Imaging Materials for a Brighter World

CANADIAN CENTRE FOR ELECTRON MICROSCOPY
DATA MANAGEMENT AND CYBERSECURITY FRAMEWORK FY25.1

BREAKTHROUGH MATERIALS RESEARCH

LEADING IN EDUCATION AND TRAINING

PROVIDING SPECIALIZED SERVICES

McMaster University

CCEM
Canadian Centre for Electron Microscopy

Table of Contents

1. Introduction

CCEM is a national research facility whose users and staff produce vast amounts of scientific data. The proper and secure handling, documentation, storage, sharing, processing, and publication of data are of high importance to CCEM.

CCEM's Strategic Plan 2023-29 provides direction with respect to CCEM's data management and cybersecurity posture. Specifically, under the Strategy's goal to re-imagine CCEM's business model to drive impact, inclusiveness, and innovation, CCEM has identified the following Sub-Goals:

- Transition from data-centric to secure knowledge-centric projects. We will grow our capabilities to translate our partners' datasets into knowledge and solutions through custom-tailored consulting, expanded data acquisition, and processing services, thus improving research context and impact; and

- Expand our virtual environment. We will expand our ability to engage researchers and clients from across Canada and internationally through virtual access to our facilities and services.

The Data Management and Cybersecurity Framework, approved by the Governing Board, outlines best practices adopted by CCEM to ensure data and intellectual property are secure and provides guidance for staff and users regarding expectations consistent with the goals of the Strategic Plan. Further, this document codifies the process of data collection, processing, storage, documentation, backup, the types of data, roles and responsibilities of CCEM users and Management, data ownership and sharing, and cybersecurity.

As part of its annual business planning process, CCEM Management will assess the state of its data management and cybersecurity capabilities and practices relative to the expectations outlined in this Framework. The Annual Business Plan will identify actions to be taken and investments to be made in order to close identified gaps.

2. Principles

Threats to CCEM's information technology (IT) infrastructure pose a constant risk to the facility's reputation and must be managed professionally with the everchanging situation in mind.

The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) have released guidelines for research data management (full text can be found here: https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html). These guidelines require research institutions to develop and follow an institutional strategy on data management and provide guidance for the establishment of this framework.

Since CCEM is part of McMaster University, it follows the policies and guidelines setup by the university and relies on IT resources that the university offers. In addition, this Framework outlines the specific steps the facility takes to ensure data are handled according to best practices. University guidelines on data management, which serve as a basis for this document, can be found here: https://rdm.mcmaster.ca/

Integrity, security and archival of research data are important to ensure that CCEM enabled research follows good scientific practice and can be used to extract the maximum amount of information which in turn can be processed into publishable knowledge. Data at CCEM are handled according to FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable).

3. Data Life Cycle

3.1 Data Types

At CCEM, the following types of data are distinguished:

Scientific Data: These data may be generated by CCEM staff, as part of jobs, or independently by CCEM users. *Scientific data* can take the following forms:

Raw Data: A form of *scientific data* as-collected by the instrument, often using vendor-specific proprietary formats.

Meta Data: A form of *scientific data* such as experimental conditions or instrument parameters used to acquire the datasets, typically affiliated with another form of *scientific data* (e.g. *raw data* or *processed data*).

Method Data: A form of *scientific data* that is used to describe the inputs, techniques and/or manner by which staff or users acquire, analyze, or process other *scientific data.*

Processed Data: Datasets that have been processed, filtered, edited, or annotated in some form, typically to extract knowledge and to be included in manuscripts for publication.

Algorithmic Data: A form of *scientific and processed data* that is used to describe the techniques, processes, and/or scripts developed to process and analyze data.

Administrative Data:

Employee Data: This includes all employee emails and other digitally tracked staff, vendor, facility and user interactions.

Password Data: This includes all necessary passwords required to be stored in an encrypted and secure fashion using third party software to protect all types of data in this framework.

Business Data This includes all data necessary for operating the facility. Examples are customer records, billing information, statistics for reporting to funding agencies.

3.2 Data Locations

CCEM data locations are defined as:

Individual Workstation: Desktop or CCEM-supplied laptop given to individual staff to conduct their daily tasks. This includes job information or notes; communication with users, service engineers, or other staff; report writing; or data processing.

Instrument Computers: Computers provided by instrument vendors or supplied by CCEM for the purposes of operating the instruments, and the acquisition and temporary storage of raw scientific data. Instrument computers are typically located in the same room as the instrument or in a nearby operator room.

CCEM Data Server: A temporary online data storage solution available to users and staff for the purposes of i) safely moving data from instrument computers to a user's personal storage device, ii) working storage for data being processed using other CCEM compute resources, iii) centralizing user data affiliated with CCEM jobs and sharing with those users. The CCEM Data Server is located and maintained by RHPCS.

MacLIMS: A laboratory management system to collect, store and maintain business data. Staff may upload additional job information to MacLIMS to be shared with the user or other staff. The MacLIMS servers are maintained by both RHPCS and UTS.

Virtual Workstations: Virtual workstations are provided to staff and users to aid in processing instrument and job data, by granting access to software and processing hardware hosted by CCEM. The virtual workstations are accessed remotely and hosted on servers maintained by RHPCS.

Lab Computers: Lab computers are provided to staff and users to aid in processing instrument and job data. These may be either desktop workstations with various software dedicated to the post-processing and analysis of scientific data, or thin client workstations provided for access to the virtual workstations, with no local storage. The lab computers are located within the shared CALM/CCEM computer area (ABB B160/D) and the CCEM X-Ray Suite (ABB CB105).

3.3 Collection

The following data types are collected at CCEM through the these means:

Scientific Data:

Raw Data are collected on instruments, often using vendor-supplied proprietary software.

Meta Data are collected through the headers of the datasets or in the laboratory management system MacLIMS.

Method Data are collected by staff and users, often using software external to the instrument (e.g. digital lab book or note-taking software).

Administrative Data:

Employee Data are collected through day-to-day communication with users and staff.

Password Data are collected and stored using the password management system LastPass.

Business Data are collected using the laboratory management system MacLIMS.

3.4 Processing

Data processing at CCEM may take in the data defined in 3.3 Collection, and generates the follow data types as outputs: *processed data*, or *algorithmic data*.

Processing of data are performed on individual workstations or virtual workstations. The virtual workstations are actively monitored for suspicious activity by McMaster IT and are only accessible using a Virtual Private Network (VPN).

CCEM offers a range of instrument vendor-supplied and third-party software packages for use on CCEM systems.

Users can also opt to utilize their own capabilities or use cloud-based services at commercial third-parties or Compute Canada.

During data processing, temporary datasets are often generated to save intermediate results in the processing workflow. These temporary datasets are to be deleted after use to save space.

Processing algorithms, especially if developed by CCEM staff or users, should be well documented and described in the manuscript if the data are published. Users are also encouraged to use and contribute to the open-source software community.

3.5 Documentation

Scientific data should be annotated with metadata (instrument parameters, experiment description, sample origin, etc.).

Processed data should be thoroughly documented for reproducibility such that other staff and users may use it.

Instruments produce a variety of data with proprietary software. If long-term storage is required, open formats (e.g., tagged image file (TIF) for images) should be used.

Datasets are stored in libraries on the CCEM Data Server. This system is linked with the laboratory management system (MacLIMS) to automatically provision space, user access rights and links to user information and sample descriptions.

3.6 Storage and Backup

The McMaster Research Ethics Board (MREB) Data Storage & Security Guide of McMaster guidelines must be followed. All data are treated as high-risk by default. This ensures maximum security of all data and simplifies data management on the storage systems.

Scientific data are temporarily stored on the instrument computer and moved to central storage immediately after the instrument session. Unprocessed raw data are stored whenever feasible, to maintain scientific integrity and reproducibility of experiments. Should raw datasets be too large to be practically stored, data reduction can be performed before saving. This process must be documented using a Standard Operating Procedure, which will be saved together with the datasets.

Often, vendors by default save data in proprietary formats which are not suitable for long-term archival. If possible, saving data in open or documented formats is preferred and users and staff are encouraged to do so. Additionally, CCEM provisions a virtual compute platform that is accessible to all users and contains relevant vendor-supplied software packages. CCEM maintains these systems with long-term availability of the original software packages in mind. This way, datasets in proprietary formats are accessible as long as technology reasonably allows (typically 5-10 years).

Storage of data on instrument computers and virtual workstations is only permitted short term, and must be moved as soon as possible, ideally at the end of the instrument session. CCEM is aware that immediate removal is not always possible, but given proper training, staff may purge any leftover data on these systems at their discretion, to ensure no unprotected data are exposed to other users.

Whenever a user utilizes a CCEM service they are provided with access to the CCEM Data Server. User-generated scientific data may be stored using their data server allotment until the user becomes inactive. Data relevant to the user jobs are stored by CCEM staff using libraries on the data server, and kept for a storage term of five years. This data storage lifecycle is to be acknowledged by the user in a user agreement. This term excludes data relevant to preservation of *method data*, if CCEM staff determines such data could be relevant in the future.

To ensure security, all data must be stored in the approved systems listed under section 3.2 Data Locations. If a user requires data to be transferred off-site, CCEM will offer the appropriate encrypted communication channels to move data to the user’s preferred storage device. Once the data is transferred off-site the user acknowledges that CCEM is no longer liable for any lost or stolen data.

Access to storage libraries on the CCEM Data Server is setup so only specified authorized users and staff can access data relevant to them. To facilitate collaboration, users can generate share links to specific people.

Administrative data are stored indefinitely across various locations depending on the data type.

Password and business data are necessary for daily operations and are thus catalogued and owned by CCEM. An employee may choose to encrypt their files, if they do, the employee is required to disclose any encryption keys or passwords associated with the encrypted files to appropriate CCEM management for safe storage.

Employee data is regularly maintained through University Technology Services (UTS) storage systems and are property of McMaster University.

3.7 Retention and Preservation

To ensure reliable archival of data, CCEM Data Server is backed up and monitored professionally by RHPCS of McMaster. Long-term archival of published open data should be done in an appropriate repository (e.g., the Federated Research Data Repository (FRDR)). Data can be accessed by user, supervisor or CCEM management. Data will be deleted after the retention time has elapsed as outlined in the project information stored in the laboratory information management system.

3.8 Publication

Data can only be published by or in collaboration with users who submitted the project request or their principal investigator. Publication of CCEM acquired data must follow the guidelines document. CCEM must be acknowledged and CCEM staff must be included in the author list if deemed that their contributions to the research have been significant. Users are responsible for data interpretation, excepting when CCEM staff are the sole researcher involved in this process.

4. Roles and Responsibilities

Users and staff take shared responsibility for data integrity and safe storage. CCEM provides appropriate data storage and processing facilities and ensures backups are available in case disaster recovery is needed. CCEM is responsible for non-disclosure agreement (NDA) data safety and the safe deletion of expired data. On project start, CCEM requires user to define data retention period and define rules around sharing and transfer. This information is stored in the laboratory information management system.

Users are responsible to use the CCEM Data Server according to SOPs and to not store any data in unsecured locations. Users are also responsible to transfer datasets to their own infrastructure as soon as possible and ensure data is archived appropriately. CCEM does not guarantee data availability on its own systems for any given time.

5. Ethics

In most cases, the project leader (user) is responsible for ethics approval if applicable (tissue samples, human samples, animals). Proof of ethics approval is required by CCEM before starting the project. In case of CCEM leading a research project, MREB approval will be sought.

6. Intellectual Property and Ownership

Generally, in research projects where users and CCEM staff are contributing collaboratively, unprocessed data types are owned by the user, methods and algorithms are owned by CCEM. This ensures that CCEM staff expertise can be used on future projects for the benefit of the entire user community.

For industry projects, NDA can be setup if required by the company, CCEM staff will not discuss projects with other users unless approved by customer. CCEM will use the resources of the McMaster Industry Liaison Office (MILO) and the McMaster Legal Office to develop and negotiate NDAs.

Data are kept confidential, if necessary, on separate system depending on NDA conditions. Business data are owned by CCEM, and may be used for operations (billing, booking etc.).

Administrative data are the property of CCEM.

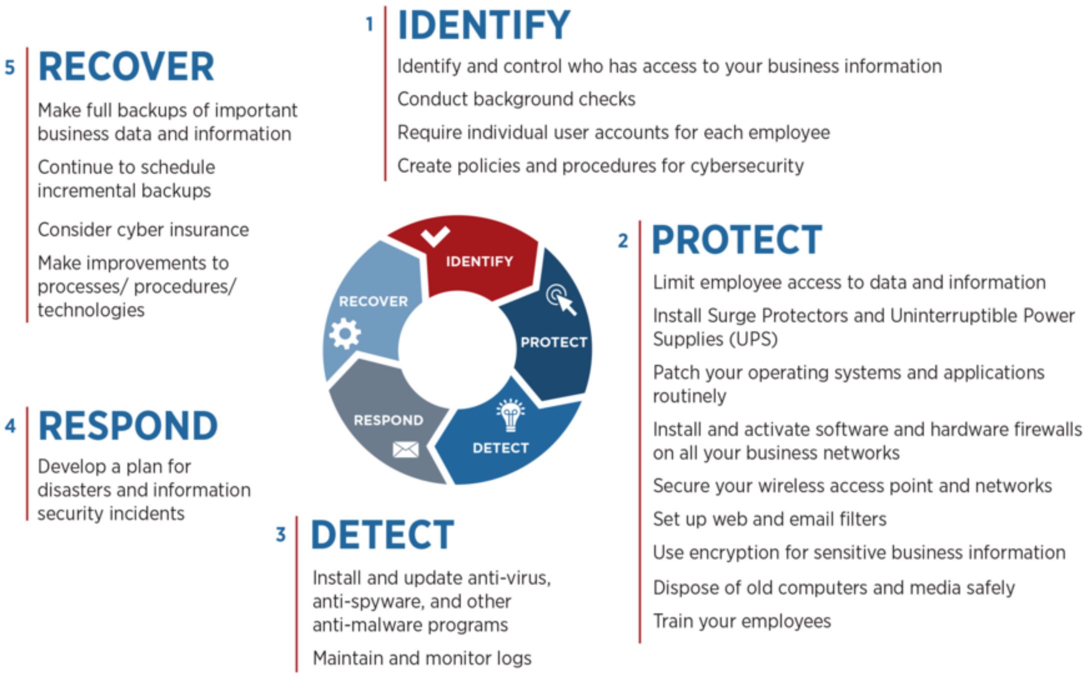
7. Sharing

Transfer of data between CCEM systems is generally done by secure network connection using state of the art encryption (transport layer security (TLS)). Sharing between CCEM and user can be done by secure network link, or if required by physical media (hard disk, etc.). CCEM is promoting open data principles where reasonable and is offering staff assistance with depositing datasets into repositories, curation, etc.

8. Cybersecurity

Cyber-attacks on CCEM infrastructure pose a serious risk to CCEM's operations and reputation.

CCEM implements the National Institute of Standards and Technology (NIST) Cybersecurity Framework Version 1.1. The elements of the framework are:



1 **IDENTIFY**

- Identify and control who has access to your business information
- Conduct background checks
- Require individual user accounts for each employee
- Create policies and procedures for cybersecurity

2 **PROTECT**

- Limit employee access to data and information
- Install Surge Protectors and Uninterruptible Power Supplies (UPS)
- Patch your operating systems and applications routinely
- Install and activate software and hardware firewalls on all your business networks
- Secure your wireless access point and networks
- Set up web and email filters
- Use encryption for sensitive business information
- Dispose of old computers and media safely
- Train your employees

3 **DETECT**

- Install and update anti-virus, anti-spyware, and other anti-malware programs
- Maintain and monitor logs

4 **RESPOND**

- Develop a plan for disasters and information security incidents

5 **RECOVER**

- Make full backups of important business data and information
- Continue to schedule incremental backups
- Consider cyber insurance
- Make improvements to processes/ procedures/ technologies

CCEM partners with UTS and RHPCS of McMaster University for safe operations and intrusion detection. McMaster UTS performs IT security services (firewall, VPN, virus scanners, email security, incident investigation, etc.). Staff are trained in basic IT security principles and users are made aware of the same (e.g., phishing awareness, data storage standard operating procedures (SOPs)). When designing computer systems, security is given special consideration (Linux based servers, regular updates).

CCEM’s risk register includes cybersecurity risks. These risks are re-assessed and updated in accordance with the schedule outlined in the Risk Management Framework.

9. Annex

9.1 Acronyms Defined

CALM	McMaster Centre for Advanced Light Microscopy
CCEM	Canadian Centre for Electron Microscopy
FAIR	Findable, Accessible, Interoperable, and Reusable
FRDR	Federated Research Data Repository
IT	Information Technology
MILO	McMaster Industry Liaison Office
MREB	McMaster Research Ethics Board
NDA	Non-disclosure Agreements
NIST	National Institute of Standards and Technology
RBAC	Role-based Access Control
RHPCS	Research High Performance Computing Service
SOP	Standard Operating Procedure
TIF	Tagged Image File
TLS	Transport Layer Security
UPS	Uninterruptable Power Supplies
USB	Universal Serial Bus
VPN	Virtual Private Network

9.2 Storage Types and Locations

Storage Type	Data Type(s)	Data Owner	Compliance Responsibility	Storage Term
Individual Workstation	Employee Data	McMaster	McMaster/CCEM	Indefinite
	Raw Processed	User	CCEM	5 years
	Meta Method Algorithmic Password Business	CCEM		Indefinite
Instrument Computers	Raw	User	CCEM	None
	Meta Method Password	CCEM		
CCEM Data Server	Raw Meta Method Processed Algorithmic	User	RHPCS/CCEM	5 years or User Inactive
	Meta Method Algorithmic	CCEM		Indefinite
MacLIMS	Processed Business	CCEM	RHPCS	Indefinite
Virtual Workstations	Raw Meta Method Processed Algorithmic	User	RHPCS/CCEM	None
	Meta Method Algorithmic	CCEM		None
Lab Computers	Raw Meta Method Processed Algorithmic	User	CCEM	None
	Meta Method Algorithmic	CCEM		None