share>/archive/3YR/

The directory/folder structure for data in the archive directory will be maintained after files are moved to archive storage for ease of data retrieval if needed. Do not delete the directories. Keeping the directories will help ensure that retrievals are processed faster, and you can point us to the location of the data before it was archived.

We recommend compressing your datasets before moving them to the archive directory.

ROLES

Title	Role
Principal Investigator (PI)/Lab Manager	Lab owner who approves share access
Technical Contact	Technical contact in charge of lab operations
Billing Contact	Processes payment for storage space

IMPROPER USE

In the event that the CRI determines that inappropriate use of storage is impacting other customers and/or the CRI's ability to manage the storage system, the CRI has the right to terminate the service by providing two weeks' notice to the identified customer contact.