# MCRI Project Data Management Plan: <Project Ref>

The [Data Management Coordinator](https://intranet.mcri.edu.au/Pages/Research/CRDO%20and%20CEBU/CEBU/CEBU.aspx) is keen to receive feedback on the content of the template, including ideas for new sections or for how the guidance might be improved.

Text in this grey Instruction style may be removed.

## Contents

[Contents 1](#_Toc32236253)

[1. This Document: Project Data Management Plan 3](#_Toc32236254)

[1.1. Project Details 3](#_Toc32236255)

[1.2. Document Details 3](#_Toc32236256)

[1.3. Related Documents 3](#_Toc32236257)

[2. Ownership of Data and Intellectual Property 4](#_Toc32236258)

[2.1. IP Ownership for this Project 4](#_Toc32236259)

[2.2. Considerations for Data or Materials Sourced from Outside this Project 4](#_Toc32236260)

[3. Electronic File Formats, Organisation and Metadata 5](#_Toc32236261)

[3.1. Electronic File Formats 5](#_Toc32236262)

[3.2. File naming conventions 5](#_Toc32236263)

[3.3. Version control 6](#_Toc32236264)

[3.4. Metadata, metadata standards, e.g. for variable naming 6](#_Toc32236265)

[3.5. Project network directory structure 6](#_Toc32236266)

[3.6. Standard Operating Procedures (SOPs) relating to data management 7](#_Toc32236267)

[4. Data Generation, Collection and Use 8](#_Toc32236268)

[4.1. Use of the data 8](#_Toc32236269)

[4.2. Data to be created 8](#_Toc32236286)

[4.3. Existing data 8](#_Toc32236287)

[4.4. Source documents 9](#_Toc32236288)

[4.5. Other associated materials 9](#_Toc32236290)

[5. Data Storage and Backup During the Project 10](#_Toc32236291)

[5.1. Storage of non-digital data 10](#_Toc32236292)

[5.2. Storage of digital data/electronic files during the project 10](#_Toc32236293)

[5.3. Study database 10](#_Toc32236294)

[5.4. Change control 10](#_Toc32236295)

[5.5. Backup 11](#_Toc32236296)

[6. Access to and Disclosure of Data During the Project 12](#_Toc32236297)

[6.1. Who should have access to the data? 12](#_Toc32236298)

[6.2. Who should have access to each type of restricted data? 12](#_Toc32236299)

[6.3. Specific precautions taken to secure personal or sensitive data 12](#_Toc32236300)

[6.4. Collaboration and data transfer 13](#_Toc32236301)

[6.5. Disclosure of data 13](#_Toc32236302)

[7. Quality assurance during the project 14](#_Toc32236303)

[8. Post-project - Data and Document Archiving, Custodianship, and Data Re-use / Sharing 15](#_Toc32236304)

[8.1. Minimum archiving period 15](#_Toc32236305)

[8.1.1. Destruction after minimum retention period (if applicable) 15](#_Toc32236306)

[8.2. Long-term custodianship (after archive period finished) 16](#_Toc32236307)

[8.3. Responsibilities 16](#_Toc32236308)

[8.4. Data Re-use and Sharing (Internal and External) 16](#_Toc32236330)

[9. Budget Impacts 18](#_Toc32236331)

[9.1. People 18](#_Toc32236332)

[9.2. Other costs or charges 18](#_Toc32236333)

## This Document: Project Data Management Plan

### Project Details

Short Title/Ref: <project reference, acronym>

Full Title: <full project title>

Principal Investigator (PI) / <project PI/CPI name>
Coordinating Principal Investigator (CPI) /
Sponsor-Investigator (*clinical trials*)

### Document Details

Version: <version of this document>

Date: <date of this version>

Author: <author of this document>

### Related Documents

List locations (with file names if known) of, for example:

Version of study protocol applicable for this DMP

Standard documents for your department or research group (e.g. Project Management Master Checklist Version for Community Child Health)

Standard Operating Procedures (e.g. for database testing and change management, for collecting, handling, storing and accessing data and/or biological samples)

Research Agreements, Memoranda of Understanding, Materials Transfer Agreements

## Ownership of Data and Intellectual Property

.

### IP Ownership for this Project

What is the ownership of the project’s data and IP? Typically this will be MCRI (see the [IP policy](https://intranet.mcri.edu.au/sites/policies/intellectual-property-policy)). For joint projects you will require a research agreement where (amongst other things) IP ownership will be defined (contact MCRI Legal for advice if needed).

### Considerations for Data or Materials Sourced from Outside this Project

Are any additional requirements specified by owners of externally sourced data?

Specify any boilerplate citation text or co-authorship credit, e.g. for

 REDCap (web-based database)

 DSS datasets such as LSAC

Ensure that there are signed agreements with other parties if data or samples are to be shared with external parties. At a minimum, an MCRI Material Transfer Agreement will be required. Contact MCRI Legal for advice. Refer also to the RCH Research Ethics Governance website <https://www.rch.org.au/ethics/>

Are licences required e.g. for use of standard measures such as

PedsQL

Pearson US BITSEA

## Electronic File Formats, Organisation and Metadata

Consider and document how your project team will organise your electronic data and documents.

###  Electronic File Formats

Refer to section 3 where you outlined the types of documents and data systems you expect to use.

For each item that will utilise electronic files, what file types will be involved? Is it possible to use a “standard” or open format that does not require specific hardware or software in order to access it? Consider both.

During the research, researchers should use the most suitable data management software according to planned analyses. The software selected must provide an audit trail (i.e. a trail that can be audited) of the original entry (identity of person entering data, date and time) and any amended data (identity of amender, date, time and reason for the amendment). Note that original data should remain visible.

Once data analysis is completed and the data have been prepared for storing, researchers should consider converting their research data to standard, interchangeable and longer lasting formats.

* Data files: e.g. raw data will be future-proofed by saving as a plain text data file (tab- or comma-delimited) plus a plain text syntax file (.do file for Stata, .sps for SPSS) rather than as a binary file that is specific to a particular stats package and version. Can you be certain you will be able to open a Stata 10 file in 20 years’ time, for example? **Data should not be entered and stored in software programs such as Word, Excel or Access - these do not provide an audit trail.**
* Documentation files: e.g. Word or Excel files. There are open equivalents to Microsoft file formats, but many people may well not be familiar with them. Another option is to store copies of files in an alternative format such as Adobe PDF-A (long-term durability – see ISO 19005-1:2005 and updates) or Rich Text Format (RTF).

For examples of recommended formats see the UK Data Archive’s Guide to Managing and Sharing Data. A link is available from the Resources section of the [Data Management Planning](https://intranet.mcri.edu.au/Pages/Research/CRDO%20and%20CEBU/CEBU/Data-Management-Planning.aspx) intranet page.

### File naming conventions

Define a convention for naming project documents (and use it!)

Suggested is a format that includes version numbering and/or date in year-month-day format (which has the dual benefits of being non-ambiguous and ensuring that name order = date order).

For example:

 MYPROJ\_ Data\_Management\_Plan\_2019-05-15\_LS\_draft.docx

 MYPROJ\_ Data\_Management\_Plan\_2019-06-28.docx

### Version control

How will document versions be labelled and controlled, particularly when more than one person is working on it?

Think about your data files and analysis/cleaning scripts as well as general project documents. Can you use version control software such as git?

### Metadata, metadata standards, e.g. for variable naming

The metadata is the documentation of your project methods, data collection and data preparation as well as a comprehensive code book. Consider what you should document and how it can be made available.

Define a convention for how project variables are to be named and coded (e.g. always use 0=No, 1=Yes)

Can you use a standard metadata format e.g. [DDI](http://www.ddialliance.org/) for social and behavioural science, [CDISC](http://www.cdisc.org) CDASH for clinical research?

### Project network directory structure

List your folders and state their function (if not obvious from the name!). Use README files within folders to explain what is stored there.

Refer to the CRDO **standard operating procedure on filing of essential documents for a study** (available for paper and electronic document filing) – check the [CRDO site](https://www.mcri.edu.au/research/facilities-resources-and-training/clinical-research-development-office-crdo).

CEBU have an internal SOP that specifies an appropriate directory structure for organising data files for analyses:

 \Analyses

 \Analysis\_name

 \yyyymmdd

 \Data

 \Analysis\_Dataset

 \Source\_Data

 \Do\_Files

 \Log\_Files

 \Results

 \Graphs

 \Paper

 \Statistical\_Report

 \Tables

### Standard Operating Procedures (SOPs) relating to data management

Part of the function of a project protocol is to provide sufficient detail on key areas to ensure that the conduct of the project is standardised among the members of the research team at a site, and between sites. The protocol can, and should, be supplemented with project-specific SOPs to ensure that all areas of project conduct are covered. With regards to data management, consider what SOPs you will need to develop. Examples are:

* Database testing/validation and change control
* Data collection
* Biological sample collection, handling and storage
* Operation of devices
* Data processing - how you plan to manage the data cleaning process from source data through to cleaned data files ready for analysis.

In addition to what is done and how, you need to think about who (i.e. which role/position) is responsible for carrying out each process.

Refer to the [CRDO site](https://www.mcri.edu.au/research/facilities-resources-and-training/clinical-research-development-office-crdo) for a template for creating and maintaining SOPs.

## Data Generation, Collection and Use

In this section, you need to address aspects such as: What will the data be used for? What sorts of data do you expect to be handling? What type of data is this plan covering and how will the data be kept confidential and private? Which data will your project create, and what existing data will you acquire from an internal or external source? What data and documents will be considered your primary sources? What additional materials or other documents or materials will be associated with the project data?

### Use of the data

Specify how the data will be used (e.g. for the purposes and analyses specified in the protocol ± the Statistical Analysis Plan).

Following the completion and analysis of the project, specify whether the data will be retained following the mandatory archive period for long-term use in future research projects.

### Data to be created

Summarise data sources, collection instruments and systems you are using or intend to use and specify who (i.e. the position) will collect the data:

Questionnaires (paper and/or electronic – if electronic, what systems/software/file types)

Assessments (paper and/or electronic – if electronic, what systems/software)

Standard measures

Multimedia sources such as video, audio, image

Externally sourced data e.g. lab results, previous projects

Tracking data e.g. recruitment, longitudinal follow-up, biological sample storage

Result sets – cleaned and derived data, analysis datasets

### Existing data

You want to collect data as close to the source as possible (N.B. *ensure that you have appropriate participant consent* *for access*). Again, specify who (i.e. the position) will have access to the resources to collect the data.

Local resource, e.g.

RCH electronic medical record (EPIC)

MCRI research project

RCH/MCRI repository/databank

RCH clinical database (RCH Ethics maintain a register of such databases: include Database Registration Number, if one exists, and the Data Access Request reference, when received).

External resource, e.g.

Non-RCH/MCRI medical record/repository/databank

Non-MCRI research project

### Source documents

For each type of data you will be collecting or generating Identify the source, or provide here a reference to a separate Source Document Plan (see template located on the [CRDO](https://www.mcri.edu.au/research/facilities-resources-and-training/clinical-research-development-office-crdo) website).

### Other associated materials

State any other documents or materials that will be associated with the project data, such as:

* Code book, data dictionary (i.e. providing a detailed description for each data variable - the source of the variable, external coding information if used e.g. MedDRA, and expected ranges [if relevant]). You can also export this information from the project database(s) (e.g. REDCap).
* Standard Operating Procedures for aspects not already referred to in this DMP (e.g. SOPs for collecting/processing biological samples, data access requests)
* Statistical Analysis Plan

## Data Storage and Backup During the Project

Obvious, but important, is considering where to store your data. Highest security, most restricted access is what you aim for, but consider your accessibility requirements: who needs access to what data and when? There will be different answers to these questions for different data/document types/purposes.

### Storage of non-digital data

Where will you store hard copies of paper documents such as questionnaires, consent forms?

Consider both blank and completed versions of documents

Will you be collecting biological samples? What are the associated protocols and procedures, and where can they be found?

### Storage of digital data/electronic files during the project

How and where will the following be stored? How will access be restricted?

Data files

Documentation files

Is a shared folder on a network file server sufficient? Does anyone require remote access via VPN or a web-based document management system or database? Note that folders in your Group drive may be shared with external people using OwnCloud. Contact IT for assistance if needed.

### Study database

What software package(s) will be used for processing and storing study data or for study management. During the research researchers should use the most suitable data management software according to planned analyses. The software selected must provide an audit trail (i.e. history) of all data entered indicating the identity of person who entered or updated the data, and the date and time). Once data analysis is completed and the data have been prepared for storing, researchers should consider converting their research data to standard, interchangeable and longer lasting formats.

See the list of examples below (see also the [Data Management Software](https://intranet.mcri.edu.au/Pages/Research/CRDO%20and%20CEBU/CEBU/Data-Management-Software.aspx) pages on the CEBU website).

[REDCap](http://intranet.mcri.edu.au/?page_id=32813)

[WebSpirit](http://intranet.mcri.edu.au/?page_id=32817)

OpenSpecimen

[EpiData](http://intranet.mcri.edu.au/?page_id=32818)

### Change control

Describe your process for managing and controlling changes to the project database(s) or reference an external SOP document where this process is described. For example, changes to the REDCap trial database will be developed and tested in a copy of the main trial database. Once complete and acceptance by the [Study Coordinator/PI?] is documented the [Study Coordinator] will merge the changes from the copy into the main trial database. The copy will then be deleted.

### Backup

What is the backup regimen for the location(s) where digital data is stored? See details of MCRI/RCH file server backup regimens on the CEBU [intranet page](https://intranet.mcri.edu.au/Pages/Research/CRDO%20and%20CEBU/CEBU/Back-Up.aspx)

Consider local solutions such as OwnCloud.

## Access to and Disclosure of Data During the Project

### Who should have access to the data?

Access to all data and information should be restricted to those who require access in order to complete project-specific tasks. This includes the immediate project team but also specified personnel outside the immediate project team (e.g. statistician, sponsor, HREC, Research Governance Office and regulatory agencies). However, the **level of access** should be determined for each role (i.e. within and beyond the immediate project team) in order to protect participant confidentiality and privacy. Implement the “Principle of Least Privilege” – personnel need sufficient privilege to perform their tasks, and no more.

Consider each of the different sources of data listed in sections 4.2 and 4.3.

Will your project contain data in the following categories?

* Personal information (data that could potentially be used to identify a specific individual, such as demographic data – dob, URN, address etc. but also rare diseases when combined with even broad-scale location data)
* Sensitive information (data on a participant’s the racial or ethnic origin, religious, political or philosophical beliefs, sexual preferences, criminal record, disease/diagnosis (e.g. HIV), consanguinity, indigenous status etc.)
* Other culturally sensitive information (e.g. recordings of deceased indigenous persons)

Consider also the conditions of use for access to external data.

### Who should have access to each type of restricted data?

Detail access permissions by role - you do not need to list names of individual personnel.

In the next section, detail how you will assign user permissions.

### Specific precautions taken to secure personal or sensitive data

Specify the precautions taken to comply with your responsibilities for protecting such data e.g. maintain personal data in a separate database to questionnaire response data.

The following are some examples for electronic data.

*General strategies*

* minimising the number of variables collected for each individual;
* separation and separate storage of identifiers and content information; and
* separating the roles of those responsible for management of identifiers and those responsible for analysing content.” (NS 3.1.41)

*Specific strategies*

* MCRI Network directory permissions – access to a shared network drive is controlled via MCRI username and password. Note that you can submit a request to MCRI IT to limit access to folders and sub-folders to specific project personnel.
* REDCap
	+ Establish a REDCap project database - This can be set up on request. Access to the REDCap database is provided on request via an MCRI user account or (for external collaborators) via a REDCap user account created by the MCRI system administrator. The permissions granted to each user within each REDCap project are controlled by, and are the responsibility of, the project team member delegated this task by the Principal Investigator. REDCap has functionality that makes adding and removing users and managing user permissions straightforward. Personnel entering or editing data in the project database must read the instructions and sign a training log signed prior to access.
	+ Within the REDCap project database, you can restrict user permissions to particular variables (e.g. those containing personal identifiers). This also allows you to exclude personal identifiers from when sending database extracts to personnel outside the immediate project team (e.g. statistician) (use the ‘remove tagged fields’ option when exporting). Individual records can be identified by a unique code assigned to the participant for the project.

For precautions to be taken with hardcopy data, refer to the general strategies for electronic data and indicate where and how hardcopy data will be stored and access restricted.

### Collaboration and data transfer

Will there be documents and/or data files to send to and receive from external collaborators?

How will this be achieved? Email? OwnCloud? Dropbox? **\*NEVER USE AN EXTERNAL SYSTEM FOR CONFIDENTIAL DATA\***

### Disclosure of data

Describe whether there are any situations in which personally identifiable information or data will be released to third parties. For example:

* “The study protocol, documentation, data and all other information generated will be held in strict confidence. No information concerning the study or the data will be released to any unauthorised third party, without prior written approval of the sponsoring institution. Clinical information will not be released without written permission of the participant, except as necessary for monitoring by the HREC, Research Governance Office or regulatory agencies.”

## Quality assurance during the project

Data should be attributable, legible, contemporaneous, original, accurate (ALCOA) as well as enduring, and complete; these terms are explained below. Any changes should be traceable, should not obscure the original entry, and should be explained if necessary (e.g. via an audit trail).

* **Attributable**: The sources of the data are known and recorded.
* **Legible**: The data are human readable.
* **Contemporaneous:** The source data are recorded when they are generated.
* **Original:** All data come from the original source.
	+ Copies and transformations of the data are accurate and complete, do not overwrite original data, and are traceable back to original data.
* **Accurate:** The data are correct.
* **Enduring:** The data are available for the entire time they are required to be kept.
* **Complete:** All available data are included.

REF: <http://www.ofnisystems.com/clinical-data-validation/>

Describe:

* How you will perform and document the testing of project databases to ensure that they meet the requirements of the Protocol, a requirement of Good Clinical Practice (GCP)?
* Plans for data cleaning (e.g. checks for missing data, invalid characters, out-of-range values, invalid dates, data that is not consistent with data in other data fields, repeated participant IDs etc.).
* Plans for source data verification (where applicable) to assess the accuracy, completeness, or representativeness of data by comparing the data in the database to the original source of the data (not applicable for data items where data is entered directly into the database and therefore the database is also the source).
* Plans for site monitoring and auditing (where applicable)

##

## Post-project - Data and Document Archiving, Custodianship, and Data Re-use / Sharing

### Minimum archiving period

What is the minimum, mandatory retention period for the data for this project? The time period for which study data, information and documents must be retained (the archive period) is determined by the type of research and relevant legislation, code and guidelines.  Where more than one legislation/code/guideline is relevant, the one with the longest retention period applies. **However, also keep in mind the importance placed in the updated (2018) National Statement on collecting and retaining data and information for use by future research projects so that the benefits of research can be shared [NS 3.1.50).**

Below is some guidance on current minimum retention requirements for research in Australia - contact the RCH Research Ethics Governance group to further discuss the requirements for your particular study. You must also comply with [MCRI’s data management policies](https://intranet.mcri.edu.au/Pages/Research/CRDO%20and%20CEBU/CEBU/Data-Management-Policies.aspx).

* All research – in general at least 5 years from publication (The Australian Code for the Responsible Conduct of Research 2007)\*
* All research – retention of any new health data for at least 7 years for adults or until age 25 for children [VIC HRA]
* Clinical trials - must archive for at least 15 year post-trial completion (TGA) or until child aged 25 years (whichever is the later) (VIC HRA)
* Gene therapy research data - must retain permanently (The Australian Code for the Responsible Conduct of Research 2007)\*
* Research that has community or heritage value - must retain permanently, preferably within a national collection (The Australian Code for the Responsible Conduct of Research 2007)\*

\* The revamped Code (2018) does not include guidance on required data retention periods - we are awaiting the release of the supporting guide “Management of Data and Information in Research”.

Describe how long and where all research data, information and documents will be kept following the end of the study. During the archive period, data should be stored in a way that allows re-identification in case this is needed (e.g. for regulatory audits). Outline how the data will be secured and how confidentiality of stored data will be ensured.

What data and documents are to be placed in MCRI's figshare system? https://mcri.figshare.com/ This is a place where project documents, datasets, publications etc. can be archived and published, either publically or privately. Private publication might be desirable where there copyright restrictions apply (e.g. for journal articles) or (e.g. for datasets) where an access request process applies (e.g. for datasets).

State who (i.e. person’s position) will be the custodian during the archive period, who will have access to the stored data and outline any procedures that may be followed to dispose of the data at the end of the archival period.

Where will electronic and non-electronic data be stored long term?

What are the retrieval implications (e.g. lead times, costs)

###  Destruction after minimum retention period (if applicable)

The National Statement specifies that research data should be retained and made available for future research projects, except where there are justifiable ethical reasons. Indicate the plan for long-term data retention for this research project.

If the plan is to destroy data and documents after the required archive period, state this here and specify that records should not be destroyed without the written consent of the Site Principal Investigator / Coordinating Principal Investigator\* / Sponsor-Investigator\*.

\*In multi-site studies, the Coordinating Principal Investigator / / Sponsor-Investigator should inform the Site Principal Investigators when these documents no longer need to be retained.

Describe the planned method of destruction. Secure destruction of research data involves using irreversible methods to ensure that the data is no longer usable. It is particularly critical that confidential or sensitive data is made unreadable.

* Hardcopies should be disposed of via a confidential shredding process.
* For electronic data, note that deleting files does not destroy the information completely; it may be necessary to utilise software which permanently erases data\* (Seek guidance from MCRI IT). Consider also other data devices.

\* It may not actually be possible to completely expunge data from institutional backups [i.e. back-up tapes held off-site].

### Long-term custodianship (after archive period finished)

As noted above, the National Statement specifies that research data should be retained and made available for future research projects, except where there are justifiable ethical reasons. Indicate the plan for long-term data retention for this research project. After the archive period, the data may be anonymised for preservation to reduce the risk of re-identification. As outlined previously, technological advances mean that identification can occur even where data and/or information has never been labelled with individual identifiers or from which identifiers have been permanently removed (e.g. linking to other data sets that contain identifiers). The risk of re-identification is related to the data context as well as what it will be used with and for.

### Responsibilities

Think about how responsibilities should be passed on at the completion of the project.

State who (i.e. person’s position) will be the long-term custodian following the archive period.

### Data Re-use and Sharing (Internal and External)

Will data be made available to other researchers? Under what conditions? As mentioned previously, the National Statement requires research data be made available for future research projects except where there are justifiable ethical reasons, (see extract below). Indicate the plan for whether data will be shared following completion, analysis and publication of this study; justify if the plan is not to share the data.

* “In the absence of justifiable ethical reasons (such as respect for cultural ownership or unmanageable risks to the privacy of research participants) and to promote access to the benefits of research, researchers should collect and store data or information generated by research projects in such a way that they can be used in future research projects. Where a researcher believes there are valid reasons for not making data or information accessible, this must be justified.” (NS 3.1.50).

Data sharing statements should indicate the following: whether individual de-identified participant data (including data dictionaries) will be shared; what data in particular will be shared; whether additional, related documents will be available (e.g. study protocol, statistical analysis plan, etc.); when the data will become available and for how long; by what access criteria data will be shared (including with whom, for what types of analysis). Ensure you seek appropriate consent (i.e. extended or unspecified consent) for this.

Note that MCRI is currently (as of July 2019) developing a policy/procedure for the process involved in sharing data for future ethically-approved research. This will include aspects such as:

* Discoverability (e.g. how will people know your project exists and how will they find out more? Will it be via project/metadata registration e.g. MCRI Research Repository, RCH Research Ethics Governance office, clinical trial registries (e.g. clinicaltrials.gov)?
* Approval (i.e. who will receive the requests and who will approve release?)
* Transfer – How will data be transferred? Might it be via OwnCloud, REDCap SendIt, dedicated REDCap project, email (with password protection e.g. using 7Zip) or other service?

In the meantime, see below for an example of a data sharing statement which covers some aspects of data sharing. Aspects such as who will approve the access and how it will be transferred have not yet been included. Review and customise for your study.

*“Beginning ‘x’ months following analysis and article publication, the following will be made available long-term for use by future researchers from a recognised research institution whose proposed use of the data has been ethically reviewed and approved by an independent committee and who accept MCRI’s conditions for access:*

* *Individual participant data that underlie the results reported in this article after de-identification (text, tables, figures and appendices)*
* *Study protocol, Statistical Analysis Plan, PICF”*

## Budget Impacts

### People

Does your team include people with the skills required to achieve what is required?

Sufficient organisational capabilities

Sufficient time and expertise for database development and maintenance

Experience and capabilities in data management (!)

Do you need a Data Manager as well as an RA?

* Include provision for data management activities in grant applications

### Other costs or charges

Do any of the items above have cost implications?

Software licences

Data storage capacity, backup charges

Web hosting, SSL certificate, SMS service

Database development

Purchase or rental of hardware such as iPads, laptops, audio equipment

If so, how much, how frequently and over what period of time?

For example, storage in your group drive (G:\) is “free” but may have per-project capacity limit. Extra MCRI storage can be negotiated with IT.

Is there scope to recover any of the costs later?

 Sale of equipment

 IP, commercialisation