

## Schedule 4 - National Cybersecurity Consortium Data Management Plan Template for Ultimate Funding Recipients

The National Cybersecurity Consortium's (NCC) Data Management Policy for funding programs, which may be found [here](#), outlines the NCC's data management requirements of ultimate funding recipients (i.e., project leads who signed funding agreements with the NCC). It requires ultimate funding recipients (UFRs) to commit to ensuring that data management planning is undertaken at all stages of a project that involves data - from design and inception to completion - and that data management plans are an essential part of this work.

The NCC has created this template for UFRs to complete in order to meet their data management requirements as specified by the NCC. This template is based on the guidance provided through the [Portage Network's Data Management Plan Assistant](#). A reminder of the NCC's data management-related monitoring and reporting requirements for UFR may be found at the end of this document.

### Data Collection

1. What types of data will you collect, create, link to acquire and/or record (e.g., numeric, images, audio, video, text, tabular data, modeling data, spatial data, instrumentation data)?
2. What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

*Notes:* Proprietary file formats requiring specialized software or hardware to use are not recommended, but may be necessary for certain data collection or analysis methods. Using open file formats or industry-standard formats (e.g., those widely used by a given community) is preferred whenever possible. Read more about file formats: [Format: University of British Columbia \(UBC\) Library](#) or [Format Your Data: United Kingdom \(UK\) Data Service](#).

3. What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

*Notes:* It is important to keep track of different copies or versions of files, files held in different formats or locations, and information cross-referenced between files. This process is called 'version control'. Logical file structures, informative naming conventions, and clear indications of file versions, all contribute to better use of your data during and after your research project. These practices will help ensure that you and your research team are using the appropriate version of your data and minimize confusion regarding copies on different computers and/or on different media. Read more about file naming and version control: [Organize: UBC Library](#) or [Format Your Data: UK Data Service](#).

### Documentation and Metadata

1. What documentation will be needed for the data to be read and interpreted correctly in the future?

*Notes:* Typically, good documentation includes information about the study, data-level descriptions, and any other contextual information required to make the data usable by other researchers. Other elements you should document, as applicable, include:

research methodology used, variable definitions, vocabularies, classification systems, units of measurement, assumptions made, format and file type of the data, a description of the data capture and collection methods, explanation of data coding and analysis performed (including syntax files), and details of who has worked on the project and performed each task, etc.

2. How will you make sure that documentation is created or captured consistently throughout your project?

*Notes:* Consider how you will capture this information and where it will be recorded, ideally in advance of data collection and analysis, to ensure accuracy, consistency, and completeness of the documentation. Often, resources you've already created can contribute to this (e.g., publications, websites, progress reports, etc.).

It is useful to consult regularly with members of the research team to capture potential changes in data collection/processing that need to be reflected in the documentation. Individual roles and workflows should include gathering data documentation as a key element.

3. Are you using a metadata standard and/or tools to document and describe your data? If so, please list.

*Notes:* There are many general and domain-specific metadata standards. Dataset documentation should be provided in one of these standard, machine readable, openly-accessible formats to enable the effective exchange of information between users and systems. These standards are often based on language-independent data formats such as XML, RDF, and JSON. There are many metadata standards based on these formats, including discipline-specific standards.

Dataset documentation may also include a controlled vocabulary, which is a standardized list of terminology for describing information. Examples of controlled vocabularies include the [Subject Headings: Library of Congress Subject Headings](#) (LCSH) or NASA's [Global Change Master Directory \(GCMS\) Keywords](#).

Read more about metadata standards: [Disciplinary Metadata: Digital Curation Centre](#) (UK)

## **Storage and Backup**

1. What are the anticipated storage requirements for your project, in terms of storage space (e.g., in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

*Notes:* Storage-space estimates should take into account requirements for file versioning, backups, and growth over time. If you are collecting data over a long period (e.g., several months or years), your data storage and backup strategy should accommodate data growth. Similarly, a long-term storage plan is necessary if you intend to retain your data after the research project.

2. How and where will your data be stored and backed up during your research project?

*Notes:* The risk of losing data due to human error, natural disasters, or other mishaps can be mitigated by following the 3-2-1 backup rule:

- I. have at least three copies of your data;
- II. store the copies on two different media; and
- III. keep one backup copy offsite.

Data may be stored using optical or magnetic media, which can be removable (e.g., DVD and USB drives), fixed (e.g., desktop or laptop hard drives), or networked (e.g., networked drives or cloud-based servers). Each storage method has benefits and drawbacks that should be considered when determining the most appropriate solution. Further information on storage and backup practices is available from the [Data Storage: University of Sheffield Library](#) and [Store Your Data: UK Data Service](#).

3. How will the research team and other collaborators access, modify, and contribute data throughout the project?

*Notes:* An ideal solution is one that facilitates cooperation and ensures data security while also able to be adopted by users with minimal training. Transmitting data between locations or within research teams can be challenging for data management infrastructure. Relying on email for data transfer is not a robust or secure solution. Third-party commercial file sharing services (such as Google Drive and Dropbox) facilitate file exchange, but they are not necessarily permanent or secure, and are often located outside Canada. Please contact your library or data manager to develop the best solution for your research project.

## Preservation

1. Where will you deposit your data for long-term preservation and access at the end of your research project?

*Notes:* The issue of data retention should be considered early in the research lifecycle. Data-retention decisions can be driven by external policies (e.g., funding agencies, journal publishers), or by an understanding of the enduring value of a given set of data. The need to preserve data in the short-term (i.e., for peer-verification purposes) or long-term (for data of lasting value), will influence the choice of data repository or archive. A helpful analogy is to think of creating a 'living will' for the data, that is, a plan describing how future researchers will have continued access to the data.

If you need assistance locating a suitable data repository or archive, please contact your library. The directory [re3data.org](http://re3data.org) provides potential open data repositories. Verify whether the data repository will provide a statement agreeing to the terms of deposit outlined in your Data Management Plan.

2. How will you ensure your data is preservation ready?

*Notes:* Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation. Some data formats are optimal for long-term preservation of data. For example, non-proprietary file formats, such as text ('.txt') and comma-separated ('.csv'), are considered preservation-friendly. The [Format Your Data: UK Data Service](#) provides a useful table of file formats for various types of data. Keep in mind that preservation-friendly files converted from one format to another may lose information (e.g., converting from an uncompressed TIFF file to a compressed JPG file), so changes to file formats should be documented.

Identify steps required following project completion in order to ensure the data you are choosing to preserve, or share is anonymous, error-free, and converted to recommended formats with a minimal risk of data loss. Read more about anonymization: [Anonymize and De-identify: UBC Library](#) or [Data Protection: UK Data Service](#).

## Sharing and Reuse

1. What data will you be sharing and in what form? (e.g., raw, processed, analyzed, final).

*Notes:* Raw data are the data directly obtained from the instrument, simulation, or survey.

Processed data result from some manipulation of the raw data to eliminate errors or outliers, to prepare the data for analysis, to derive new variables, or to de-identify the human participants.

Analyzed data are the results of qualitative, statistical, or mathematical analysis of the processed data. They can be presented as graphs, charts, or statistical tables.

Final data are processed data that have, if needed, been converted into a preservation-friendly format.

Consider which data may need to be shared to meet institutional or funding requirements, and which data may be restricted because of confidentiality/privacy/intellectual property considerations.

2. Have you considered what type of end-user license to include with your data?

*Notes:* Licenses determine what uses can be made of your data. Funding agencies and/or data repositories may have end-user license requirements in place; if not, they may still be able to guide you in the development of a license. Once created, please consider including a copy of your end-user license with your Data Management Plan. Note that only the intellectual property rights holder(s) can issue a license, so it is crucial to clarify who owns those rights.

There are several types of standard licenses available to researchers, such as the Creative Commons licenses and the Open Data Commons licenses. In fact, for most datasets it is easier to use a standard license rather than to devise a custom-made one. Note that even if you choose to make your data part of the public domain, it is preferable to make this explicit by using a license such as [Creative Commons' CC0](#). For more information, read [How to License Research Data: Digital Curation Centre](#) (UK).

3. What steps will be taken to help the research community know that your data exists?

*Notes:* Possible tools for making data available to others include: data registries, repositories, indexes, word-of-mouth, publications.

4. How will the data be accessed (e.g., web service, file transfer protocol, etc.)?

*Notes:* If possible, choose a repository that will assign a persistent identifier (such as a Digital Object Identifier (DOI)) to your dataset. This will ensure a stable access to the dataset and make it retrievable by various discovery tools.

One of the best ways to refer other researchers to your deposited datasets is to cite them the same way you cite other types of publications (articles, books, proceedings). The UK Digital Curation Centre provides a guide for how to [Cite Datasets and Link to Publications](#). Some data repositories also create links from datasets to their associated papers, thus increasing the visibility of the publications. Contact your library for assistance in making your dataset visible and easily accessible.

## Responsibilities and Resources

1. Who will be responsible for managing this project's data during and after the project?

*Notes:* It is important to identify important data activities in your project and who will be responsible -- individuals or organizations -- for addressing these parts of your data management plan. This should include timelines associated with these staff responsibilities and any training needed to prepare staff for these duties. For all NCC data management reporting, UFRs must identify an individual who is the project's point-of-contact for all project-related activities and issues, including data management.

2. How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

*Notes:* Indicate a succession strategy for these data in the event that one or more people responsible for the data leaves (e.g., a graduate student leaves after graduation). Describe the process to be followed in the event that the Principal Investigator leaves the project. In some instances, a co-investigator or the department or division overseeing this research may assume responsibility.

3. What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

*Notes:* This estimate should incorporate data management costs incurred during the project as well as those required for the longer-term support for the data after the project is finished. Items to consider in the latter category of expenses include the costs of curating and providing long-term access to the data. Some funding agencies state explicitly the support that they will provide to meet the cost of preparing data for deposit. This might include technical aspects of data management, training requirements, file storage and backup, and contributions of non-project staff.

## Ethics and Legal Compliance

1. If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

*Notes:* Consider where, how, and to whom sensitive data with acknowledged long-term value should be made available, and how long it should be archived. These decisions should align with Research Ethics Board requirements. The methods used to share data will be dependent on a number of factors such as the type, size, complexity, and degree of sensitivity of data. Outline problems anticipated in sharing data, along with causes and possible measures to mitigate these. Problems may include confidentiality, lack of consent agreements, or concerns about Intellectual Property Rights, among others. In some instances, an embargo period may be justified; these may be defined by a funding agency's policy on research data. Restrictions may be imposed by limiting physical access to storage devices, placing data on computers that do not have external network

access (i.e., access to the Internet), through password protection, or by encrypting files. Sensitive data should never be shared via email or cloud storage services such as Dropbox.

2. If applicable, what strategies will you undertake to address secondary uses of sensitive data?

*Notes:* Obtaining the appropriate consent from research participants is often an important step in assuring Research Ethics Boards that the data may be shared with researchers outside your project. The consent statement may identify certain conditions clarifying the uses of the data by other researchers. For example, it may stipulate that the data will only be shared for non-profit research purposes or that the data will not be linked with personally identified data from other sources. Read more about data security: [Data Protection: UK Data Service](#).

3. How will you manage legal, ethical, and intellectual property issues?

*Notes:* Compliance with privacy legislation and laws that may impose content restrictions in the data should be discussed with your institution's or company's privacy officer or research services office. Outline how compliance with applicable privacy laws will be achieved and what safeguards will be in place to protect privacy.

Research Ethics Boards are central to the research process. Include a description of any processes that the project has undergone with a Research Ethics Board and what the outcomes were in terms of requirements related to legal and ethical issues.

Establishing clarity on intellectual property data-related issues is also important. If applicable, include a description concerning licensing, and intellectual property rights to the data, including a description of any ownership structures for the collection, production and sharing of data. Document terms of reuse in line with the relevant legal and ethical requirements, as applicable (e.g., subject consent, permissions, restrictions, etc.). Your data management plan must also acknowledge and reflect the NCC's Intellectual Property Policy and the NCC's Cybersecurity Policy for UFRs.

## **Monitoring and Reporting Requirements**

### **Monitoring**

UFR are responsible for ensuring that project data are tracked and monitored on an ongoing basis, and in a manner that is consistent with their data management plans.

### **Reporting**

UFRs must submit progress reports, as required to their NCC project officer as part of their quarterly financial reporting. Specifically, UFRs are asked to indicate in their quarterly reports whether the project has encountered any data issues. If it has, they are then asked to suggest solutions for addressing them and to report on their resolution in the subsequent financial report. In addition, UFRs are required to summarize any data issues, how they were resolved, and any changes to their data management plans in the project's annual progress report.

*Please note that this document is subject to minor edits and is not in final format.*