

Centre for Health Informatics
Standard Operating Procedures for Data Access, Storage, and Computing

Effective Date: September 9, 2020	Version: 9.0
	Revision Date: September 3, 2020

Table of Contents

1.0	<i>Purpose:</i>	3
2.0	<i>Scope:</i>	3
3.0	<i>Definitions:</i>	3
4.0	<i>Procedure:</i>	4
4.1	Prerequisites	4
4.1.1	University of Calgary IT account	4
4.1.2	Multi-Factor Authentication	4
4.2	Data Classification	4
4.3	Data Request and Release	5
4.3.1	Obtain CHREB Ethics Approval.....	6
4.3.2	Submit Data Request to AbSPORU	6
4.3.3	Obtain AHS Research Operations Approval	6
4.4	Data Extraction and Linkage	7
4.5	Data Storage	7
4.5.1	Academic File Share	7
4.5.2	OneDrive for Business.....	7
4.5.3	Secure Computing Data Storage (SCDS)	8
4.6	Data Computing	8
4.6.1	Academic File Share	9
4.6.2	Advanced Research Computing (ARC)	9
4.6.3	Data Analysis Tool (DAT).....	9
4.7	Data Sharing	11
4.7.1	OneDrive for Business.....	11
4.7.2	ShareFile	11
5.0	<i>Appendix:</i>	13

1.0 Purpose: To provide University of Calgary's Cumming School of Medicine researchers and affiliates with a transparent, legal, secure process to access to health data resources and analytics via Centre for Health Informatics (CHI).

2.0 Scope: To outline the established procedures for requesting, accessing, storing, and computing health data resources via CHI for academic researchers and affiliates. Respective stages and responsibilities will be outlined.

3.0 Definitions: In this procedure,

- a) Alberta Health Services (AHS) is the health authority for the province of Alberta, Canada. AHS retains control and custody of the health data collected for the province and supports adequate data life cycle management and access provisioning activities provided by the University of Calgary.
- b) Alberta SPOR SUPPORT Unit (AbSPORU) is a specialized virtual, multidisciplinary research service team that helps researchers who conduct patient-oriented research with access to data platforms and services, tools, methods, training, and connections. SPOR Data Platform members, who are AHS employees with direct access to AHS Enterprise Data Warehouse, can facilitate data extraction and linkage for researchers.
- c) Centre for Health Informatics (CHI) is a designated space within the Cumming School of Medicine with specialized health data science, epidemiology, and biostatistics personnel who work together to advance and innovate health informatics for precision health projects. CHI facilitates timely data access, provides analytic support, data science training, and collaboration.
- d) Centre for Health Informatics - Alberta Health Services Analysts (CHI-AHS analysts) are in-house AHS employees appointed to CHI who are the conduits that enable timely access to AHS data holdings for CHI supported research. These analysts fulfill the information management role as stipulated by the IMA standards and facilitate analytical workflow with study PI's.
- e) Conjoint Health Research Ethics Board (CHREB) is the governing ethics body for faculties of medicine, kinesiology, and nursing at University of Calgary
- f) Information Management Agreement (IMA) is a document that The University of Calgary has signed with AHS that allows AHS employees within CHI to analyze EMR data using the Secure Computing Platform following the data release, storage, and computing standards described within.
- g) Information Technologies (IT) department provides computing and media services in support of learning, teaching, research, and administration at the University of Calgary.

- h) Principal Investigator (PI) is the holder of an independent grant administered by a university and the lead researcher for the grant project.
- i) Research Computing Services (RCS) is a group within the wider University of Calgary IT team that plans, manages, and supports high performance computing (HPC) systems in use by researchers throughout the University of Calgary. With the primary focus of meeting the increasing demand for engineering and scientific computation, RCS offers a wide range of specialized services to help researchers solve highly complex real-world problems or run large scale computationally intensive workloads on high-end HPC resources.
- j) ServiceNow is a cloud platform Service Management tool that IT uses to track all incidents, requests, changes, approvals and configuration items. It also hosts the IT website for direct client interaction.

4.0 Procedure:

4.1 Prerequisites

4.1.1 University of Calgary IT account

Many services provided by the University of Calgary require an IT account. All University of Calgary students and employees are required to register during the on-boarding process.

To sign-up, visit [IT – Account Registration](#).

4.1.2 Multi-Factor Authentication

MFA is an authentication method in which the user is granted access only after successfully presenting two or more of the following factors:

- Knowledge - password (recommended)
- Possession - mobile device (recommended)
- Inherence - biometrics

MFA provides a second layer of access protection when using the data services provided by the University of Calgary.

To set up MFA, visit [MFA – How to Get Started](#).

4.2 Data Classification

[University of Calgary Data Security Classification](#), Level 2 and Level 4, are university-wide designated data classification standards relevant to health data. Relevant security classifications from table 4.1 are included here for quick reference.

Table 1. Security Classification

Classification	Definition	Examples
Level 2 Internal Use (anonymized health data)	<p>Information not approved for general circulation outside the University.</p> <p>Information the disclosure or loss of which would inconvenience the University although it would unlikely result in financial loss or reputational damage.</p>	<p>Unpublished researcher data.</p> <p>Anonymized or de-identified human subject data.</p>
Level 4 Restricted Use (electronic medical record data from Sunrise Clinical Manager™ (SCM) or Epic™ from Connect Care)	<p>Information that is confidential; and subject to specific privacy and security safeguards under law, policy or contractual agreement.</p> <p>Information the loss or disclosure of which could cause severe harm to individuals or the University.</p> <p>Information the loss or disclosure of which may obligate the University to report to the government or other regulating body and/or provide notice to affected individuals.</p>	<p>Health information when it can be linked to an identifiable individual including:</p> <ul style="list-style-type: none"> • information about health status • diagnostic, treatment or care information <p>Identifiable human subject research data.</p>

4.3 Data Request and Release

The following section contains instructions for requesting data from AHS holdings through the University of Calgary's platforms. This process is available to all University of Calgary researchers and if assistance is required contact chi@ucalgary.ca.

This permission process is handled differently for Level 4 Electronic Medical Records (EMR) data (e.g. SCM or EPIC data). These studies should involve CHI and AbSPORU resources. The PI must arrange for a CHI consultation via chi@ucalgary.ca. Study protocol and resources will be discussed and contracted services can be arranged based on project funding.

4.3.1 Obtain CHREB Ethics Approval

All study participants must complete the [TCPS 2: CORE](#) online training module that encompasses the 2nd edition of the Tri-Council Policy Statement and includes guidelines on how to ethically conduct research involving humans. Once completed, create an account with the online human ethics submission management system, IRISS. To register an account visit the [IRISS Homepage](#). This platform houses the CHREB ethics submission form.

The following documents should be submitted within the CHREB application:

- Study Protocol [Required]
- CHREB Budget Summary [Required]
- Department Approval [Required]
- Other forms are required based on your study type and data needed (e.g. consent forms)

For more information on CHREB guidelines, eligibility requirements, forms and templates, visit the [CHREB webpage](#).

The PI is responsible for submission of the ethics application as well as any subsequent modifications.

4.3.2 Submit Data Request to AbSPORU

Complete and submit an AbSPORU [Services Request Form](#) and return to the Data Platform via spor@albertainnovates.ca. The approved CHREB application study ID number will be needed for the completion of this form.

4.3.3 Obtain AHS Research Operations Approval

Approved CHREB ethics applications requiring AHS data will be automatically forwarded to the AHS Health Systems Access (HSA) team. The HSA team will send a link with a questionnaire to the study PI. For Level 2 data studies, a Data Disclosure Agreement (DDA) is embedded within the link. This document details the policies and guidelines required to ensure the proper use, storage and disposal of AHS data ensuring the utmost security and confidentiality. The HSA team will review the DDA and based on the questionnaire response, a Provincial Research Administration Study Summary (PRA) form is generated. This form outlines the study protocol, personnel involved, and the databases being requested. Additionally, if funding has been provided for the study, the

award letter or letter of support from your institution/department should be included at this stage.

The PI is responsible for completing and returning the questionnaire, DDA and PRA forms as well as funding documents to the HSA team for review and approval.

4.4 Data Extraction and Linkage

CHI offers a wide range of analytic services to supplement the completion of a study. To access these services, the PI must arrange a consultation with a CHI representative to determine the resources needed, the expertise required and the payment schedule agreed upon before moving forward.

Members of AbSPORU's Data Platform housed within CHI, will receive notice of the data request through an internally established process. The assigned analyst will conduct data extraction and linkage if needed, as outlined in the study protocol and AHS research operations documents.

The PI is responsible for updating the CHREB ethics submission if additional data not specified in the study protocol are required. These data cannot be released to researchers without full ethics board approval.

4.5 Data Storage

4.5.1 Academic File Share (Level 2 specific)

This is a shared storage that can be used as a shared project space that can be accessed simultaneously by users. The default quota is 100GB, but the initial allocation can be increased upon request. Academic File Share is available on campus or off-campus, using the IT-supported VPN client. Parallel or concurrent group access is well supported. Additionally, Academic FS is capable of conducting data analysis, go to section 4.6.1 to learn more.

For more information on Academic File Share and other storage options, visit [RSC – Secure Computing Data Storage](#).

To request an Academic File Share, visit [Service Now](#).

4.5.2 OneDrive for Business (All levels)

This is an Office 365 cloud-based personal storage service that is available to all active UofC accounts. OneDrive for Business cannot be used as a shared project space and is better suited for storing personal materials as only the user has access via a UofC IT login. The default quota is 5TB, but

the initial allocation can be increased upon request. OneDrive for Business is accessible on or off campus with access to the internet. Files stored are by default private but can be shared with internal and external collaborators via an emailed link. Multi-Factor Authentication is required for secure use.

For additional information and tutorials, see [OneDrive for Business – Getting Started](#).

For access to OneDrive for Business, go to the Apps launcher in [Office 365 Portal](#).

4.5.3 Secure Computing Data Storage (SCDS) (Level 4 specific)

SCDS is a service provided by RCS that allows researchers to store confidential Level 4 data on a highly secure group file share. While the SCDS group file share is rated to hold Level 4 data, it can hold any level of data. All data stored in the group file share will be treated as if it were Level 4. A single SCDS group file share should be limited to one project. The default quota is 10GB but 1TB or greater is available upon request. Access is only permitted on devices in the Managed Client Zone. Direct access is only permitted by the user while on analysis platforms like DAT or MARC.

Refer to the [Secure Compute Data Storage](#) webpage for more information on features and specifications. For further instruction on how to use this service, visit the [SCDS Training](#) webpage.

To request access to SCDS group file share, visit [ServiceNow](#).

4.6 Data Computing

RCS provide the following services for computing Level 2 and 4 data. RSC will provide maintenance and analytical support as needed.

Studies using Level 4 EMR data will require CHI and AHS approval prior to data transfer to SCDS for computing on MARC or DAT. Extracted data will be released by AHS to sanctioned study team members.

If more assistance or training is needed on the programs and services below, send an email to support@hpc.ucalgary.ca.

4.6.1 Academic File Share (Level 2 specific)

On Academic File Share, parallel or concurrent group access is well supported. When using a personal MacOS, Windows, or Linux workstation, Academic FS storage can be 'mounted' onto the work station to process data (appearing as a local drive). Most programs installed on a workstation can read/write to that mounted storage. The data/files in the shared storage can be marked as 'locked' or 'in use' when you are editing a file. This ensures that while working on a file, editing capabilities can be exclusive to a single user.

To request an Academic File Share, visit [Service Now](#).

4.6.2 Advanced Research Computing (ARC) (Level 2 specific)

This system is a Linux based HPC cluster for UCalgary researchers. It consists compute nodes (1200 nodes as of 2020), each containing between 8 and 80 cores. These are connected with a high-speed network that allows fast communication between processes on different nodes. The different types of hardware are grouped into partitions which are internally homogeneous. ARC supports batch-oriented workflows that involve running multiple processes and computational threads simultaneously. On ARC, computations have to be submitted as jobs to a job scheduler called SLURM. Submitted jobs are then assigned to appropriate compute nodes as resources become available. For internal collaboration, the user collaborator must request access to the existing ARC file system.

The [ARC Cluster Guide](#) provides a high-level introduction to this system, including the SLURM commands and sample job scripts.

Please contact support@hpc.ucalgary.ca to request an ARC account.

4.6.3 Data Analysis Tool (DAT) (Level 4 specific)

DAT is a service that enables rapid processing of smaller data sets containing Level 4 data. Researchers can analyze data and visualize results in a virtual environment. It is a Windows-based computational cluster with a graphical user interface and a persistent workflow. Data stored in DAT is written and stored back in SCDS, therefore, storage limits are restricted to that of the SCDS group file share where the data resides.

4.6.3.1 Prerequisites

The following criteria must be met before requesting access:

- Be a member of an existing SCDS group file share
- Install the Citrix Workspace App for access to the platform (To install on a University PC: Start menu > search programs and

files for ‘Software Center’ > in “Available Software” tab, select “Citrix ICA Client” and “Install”)

Refer to the [DAT](#) webpage for more information on computing resources, features and specifications. For further instruction on how to use this service, visit the [DAT Training](#) webpage.

To request access to a DAT virtual machine, visit [Service Now](#).

4.6.4 Medical Advanced Research Computing (MARC) (Level 4 specific)

MARC is a high-performance computing infrastructure for large complex data sets that is designed with controls appropriate for Level 4 classified data stored in an SCDS group file share. It is a Linux-based computational cluster made up of 8 compute nodes with 40 cores as well as 1 large memory node with 80 cores. Due to security requirements for Level 4, restrictions have been placed on MARC to prevent accidental (or otherwise) data exfiltration. MARC has a batch-oriented environment where jobs are submitted and the SLURM scheduler runs the job when resources become available. Jobs can span more than one compute node if required.

4.6.4.1 Prerequisites

The following criteria must be met before requesting access:

- Be a member of an existing SCDS group file share
- Install the Citrix Workspace App for access to the platform
- Familiarity with the Linux environment including: Linux command line, files and directories; file editors like ed,vi,vim or emacs; and the job scheduler Slurm Workload Manager is required.

4.6.4.2 Compute Resources

Home file system: /home

There is a per-user quota of 25GB under /home. This limit is fixed and cannot be increased. Each user has a directory under /home, which is the default working directory when logging in to MARC. It is expected that most researchers will work from /project and only use home for software. /home is expected to be used only for Level 2 data and not for your patient identifiable files. The identifiable files go in the appropriate directory under /project.

Project file system: /project

This file system is intended for larger projects. Directories will be created in /project named after your project ID. This name will be

the same as your SCDS share name. The expectation is that all files to do with that project will be stored in /project/projectid. Quotas in /project are flexible.

Refer to the [MARC](#) webpage for more information on computing resources, features and specifications. For further instruction on how to use this service, visit the [MARC Training](#) webpage.

To request access to MARC, visit [Service Now](#).

Please write to support@hpc.ucalgary.ca with an estimate of how much space you will require.

The [Secure Compute Platform](#) offers two analyses options for Level 4 data – Data Analysis Tool (DAT or Medical Advanced Research Computing (MARC) depending on the requirements of the research project. Visit the “[DAT or MARC](#)” webpage for a detailed comparison of each analysis platform.

4.7 Data Sharing

4.7.1 OneDrive for Business (All levels)

Files stored within OneDrive for Business can be shared with internal and external collaborators through a customizable link. The link can be customized to allow/disallow editing abilities, prevent downloading and set viewing restriction with a password or expiration date. The level of access for the recipient can be set so that anyone with the link, only people within your organization, only people with existing access to the file, or only individuals you specify can open the link. These stipulations cannot be changed once the link has been sent. Although, OneDrive tracks who the file has been shared with and access can be revoked by the file owner at any time.

For additional information on how to share and collaborate, see [OneDrive for Business – Getting Started](#).

OneDrive for Business is automatically enabled on your UCalgary account once MFA is set up. For access, go to the Apps launcher in [Office 365 Portal](#).

4.7.2 ShareFile (Level 4 specific)

ShareFile is a service for sharing and organizing files online. It allows internal and external collaborators to share Level 4 data in a secure, password protected environment. The system is similar to that of an Outlook email account with greater security measures. While ShareFile

has security specifically fit for Level 4 data, these restrictions can be customized to fit the level of data being handled. When sharing a file, the email can be customized to allow for view only, to restrict the number of views or downloads, and/or set an expiration date on the file where access is disabled after a certain amount of hours or days. These stipulations can be changed after the link has been shared allowing continuous control. MARC and DAT are both packaged with this data sharing and collaboration tool.

Refer to the [ShareFile](#) webpage for more information on features and specifications. For further instruction on how to use this service, visit the [ShareFile Training](#) webpage

Access to ShareFile is granted once a SCDS group share file is created. All SCDS share files will be accessible through a single ShareFile account. Go to the [ShareFile account login](#) to register your account.

5.0 Appendix

Information Management Agreement

<https://cumming.ucalgary.ca/centres/centre-health-informatics/data-and-analytic-services>

University of Calgary Data Classification Standards

https://www.ucalgary.ca/policies/files/policies/im010-03-security-standard_0.pdf

Information Technologies (IT)

<https://itregport.ucalgary.ca>

Multi-Factor Authentication (MFA)

https://ucalgary.service-now.com/it?id=kb_article&sys_id=74c52bfbdb901b806b3e777a8c9619c4

TCPS 2 CORE

<https://research.ucalgary.ca/conduct-research/ethics-compliance/tcps2-core-tutorial>

Institutional Research Information Services Solution (IRISS)

<https://www.ucalgary.ca/iriss/>

Conjoint Health Research Ethics Board (CHREB)

<https://research.ucalgary.ca/conduct-research/ethics-compliance/human-research-ethics/conjoint-health-research-ethics-board>

Research Computing Services - Secure Data Storage

https://ucalgary.service-now.com/kb_view.do?sysparm_article=KB0030163

Academic File Share

https://ucalgary.service-now.com/kb_view.do?sysparm_article=KB0030163

OneDrive for Business

https://ucalgary.service-now.com/it?id=kb_article&sys_id=60994170db2da7487cab5068dc961900

Secure Compute Data Storage (SCDS)

General Information: <https://it.ucalgary.ca/secure-computing-platform/store-your-data>

Training: <https://it.ucalgary.ca/secure-computing-platform/training/secure-computing-data-storage-scds-training>

Request access: https://ucalgary.servicenow.com/it?id=sc_cat_item&sys_id=fe66b3a7db297300897e4b8b0b96199d

Advanced Research Computing (ARC)

General Information: https://rcs.ucalgary.ca/index.php/ARC_Cluster_Guide

Request access: support@hpc.ucalgary.ca

Data Analysis Tool (DAT) virtual machines

General Information: <https://it.ucalgary.ca/secure-computing-platform/analyze-your-data/data-analysis-tools-dat>

Training: <https://it.ucalgary.ca/secure-computing-platform/training/data-analysis-tools-dat-training>

Request access: https://ucalgary.servicenow.com/it?id=sc_cat_item&sys_id=890dd8b51bacc810d61c4000cd4bcb8c

Medical Advanced Research Computing

General Information: <https://it.ucalgary.ca/secure-computing-platform/analyze-your-data/medical-advanced-research-computing-marc>

Training: <https://it.ucalgary.ca/secure-computing-platform/training/medical-advanced-research-computing-marc-training>

Request access: https://ucalgary.service-now.com/it?id=sc_cat_item&sys_id=9080827cdb974c10897e4b8b0b96192a

ShareFile

General Information: <https://it.ucalgary.ca/secure-computing-platform/share-your-data>

Training: <https://it.ucalgary.ca/secure-computing-platform/training/data-sharing-and-collaboration-sharefile-training>

Register account: <https://ucalgary.sharefile.com/Authentication/Login>