Data management and data sharing

1. Data Management Plan

An effective data management plan in the behavioural and life sciences must address the following principles:

- Protection of human subjects - Show evidence of a disclosure risk limitation strategy that addresses respondent privacy and confidentiality across the research data life cycle, taking into account not only informed consent and the IRB submission for data collection, but also evaluation and limitation of disclosure risk in the final analytic data files. If a restricted-use file will be created, detail the administrative controls on data access and use that are planned.
- Comprehensive standardized documentation - Provide a plan for creating study and variable level metadata that documents all variables, including derived, imputed, and recoded items, in a clear and transparent way. Documentation should conform to international documentation standards in the scientific field in which data collection occurs and should facilitate interoperability and metadata reuse, avoiding the use of proprietary software.
- Enduring access - Present a recommendation for widespread and fair access to the data for all eligible users. Criteria for eligibility and the legal terms of data re-use should be clearly stated in the plan along with any licensing proposed. The technology underlying the dissemination mechanism should be designed to ensure stable, continuous, and secure access to the data files.
- Long-term preservation - Provide a plan for preserving the data over time and ensuring that the data are migrated as appropriate and kept usable for a reasonable period. This plan can incorporate a strategy that moves the data to a digital or institutional repository, like DSDR. Any barriers to archival storage of data in specific locations should be noted.
- Usage metrics - Outline a method for tracking data use and the characteristics of data users. Effective data sharing must address the needs of the secondary users. In the absence of information about the volume and nature of the user community, data dissemination strategies may be costly and ineffective.

Data arising from research projects should adhere to FAIR principles. It should be:

- Findable
- Accessible
- Interoperable
- Reusable/Reproducible

Your data management plan should help you to achieve this, by ensuring that your data are appropriately annotated with the necessary metadata, in standard formats and submitted (where appropriate) to a public repository in a timely fashion. This can be achieved in a data dictionary.

There are a number of technical areas in a data management plan that require specific input about the types and volumes of data that you intend to generate, the formats you will store the data in,

the associated standards that will be adhered to within your metadata to ensure that your data are understandable and fully re-useable, and the appropriate public repositories that you may use to disseminate your data in the public domain. These areas tend to be highly discipline-specific. The Life Sciences and Biomedical data areas are particularly rich in the numbers of available public repositories, data standards and recognised data formats.

2. Data access management

Best Practices for Data Access Management:

- Create a complete inventory of your users and resources and keep it up to date.
- Work with department managers and other business owners to determine where your sensitive data is and who owns it.
- Determine who has access to what and who owns what data in your collaboration. For example, you can export access lists of your file servers using PowerShell scripts or third-party software.
- Establish a security structure by creating security groups and making users members of the appropriate groups.
- Assign each group appropriate access to shared data.
- Empower data owners to control access rights to the data they own.
- Audit the actions of data owners to be sure that all operations are authorized.
- Establish an access request workflow (such as a request portal) so users can easily request access to data they need to do their jobs.
- Audit and report on access to sensitive data as well as changes to it.

3. Data ownership

- You have authority within your organisation to share the data
- You have ensured that there is no personal information
- You have ensured that there are no legal restrictions preventing the sharing of the data
- You have conducted due diligence to ensure your agency owns the data

4. Standards and procedures for data sharing

One possibility is to assign a data steward who is charged with the role of ensuring the Data Access and Sharing Policy is followed within their area of responsibility.

Data steward responsibilities include:

- Define restricted data for department/unit, collaboration.
- Ensure employees within department/unit or collaboration are trained on expectations for restricted data.
- Oversee that restricted data is limited to those with authorized roles in a 'need to know' responsibility.
- Perform annual internal review to confirm appropriate user access with respect to restricted data being used within unit/department or collaboration.
- Ensure operational procedures adhere to outlined access, transmission and storage protocols for restricted data.
- Resolve stewardship issues and use of data elements that cross multiple operational units/departments or collaborations.
- Understand laws, regulations, retention requirements that are specific to data assigned to Data Steward.

Data Classification Protocols		
	Restricted Data	Unrestricted Data
Data Classifications	Data is classified as “restricted” if data protection is required by federal or state law/institutional policy and/or data is defined as restricted by Data Steward. Examples: SSN, Credit Card data, Protected Health Information	Data is classified as “unrestricted” if it is not considered to be restricted. Examples: Admissions Requirements, Institutional Report
Access Protocol	Data access is limited to those with authorized roles in a ‘need to know’ function.	At the discretion of the data steward, anyone may be given access to unrestricted information. However, care should always be taken to use data appropriately and to respect all applicable laws. Data that is subject to copyright must only be distributed with the permission of the copyright holder.
Storage Protocol	Electronic restricted data is to be stored only on OneDrive. Electronic restricted data is not to be stored on C: drive, nor on removable media. Restricted data in paper form should be secured via secure print at multi-function print stations.	No storage requirements.
Identifiable Human Subjects Research Protocol	Identifiable Human Subjects research data. Any human subjects research data set containing data elements that would allow the human subjects/participants to be identified is considered restricted data, and must conform to the	De-identified Human Subjects research data are not considered restricted data for the purposes of this policy. De-identified means that the information does not identify an individual, and there is no reasonable basis to

Data Classification Protocols		
	outlined access, transmission and storage protocols outlined within this policy.	believe that the information can be used to identify an individual. Information is considered de-identified under this policy if no code exists enabling the linkage of the identifying information to private information or specimen. Coded Human Subjects research data are not considered restricted data for the purposes of this policy, so long as the code and the data are separately stored.
Institutional Research data	Sexual assault survey results, alumni data survey, institutional reports	Research Coordinator
Sponsored Projects and Research data	Employee history, financial conflict of interest in research screening and disclosures	Research Coordinator

5. Data protection and security

Data protection and security checklist

- Cloud backup is preferred.
- Archiving data separate from backup

- Disaster recovery plan for backup restoration
- Creating snapshots and replication for faster recovery
- Implement the before mentioned actions into a Continuous Data Protection (CDP) plan.

6. Data exploitation strategy

- **Planning:** establishing the vision and requirements followed by agile development and use of a framework such as Scrum. Complex requirements are captured, distilled into planned incremental updates through to full delivery, maintain a focus on an end user minimum viable product or setup that can leveraged as soon as possible and a style that is flexible and effective for the organisation you are working with.
- **Training:** Training packages to support the needs of varying levels of users and situations. This can be classroom based, coaching and mentoring, upskilling around problem solving and delivery of activity based workshops to enable practical approaches to problems.
- **Setup and migration:** Understanding the best solution for each situation, deciding on the most appropriate setup and take this through to delivery either by way of migration from a previous setup or by creation of a new platform. Understanding of the practicalities and technicalities of migrations.
- **Quality assurance and performance testing:** The use of advanced analytics. Reduce risk, maintain clear lines of contact with all stakeholders, perform rigorous, well documented testing at every stage of delivery with client stakeholders and users and operate in a transparent agile fashion such that everyone is aware of progress.
- **Security Testing**