Marquette University

Research Data Management Plan Guide and Tool Kit

TABLE OF CONTENTS

1. PURPOSE

The purpose of the Tool is to provide guidance for Marquette University (**MU**) students, faculty, and staff on how to develop a data management plan for a research project and handle data as part of their MU research. Research data management plans (**DMPs**) are critical because they help to manage data risks, such as data loss and data corruption, along with user access and data sharing. Data includes measurements, facts, statistics, and other information collected for reference or analysis.

Most research datasets are sensitive and are being analyzed in a variety of ways to capture the newest insights in ways that could affect technology development, medical practices, or people's daily lives. Data theft is real and is happening. However, data theft is not the only concern—so is inadvertent modification of data or the disclosure of data. If data cannot be stolen, tampering can be the next best thing to disrupting research, because if the integrity of a researcher's data is lost, the researcher's reputation and findings can be tarnished.

Data management plans are also helpful as they aid researchers in organizing their data and can foster collaboration. It is important to know about data management options to pick the best tool for use. This Research Data Management Plan Guide and Tool Kit assists Marquette researchers in establishing a culture of data security and management to ensure that best practices and common tools are used to reduce the likelihood of data being lost, corrupted, tampered with or inadvertently disclosed to protect the research.

If assistance is needed with this guide, please submit your question to:

- Data_management@marquette.edu

2. DATA CLASSIFICATION AND AVAILABILITY

2.1. WHAT IS DATA CLASSIFICATION

Data classification is the process of analyzing data and organizing it into categories. MU relies on electronic and physical information to conduct its research operations and to achieve its research objectives. To advance the university's interests and protect the privacy of its students, employees and patients, Marquette must protect its sensitive and critical information.

For research purposes, there are four general categories of data that are used and created:

- **Category 1**: Publicly Available Information: This includes publications, unrestricted databases and other materials available on-line that are not identified as confidential by the provider of the information.
- **Category 2**: Compiled or Generated Restricted: This data is generated during research or compiled as part of the research. This may include data sets created through measurements of materials, insects or animals, compiled from publicly available information, lab notes, images, survey results.
- **Category 3**: Compiled or Generated Controlled: This is data that is not only confidential, but for which access is also controlled by federal or state law, such as personally identifiable information, personal health information, student records, or controlled unclassified information.

- **Category 4**: Licensed Data: This is data that has been licensed from a third party for use in conducting a research project. The use, protection and disclosure of the data will be controlled by the license agreement.

2.2. WHAT IS DATA AVAILABILITY?

Data availability refers to the timeliness and reliability of access to data. To develop the DMP, it is important to first understand who needs access, when they will need the access and from where they will access the data. The following items should be considered when managing data availability:

- Whether data needs to remain accessible constantly (e.g., 24 hours 7 days a week)
- Who needs to have access to data (e.g., are they internal to MU or external collaborators)
- Should access that is granted to collaborators be the same or different
- Which storage methods and locations meet accessibility needs
- Whether data is easily re-acquired or re-created if lost, stolen, or destroyed

Understanding what categories of data will be generated and who requires access to the data is the first step in developing the DMP as there are different levels of risk and data exposure depending on access.

For example, if all research activity and data is maintained within the MU environment and all collaborators are MU faculty and staff there is a lower likelihood of data risk or exposure due to knowing the environments the data resides. On the other spectrum, there is a higher level of data risk and exposure when research is completed with a variety of researchers throughout the world, and each researcher has different access profiles and stores their data in a variety of cloud and local storage solutions. Data risk and exposure becomes even more important to understand when factors such as sensitive data or export controls are part of the data environment.

3. RESEARCH DATA LIFE CYCLE

3.1. WHAT IS THE RESEARCH DATA LIFE CYCLE?

The research data life cycle is the process that a researcher takes to complete a project or study from its inception to its completion. Research today not only has to be rigorous, innovative, and insightful; it must be organized. As improved technology creates more capacity to create and store data, it increases the challenge of making data FAIR: Findable, Accessible, Interoperable, and Reusable (*The FAIR Guiding Principles for scientific data management and stewardship*).

To address the challenge of making data FAIR, select funding agencies require data management plans to be submitted with grant applications. Additionally, many academic journals require the submission of relevant data with manuscripts to promote open access and reproducibility of research. This link is to a data management cautionary tale: https://www.youtube.com/watch?v=66oNv_DJuPc&t=161s

A research data life cycle typically consists of the following phases:

- Phase 1: plan and design: DMP development and the data organization plan

- Phase 2: collect and create: Data gathering and development of organizational foundations
- Phase 3: analyze and collaborate: Reviewing of data to develop informed conclusions
- Phase 4: store and manage: Process for securing and maintaining data
- Phase 5: evaluate and archive: Assessment of data cleansing and data retention
- Phase 6: share and disseminate: Repository storage and publication communications
- Phase 7: access and reuse: Access permissions to data and how it can be used

3.2. PHASE 1: PLAN AND DESIGN: CONCEPTS

The Data Management Plan (DMP) documents the process of organizing a researcher's data as they progress though the research data life cycle. The DMP may need to be altered as the course of research changes. A DMP is critical to help think through the entirety of the research data life cycle, and should consider:

- Type and format of data being used
- How data security will be managed throughout the project
- Assessment of current and future data costs

The standard tool to be used, no matter the data classification type, is the DMP Tool, which is supported by the Raynor Memorial Libraries. The DMP Tool should be used to help develop a DMP and address at least the following areas:

1. What type of data is used (e.g., electronic, video, paper, audio, etc.) and what is the data format and average size of the data file? [Phase 1: plan and design]
2. What are the data naming conventions/meta-data that will be used when storing data to ensure version control is managed? [[Phase 1: plan and design]]
3. Where is information being obtained from and are there any special requirements for housing the data due to the data's sensitivity or the data owners' requirements? [Phase 2: collect and create]
4. Where is the data being stored (e.g., locally, third party web storage, campus-managed network drives or software) [Phase 4: store and manage]
5. Is collaboration with other people at the university or with other external resources going to happen? If so: [Phase 4: store and manage]
    a. What are the roles, responsibilities, and tasks for each person involved?
    b. What is the process for granting and removing user access?
    c. Who has the authority to grant or remove user access?
    d. What is the expected data availability for data collaboration (e.g., on demand, one person at a time, or multiple users in a file at the same time)?
    e. Does everyone need the same level of data access?
    f. Who owns the data?
6. Are there data retention standards being used? If so: [Phase 5: evaluate and archive]
    a. What is the established frequency of storing data before destruction?
    b. Is any data being stored for a period of time more than the standard data retention policies [UPP 1-12: Records Management]
    c. Is keeping data too long a liability?
7. Is there a process for destroying data? If so: [Phase 5: evaluate and archive]
    a. What is the process for physical documentation?
    b. What is the process for electronic data?
    c. What is the process for physical hardware?
8. Where is the final research (e.g., findings, data populations, analysis) being stored? [Phase 6: share and disseminate]
9. Should access for everyone on the project be the same? If not: [Phase 6: share and disseminate and Phase 7: access and reuse]
    a. Who will manage user access?

    b. What are the various user roles and what access does each role need?
10. Is there a data backup plan? If so: [Phase 7: access and reuse]
    a. Where is the data being backed up?
    b. How frequently will the data be backed up?
    c. Who owns/performs the data backup process?
    d. How often will the backup data be verified as usable in case a restore of data is necessary?
    e. What controls are in place to ensure that data is not inadvertently overwritten?
11. What security protections exist for physical documentation or research done on campus premises (e.g., labs, physical lab notebooks)? [Phase 7: access and reuse]

To manage data across a project and a team, a checklist has been developed to outline the important steps for onboarding a new researcher / trainee to a new project. The Research Data Management Onboarding Checklist is a general guide that focuses on policies and resources; refer to section 7 Research Data Management Onboarding Checklist.

3.3. PHASE 1: PLAN AND DESIGN: TOOLS

In this phase the tools are all standard as these steps are critical to the organizational steps of a project, no matter the data classification or sensitivity.

The most critical tool in this phase is the DMP Tool, which helps in the development of the data plan. The tool is used to develop a DMP as part of the grant application process. Use of the DMP Tool is a higher education best practice used by many fellow universities and other submissions of data management plans can be accessed. After initial creation of the DMP using the DMP tool, the DMP should be continually reviewed as a researcher progresses through the research life cycle. This ensures the DMP is still accurate or it is being modified as appropriate.

Any questions on how to use the DMP Tool should be directed to the Head of Research and Instructional Services or the Coordinator of Digital Scholarship and Programs, at the Raynor Memorial Libraries.

3.4. PHASE 2: COLLECT AND CREATE: CONCEPTS

Data should be collected and organized in accordance with the project's organizational foundations to ensure data is properly managed. Organizational foundations, such as naming conventions, directory structures, and version control procedures, should be developed. Organizational foundations are essential for communicating the data amongst team members and reducing the likelihood of data loss.

3.4.1. NAMING CONVENTIONS

A naming convention is a framework for naming files in a way that describes what they contain and how they relate to other files. Naming conventions help a researcher stay organized and make it easier to identify files. Naming conventions should be used no matter the data classification. By consistently organizing files, a researcher can quickly find what is needed. It also allows for easier collaboration in a group file-share setting, if applicable. It is essential to establish a naming convention before beginning to collect files or data to prevent a backlog of unorganized content that can lead to misplaced or lost data.

While the tips below are examples of naming conventions there is no single formal standard that must be adhered to. In fact, most disciplines recommend their own standards.

Naming convention tips:

- Do not use research participant names in file name titles
- Ensure standardization of shortened information, dates, or file groupings
- What files will a researcher want and how will the data be grouped or searched for?
    - For example: Identify what group of files a naming convention will cover and use different naming conventions for a different file group. Additionally, what type of data should be at the beginning of a naming convention for easy identification.
- What type of metadata may be excellent to capture?
    - Identify the most critical sorting type(s) for a research project: experiment conditions, data type, research name, lab location, date range of experiment (YYYYMMDD), experimental number, sample ID
        - Note: Two- and three-letter abbreviations can help to reduce naming convention length while also allowing sorting to occur.
- What types of characters will be used to separate metadata fields?
    - Symbols that could be used: dashes, periods, underscores, camel casing (upper letter for each section (e.g., FileName.xxx)
        - Note: Avoid using special characters: ~ ! @ # $ % & * ( ) ' ; < > ? [ ] { } ` "
- How will versioning be tracked, within a name or with software?
    - Option 1: Track versions by adding the version number at the end (e.g., _v02)
    - Option 2: Track versions by adding the version date at the end (e.g., _YYYYMMDD)
    - Option 3: Leverage software that auto tracks versions, such as SharePoint, reducing the need for copying files and adding new versions.
- How will naming conventions be formally documented: a ReadMe.txt file, DMP, or the onboarding checklist for the project?

- Example: My naming conventions for this group of labs: [SA-LAB-ID]_[YYYYMMDD]_###_Status.TIF_Version xx

Naming convention examples
Here are a few examples of naming conventions. Supplemental guidance to the naming convention guidance in this section can be found in section 9 File Naming Checklist.

- 20160104_ProjectA_Ex1Test1_SmithE_v1.xlsx
- 20160104_ProjectA_MeetingNotes_SmithE_v2.docx
- ExperimentName_InstrumentName_CaptureTime_ImageID.tif

3.4.2. DIRECTORY STRUCTURES

Researchers are advised to structure their folders (whether paper or electronic) to correspond to how the records were generated and to complement proposed or existing workflows no matter the data classification. Filing structures enable researchers to be more transparent, make it easier for researchers to determine where files should be saved, and ultimately make retrieval and archiving more efficient. Before collecting or working with data, a researcher should decide the structure, allowing for standardized data collecting and analysis by team members.

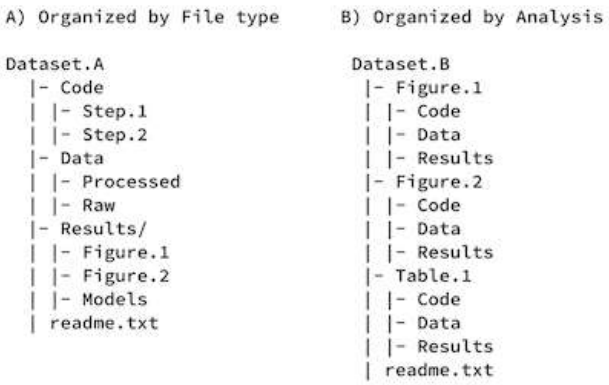


Image: Dryard FAIR Data practices to organize files in a logical schema

One of the most common ways to sort files is by function. Since all records must be retained for a certain period (both to meet university recordkeeping requirements and satisfy grant-mandated retention periods), maintaining records functionally will enable records retention periods to be assigned to groups of records. Tips for organizing data are:

- Organize data hierarchically and identify ways to divide data into categories/attributes (e.g., project, time, location, file type)
- Organize files within folders by a common methodology, such as chronically, classification code, or alphabetically
- Name folders and subfolders to reflect the content of the folder, not the names of researchers, staff, or research participants
- Document the file directory structure and describe the kinds of records that should be maintained in those folders to ensure compliance
- Leverage naming conventions to organize folders, subfolders, and data content

3.4.3. VERSION CONTROL PROCEDURES

Version control is a method used to track the changes to a file or a file set over time so that a researcher can recall older versions. Version control should be used no matter the data classification. File versioning can be as simple as adding a version number to the end of a file

name such as _v01 or _v02, or version control software can be used, such as SharePoint or GitHub. Version control software allows multiple people to work on files at the same time, have the data be saved based on different versions, and recall prior versions, if necessary.

3.4.4. README FILES

A ReadMe file is a clear and concise description of all relevant details about data collection, processing, and analysis and allows others to interpret and reanalyze datasets. ReadMe files main purpose are to explain file naming conventions and to provide clarity on files/data being deposited into repositories. However, it is best practice to create ReadMe files for each dataset regardless of data classification and whether or not the data is being deposited in a repository. ReadMe files should be recorded as a plain text file or PDF file to avoid any proprietary formats that may require certain software to read the data, such as Microsoft Word.

The following provides guidance for the development of a ReadMe file:

- Cornell University's Research Data Management Service Group: ReadMe Template
- ReadMe template: Refer to section 10 ReadMe File Checklist

3.5. PHASE 2: COLLECT AND CREATE: TOOLS

The collect and create phase is focused on obtaining the data and establishing the organizational foundations, such as the data naming organization and directory structures. There are no formal tools that should be used when developing the organizational foundations of the project. Instead, a researcher should reference the phase concepts above.

From the perspective of risk management, the critical area of concern is obtaining the data. Data can be gathered in different ways, and the following provides guidance and the preferred tools for data transferring and data surveying.

3.5.1. DATA TRANSFERRING

When there is Category 2, 3 or 4 data, the researcher should contact IT Services to determine the best way to obtain the data. Data transferring may occur through an application programming interface (**API**), external hard drive transfer, or secure-file transferring over the internet. In most cases a data exchange using SharePoint will be acceptable, but a researcher needs to be aware the data sensitivity, who can access the SharePoint site, and if the data will remain in the SharePoint site. When data is public, it is acceptable for a researcher to use an external hard drive or USB drive to obtain and transfer the data.

It is critical for a researcher to ensure the data being obtained is from a reliable source and that the data does not contain any malicious code or bugs. It is recommended that data be scanned by antivirus software before the data is stored on the university network. If the data is determined to be malicious, IT Services should be contacted immediately.

3.5.2. DATA SURVEYING

The preferred tool supported by the university is Qualtrics. The tool can be used to complete surveys, grant user access, and complete high level data analysis. It is strongly recommended that if the survey is capturing Personally Identifiable Information (**PII**) data, health information, student records or other sensitive data that the researcher work with the institutional compliance areas (i.e., **IRB** or Institutional Animal Care and Use Committee (**IACUC**)) to ensure the survey is properly designed.

3.5.3. VERSIONING CONTROL SOFTWARE:

While research is occurring a backup procedure should be established. The following tools allow for versioning to occur.

- **Preferred tool: Microsoft Teams (Teams)** is a university supported and managed collaborative application that helps the team stay organized. Teams can develop channels for conversation and file sharing, allow project management functionality, and conduct meetings and calls. Teams is established for all faculty and staff and can be accessed from http://office.mu.edu. Teams can be used for internal university employees, staff, and faculty, as well as to invite external parties to collaborate and share files. This tool can be used for all data classification types, but access must be managed to ensure sensitive data is properly restricted.
    - Note that when file sharing occurs within Teams, a SharePoint site is developed for that Team environment, which has the full capabilities of SharePoint.
- **Preferred tool: Microsoft SharePoint** is a university supported and managed application that allows for sharing and managing of content, knowledge, and application to empower teamwork, quickly find information, and seamlessly collaborate across the organization. SharePoint is established for all faculty and staff and can be accessed from http://office.mu.edu. SharePoint can be used for internal university employees, staff, and faculty, as well as to invite external parties to collaborate and share files. This tool can be used for all data classification types, but access must be managed to ensure sensitive data is properly restricted.
- **Microsoft OneDrive** is personal file storage for individual workplace productivity and organizational file storage for management of departmental document libraries and files. This is the default storage area for work done on university devices when a user signs into their account. OneDrive allows for external and internal file sharing, co-authoring, and version control. This tool can be used for all data classification types, but if data collaboration is occurring it is recommended that Microsoft Teams or SharePoint be used, and access managed accordingly.
- **GitHub** is a web-based service for Git repositories (i.e., groups of tracked files). GitHub is commonly used for managing and sharing different versions of code for programming projects, but it can be used just as effectively for version control of other types of files, such as text documents. GitHub has a huge open-source community, but Marquette does not support this application and **only data classified as category 1 should be used** with this service.
- **Dropbox** is a file hosting service that offers cloud storage, file synchronization, personal cloud, and client software. Dropbox saves all the researcher's lost files and restores older versions of files. This environment is not managed by university resources and **this service should only be used for data classified as Category 1**. Services include external/internal file sharing, co-authoring, and version control.
- **Google Drive** is the same as Dropbox in terms of security, data storage capabilities, data management, and data version control. Google Drive is not supported by the university **and this service should only be used for data classified as Category 1**.

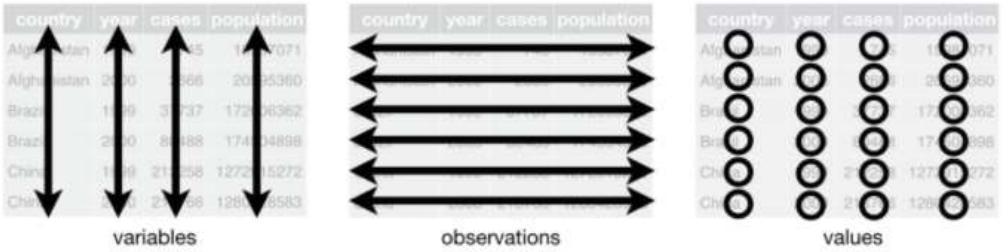
3.6. PHASE 3: ANALYZE AND COLLABORATE: CONCEPTS

Inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making are critical actions and outcomes of this phase. To effectively analyze data, not only does the data need to be organized, but it needs to be kept safe.

3.6.1. ANALYSIS READY DATASETS

Analysis-ready datasets, no matter the data classification, should be focused on the following key actions:

- Creating: Structuring data based on variables, columns, and rows using the following best practices:
  - All data should be labeled
  - Each experimental subject should have a unique study ID
  - Data should be rectangular format
    - Rows should represent the appropriate unit of analysis
    - Columns should represent the unique attributes of the row
  - Data files should contain the same number of columns in each row.
    - Note: Problems arise when data is missing, incomplete and in the middle of a row only
  - Data should be within each column without data combined into a single column (e.g., date and time)
- Formatting: Data should be tidy using the following concepts:
  - Each variable forms a column
  - Each observation forms a row
  - Each dataset contains information on only one observation's unit of analysis



- Validating: Review data and confirm that data is collected correctly. Best practices for data validation are:
  - Program valid ranges for inputting data into fields, when applicable
    - Example: If the data is only odd data, flag if data entered is even
  - Apply data formatting to fields in advance to prevent risk of inaccurate "automatic" formatting
    - Example: Format dates to always be MM/DD/YYYY instead of manually entering the date, which may result in data variations
  - Reduce/eliminate/plan for options where a data response would be "Other"
  - Reduce/eliminate/plan for "prefer to not answer" data
- Standardizing: Consistently record data that ensures the same kind and format for each data element being collected. Standardization helps to minimize data collection and analysis errors and prevents inconsistencies. Best practices consist of:
  - Coding data harmoniously and consistently

    - Standardizing free text responses into categorical data
    - Establishing a standard date and time to be used consistently (example ISO 8601: YYYY-MM-DD)

      Note: Additional guidance is provided in section 3.4.1 Naming conventions
- Cleaning: Before performing data analysis, review the data for completeness, inaccuracies, inconsistencies, or sensitive data. Cleaning allows for data errors to be rectified before analysis or unique identifiers to be used if data is sensitive. Best practices for data cleaning are:
    - Check for outliers: Ensure all data elements are in the correct formats and ranges.
    - Check for missing data: Ensure there are no data items that are missing, creating null elements. If data is missing, code data appropriately.
        - Example: If the response is no response, then record a zero or write "blank"; do not leave the cell empty.
    - Ensure that the data does not contain Protected Health Information (**PHI**) as defined under the Health Insurance Portability and Accountability Act (**HIPAA**), requires data to be protected and confidential for their patients. If HIPAA data is required for the research, work with IRB and IT to establish a secure platform/process for the data.
        - What is PHI data: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html
- Documenting: Ensure there are clear guidelines for understanding the data, the data analysis, and rules for sharing and reusing the data. Refer to section 3.4.4 ReadMe files for summary documents used for clarity. Best practices for data documentation and communication are to:
    - Store data with appropriate metadata
    - Create and use data dictionaries and ReadMe files
    - Save data as machine-readable ASCII or UNICODE files to reduce the need for specific applications to read files
    - Adopt appropriate file naming practices; section 3.4.1 Naming conventions

3.6.2. LAB NOTEBOOKS

According to the NIH, a lab notebook is a complete record of procedures, reagents, data, and thoughts to pass on to other researchers, as well as an explanation of why experiments were initiated, how they were performed, and the results. Additionally, a lab notebook is a legal document to prove patents and defend a researcher's data against accusations of fraud. Ultimately a lab notebook is a record of how a researcher is remembered during this time in their career. A misconception is that a lab notebook is a journal to write down ideas or theorize about how something works, which it is not.

Types of lab notebooks are:

- Bound/stitched: Traditional notebook with pages numbered and bound, reducing the likelihood of losing pages and making it a stronger legal document. However, it is difficult to digitize and may require references to data stored throughout the book as it may not be logically organized.
- Loose leaf/binder: Sheets can be easily reorganized to consolidate thoughts; however, pages can be lost more easily as pages are not numbered. Additionally, binders can be easily expanded to hold more volume of data than a bound/stitched binder. Due to the ability to open, add, and move pages, the authenticity of the binder can be questioned more easily than a stitched binder and this option is not usually recommended.

- Computer/Electronic: Notebooks are easy to search, read and expand data as the notebook capacity is only limited by the storage location capacity. However, data security of the notebook is critical to ensure data integrity, which means verifying user access and roles, as well as ensuring that files are not corrupted. Additionally, there could be software compatibility issues that may occur due to new releases of software. Contact IT Services to assess software options; industry standards are Labfolder, LabArchives, RSpace, Benchling, and SciNote

Data that should reside in a lab notebook includes:

- Notebook name
- Inside cover/cover page: Research name, year, general project name, lab address
- Table of contents: Must contain subject of the entry, date of the entry, where in the notebook the data resides
- Body of the notebook: Experimental entries (must contain: date, title, goal of experiment, background needed, how the experiment will occur, what observations occurred, what data analysis occurred, results / next steps for future experiments)
    - Note: When recording how the experiment will occur, ensure as much detail can be provided as possible. For example, if using a reagent, not only detail the chemical, but also the manufacturer, product name, lot number, expiration date, storage location, etc.

In addition to the NIH lab notebook guidance above, Rice University has developed a comprehensive guide for laboratory notebooks that should be referenced.

3.7. PHASE 3: ANALYZE AND COLLABORATE: TOOLS

A variety of tools are used for inspecting, cleaning, transforming, and modeling data when conducting and collaborating on data analysis. Below is guidance on a variety of standard tools that can be used no matter the data classification type.

3.7.1. PROJECT MANAGEMENT TOOLS

No formal guidelines are established; however, project management tools can be used to support researchers through the entire project lifecycle. Project management tools can be used for collaborating and allowing portions of data to be open to the public, as well as managing files, data, code, and protocols in one central location.

- Standard Marquette supported tools:
    - Microsoft Teams (Free): A collaboration tool that allows for data storage, data permissions, data communication, and integration of various tools, such as Planner or OneNote.
    - Microsoft Planner (Free): A traditional project management lite tool that allows for identification of key steps, file supporting in key steps, assignment of roles and responsibilities. Gantt charts are not available in this tool.
    - Microsoft Project (MU supported tool that is not free): A traditional project management tool that has full functionality for identification of key steps, linking steps, identifying project durations and delays, role assignments, and other topics.
- Open source tool (Not supported by MU):
    - Open Science framework: https://osf.io/

3.7.2. COLLABORATION TOOLS:

Used to manage files, data, code, and protocols in one central location, as well as collaborate with other project members.

- Standard tools supported by Marquette:
  - Microsoft Teams: Fully collaborative tool with data file storage, channel and chat collaboration, and interfacing to other applications
    - Supported features:
      - 25 TB of data max, with max file size of 100GB
      - Data permissions and user access restrictions
      - Access permissions allow both interactions within MU and outside of MU campus.
      - Enterprise wide security settings managed by IT Services
      - Remote access to the data
      - Backup plan maintained by IT Services as part of SharePoint,
      - Metadata criteria and sorting within SharePoint.
        - https://sharepointmaven.com/create-metadata-sharepoint/
  - SharePoint: Tool for data file storage and adding metadata fields to the file. This application can be integrated into a Teams site if necessary.
  - Physical research labs: Standard labs established for researchers that limit user access in a variety of physical security controls. Contact Office of Research Compliance (**ORC**) or IT Services for guidance on security protocols based on the sensitivity of the data.
- Tools used throughout Marquette campus (not supported by MU resources):
  - Note: While the tools in this category can be used, if data is in Categories 2, 3 or 4, data should not be stored in these tools as the security of the data environment cannot be verified by IT Services. However, it is understood that for a Principal Investigator (**PI**) from another university they may require the use of these tools and discussions should be conducted with the PI to ensure appropriate levels of risk management are taken.
  - RedCap: Tools used for data analysis usually in conjunction with the Medical College of Wisconsin that manages user access and the environment. Currently Marquette does not have a dedicated instance established.
    - Link: https://www.project-redcap.org/
  - Dropbox
  - Google Drive

3.7.3. COMMUNICATION TOOLS:

Used to interact with people either through video, audio, and/or text communications. All the following tools are used within the Marquette experience, but only one tool (Teams) IT Services supports with security settings for data storage and data encryption communications.

- Standard tools supported by Marquette:
  - Microsoft Teams (Free)
  - Microsoft Outlook (Free)
- Tools used throughout Marquette campus (not supported by MU resources):

- WebEx
- Zoom
- Google Meets

3.7.4. STATISTICAL ANALYSIS TOOLS:

Output of data analysis tools should be properly secured and stored no matter the data classification. Note: If category 1 data (e.g., public data) is used and analysis is completed, the output data and analysis used would become category 2 data and should be treated with more security than the category 1 data originally used.

- Standard tools supported by Marquette:
  - SPSS (Free): https://www.marquette.edu/its/help/downloads/
  - Computer clusters: Three main clusters used for specific research purposes
    - Computational cluster "Pangea" (Contact: Chair of Mechanical Engineering)
    - Computer science cluster (Contact: Director of Technology within Computer Science)
    - High performance computing cluster "Raj" (Contact: Director of Infrastructure within IT Services)
  - Microsoft Power Business Intelligence (BI), aka "DataMarq": Tool that allows for interacting with data through reports
- Tools used throughout Marquette campus (not supported by Marquette resources):
  - Stata: Paid subscription (multiple users in MU). Used for data science and graphical depiction of data.
    - Link: https://www.stata.com/
  - R: Free software for statistical computing and graphics.
    - Link: https://www.r-project.org/

3.7.5. TRANSCRIPTION TOOLS:

Software that allows for a conversation to be recorded by writing. No formal standard is set up for transcription software, but this type of software can make data entry of interviews or other research activities a less manual function. Note: That transcription data could become either category 2 or 3 data depending on the content of the transcription and data security should be handled accordingly.

- Standard tools supported by Marquette:
  - Microsoft Teams: Available only if the transcription functionality is turned on during a Teams meeting.
- Tools used throughout Marquette campus (not supported by Marquette resources):
  - Dragon Speak (Paid software)
    - https://www.nuance.com/dragon/industry/education-solutions.html
  - Temi (Paid service per minute) https://www.temi.com/
  - Otter subscription (Free basic platform): https://otter.ai/edu

3.8. PHASE 4: STORE AND MANAGE: CONCEPTS

All research data life cycle phases revolve around the management of data storage and it is imperative to ensure data remains secure and adheres to recommended safety protocols. Since data can be a variety of types, sensitivity designations, and transmission styles, it is important to ensure that data is stored effectively to enhance a culture of data security, thereby reducing the likelihood or impact of data being lost or corrupted.

- In Phase 1: plan and design a researcher should detail within the DMP Tool the plan for data storage throughout the project. Standard tools for storage are listed with the tools section of this phase.
- In Phase 2: collect and create datasets are being obtained from outside sources or generated from new data studies, storage protocols should be reviewed to ensure 3.8.1 Data Security is maintained.
- When the project is nearing completion, an additional review should be conducted as part of Phase 5: evaluate and archive to determine how data should be preserved long-term and to what degree data should be shared.

When focusing on storage the following items should be considered, which are detailed in section: 11 Data Security and Privacy Checklist

- During analysis:
    - What is the sensitivity of the data and does sensitive data need to be anonymized?
    - What is the data format and average size of a file?
    - What is the data retention need?
        - Note: If data is forever, what is the value of the data in 20 years?
    - What is the data destruction process, including naming convention, to easily identify when to purge files?
    - What are the known costs for current data analysis storage?
    - Who should have access to the data and how will this be managed / maintained?
    - What is the backup plan for data (i.e., what is deemed an acceptable loss of data that does not compromise the research)?
- End of project:
    - If data is sensitive, can it be anonymized to reduce the risk if data is lost?
    - Does the data need to be shared and if so, how will it be shared (MU repository [refer to section: 3.13 Phase 6: share and disseminate: tools], standard industry database, publication, etc.)?
    - How long does the data need to be stored?
    - What is the review process to remove/destroy data?
    - What format will the data be stored in?
        - Note: What is the process for verifying data compatibility as applications change over time?
    - What is the cost for project data storage?
    - Who should have access to the data and how will this be managed / maintained?
    - What is the backup plan for storage of data that supports the research findings?

3.8.1. DATA SECURITY

Data security is important to reduce the likelihood and impact of data being lost or corrupted. The university is committed to protecting the information that is critical to teaching, research, and the university's varied activities, business operations, and supported communities, including students, faculty, staff, and the public. These protections may be governed by legal, contractual, or university policy considerations.

As identified within section 2.1 What is data classification there are four main data classifications:

- **Category 1**: Publicly Available Information: This includes publications, unrestricted databases, and other materials available on-line that are not identified as confidential by the provider of the information.
- **Category 2**: Compiled or Generated Restricted: This data is generated during research or compiled as part of the research. This may include data sets created through measurements of materials, insects, or animals, compiled from publicly available information, lab notes, images, survey results.
- **Category 3**: Compiled or Generated Controlled: This is data that is not only confidential, but for which access is also controlled by federal or state law, such as personally identifiable information, personal health information, student records, or controlled unclassified information.
- **Category 4**: Licensed Data: This is data that has been licensed from a third party for use in conducting a research project.

When identifying data security, the Data Security and Privacy Checklist should be used to identify a holistic picture of data. This checklist should also be discussed with institutional compliance support staff (i.e., the IRB or IACUC) and or IT departments as appropriate.

For additional information on security details, visit the MU IT Services page.

A strong way to ensure data security is to reduce data access to the least amount of access needed for a person to perform their job. If a person leaves or changes roles, access should be removed or modified accordingly. Additionally, at least quarterly, access should be reviewed to ensure those with access are appropriate based on job role and job function. The quarterly access review should be performed for both physical locations, such as labs, and electronic access to applications, servers, or data storage locations.

Another way to reduce data risk, especially for restricted or confidential data, is for the researcher to encrypt the data at rest (i.e., data should be secured through encryption prior to it being stored). By encrypting data prior to storage, if the data is inappropriately obtained, the ability to access the data will be difficult and potentially impossible.

3.9. PHASE 4: STORE AND MANAGE: TOOLS

Data storage can take a variety of forms and should be used based on the data classification. Additionally, a researcher should select data storage according to their needs for backup capabilities, ownership of data, and data sharing. If questions arise while storing data, researchers should reach out to IT Services, departmental IT, or institutional compliance areas (i.e., the IRB or IACUC) and be aware of the guidance within the university policy, UPP 1-41: Cloud Computing Policy.

The following provides guidance and the preferred tools for data storage both during and upon completion of research:

- Standard Marquette supported tools:
  - During research:
    - Microsoft Teams (Free): Can be used for all data types, but for sensitive data, ensure data is stored in a private SharePoint site and user access is restricted. Microsoft Teams can be used for all categories of data, but user access should be managed to ensure data is properly secured. User access is managed at the Teams site level by the person who established the Teams site.
    - Departmental university network data storage:
      - User access: While departmental storage is stored in the MU network, permissions and user access may need to be reviewed to verify who has access based on where the data is stored. For data that is Category 2 or 3, this option is not recommended.
      - Cost: Storage may be free, but the department may be charged if storage capabilities need to be expanded.
  - Completion of research: Refer to section 3.13 Phase 6: share and disseminate: tools for details on the MU repository.
- Third party storage options:
  - During research:
    - Dropbox, Google Docs, etc.: Category 1 data only should be stored in these locations as the security parameters are unknown.
  - Completion of research: Industry/research specific repositories as needed for data sharing
    - Note: It is acknowledged that at times support will be stored in other third party locations, such as Dropbox, Git Hub, RedCap, etc., but a copy should be stored at least within the MU repository for final works as identified within section 3.13 Phase 6: share and disseminate: tools.

3.10. PHASE 5: EVALUATE AND ARCHIVE: CONCEPTS

When focusing on retention of data, a researcher must be aware of the data storage process and if the value of the data being stored outweighs the costs/efforts of the data storage process. The data storage process not only focuses on the data itself, but also the ongoing migration of data formats, storage costs, and who and how the data will be cared for, storage systems maintained, and user access services managed.

3.10.1. DATA RETENTION TYPES

No matter the data classification, essential research records should be maintained that substantiate:

- Grant applications or demonstrate compliance with contractual terms, if sponsored research
- Published research and patents
- Best practices within the field

From a data storage perspective there are two distinct storage retention disciplines:

- Permanent retention (i.e., archiving): Ongoing migration of electronic formats and storage costs, as well as care, maintenance, and user access services for records in perpetuity.
    - Note: Should be used for high value data to the researcher, MU, and society. This type of storage requires a considerable investment for MU to store.
- Long-term storage (i.e., preservation): Research data will be available to those who require access (e.g., sponsors, researchers, public, etc.) in a constant and user accessible format for a specific period as defined in the research data management plan (usually no more than 7 years).
    - Note: Assessments by the researcher should be conducted after the data retention period expires to determine if permanent retention is required.

When assessing what data should be retained and if the data should be permanently retained, the following questions should be considered:

- What are the essential records required to understand this research data and project?
- What was the impact of the research on its discipline?
- What impact has the researcher had on their field?
- Can the research data be replicated?
- Is the research data indexed, allowing future researchers to understand the data easily?
- Has the research been published? If so, where, and did the publication include applicable datasets?
- Has the data been kept in a research repository?
- Is the data sensitive and requires segregated storage or access requirements?
- Does someone else own the data?
- Has an institutional compliance area (i.e., IRB or IACUC) approved of the retention plan?
- If data was obtained from human subjects via consent, was appropriate consent obtained for the retention plan?

When identifying what type of data should be stored, the following research data and materials should be assessed for importance:

- Computer software required to read data (e.g., applications, programs, databases, etc.)
- Laboratory notes
- Organizational research documents (e.g., ReadMe texts)
- Materials submitted for approval to institutional compliance areas (e.g., IRB, IACUC, or other research oversight committees)
- Protocols, coding, algorithms, formulas, specimens, reagents, or other research specific critical data
- Datasets, graphs, charts, instrumental outputs

When in doubt, contact the Raynor Memorial Libraries Coordinator of Digital Scholarship and Programs) to help in assessing if the research data should be permanently stored in the MU repository (section: Phase 4: store and manage).

3.10.2. DATA COPYRIGHT AND INTELLECTUAL PROPERTY

If the researcher is an MU faculty, staff, or student, the researcher may create a patentable invention, copyrightable works, and other forms of intellectual property that has scholarly, scientific, and commercial value. Please note that raw data on its own is considered facts and thus cannot be copyrighted; however, data gathered in a unique and original way, such as databases, can be copyrighted or licensed. If a product or service is for public benefit, the product or service may benefit from intellectual property protections. If the researcher has any question on the copyright and intellectual property process, please work with the Technology Transfer department, specifically the executive director.

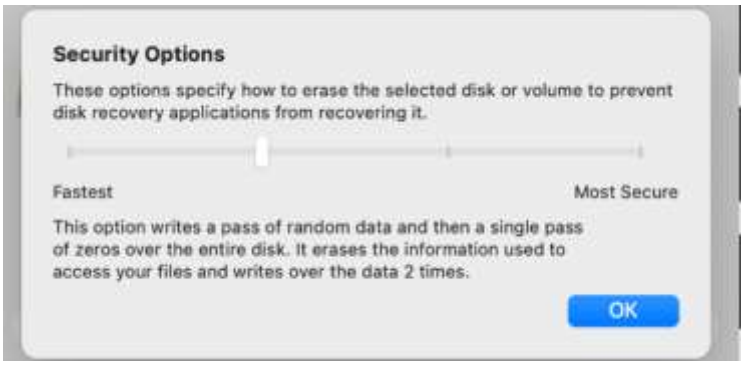
3.10.3. DATA DESTRUCTION

Destruction of data is critical as it reduces the risk to a researcher and the university in the case that data is lost or stolen and allows for storage space and storage costs to be reduced or eliminated. While the university has data retention guidelines, UPP 1-12: Records Management, the guidance may not address every data scenario. Therefore, a general best practice is that data should not be retained for more than 7 years, but the researcher should consult with the Office of General Counsel if clarity is needed.

How data is destroyed is dependent on the data classification and the mode by which data is being stored. The following are general guidelines for destruction of data; the IT Help Desk can be contacted for further assistance:

- *Paper:*
  - No matter the data classification, data should be placed in secure shred bins available across campus. Do not throw out research data in standard trash cans, especially if it contains Category 2, 3 or 4 data.
- *Electronic: SharePoint*
  - No matter the data classification, data stored on SharePoint can be removed in two ways, either as a full site or based on an individual file or folder. Guidance for remove of data based on both types is below:

        - Deleting an entire SharePoint site: Locate the applicable site, select the settings sprocket, select site information, and then select the “delete site” button.
        - Deleting a specific file/folder (not the entire site): Locate the SharePoint site, select the file or folder which should include a blue checkmark, select delete from the tool bar, and confirm file deleted.
            - Note: The file will still exist within the recycle bin. To permanently delete the file, empty the recycle bin by selecting the quick launch bar, selecting the recycle bin, and clearing the bin within the tool bar. Guidance can be found here: Emptying recycle bin.
- Electronic: Reformat hard drive or external storage device
    - Computer is part of CRP: The hard drive will be reformatted by IT Services or an authorized IT Service’s third party provided
    - Non-CRP computer: Contact IT services to verify next steps as the reformatting process varies based on if the hard drive is encrypted or not and what type of data is on the device.
        - *Note: There is a potential that the reformatting of the hardware will require only one attempt to sufficiently wipe the data if the non-CRP computer’s hard drive was encrypted or if there was no sensitive data on the computer.*
    - Non-CRP external storage device: No matter the operating system, it is recommended that the external storage device be reformatted by the device owner at least once prior to discontinuation of use, and if the external storage device had sensitive materials, the external storage device be reformatted three times to ensure data cannot be recovered using data recovery tools. *Note: If the external storage device was encrypted, then only one reformatting of the drive is required.*
    - Operating system types:
        - Windows: Go to File Explorer, right click on the drive, and select format drive. Ensure to uncheck the “quick format” box. The computer will go through the process of wiping the data and the drive can be reused.
        - Macintosh: Select the drive from the Disk Utility application and then select the GUID partition map or storage device to erase. Upon selecting the device select format and ensure that the security options for formatting is set to one step above fastest (see picture below). Then select erase to format the drive or device.



- *Electronic: Destruction of hardware*

    - Contact IT Services Help Desk and identify the hardware that requires destruction and if the hardware contains sensitive data. IT will work with the researcher and the third party vendor to identify the best process to dispose of the hardware, which may include triple degaussing or physical destruction of the hardware if the data is sensitive. Additionally, The IT Services Help Desk will work with the university preferred vendor to schedule pick-up of the hardware and obtain validation of data destruction from the vendor.
- *Electronic: Third party cloud service*
    - Refer to the applicable processes for each third party vendor on how to delete files properly. Please ensure that if recycle bins are part of the software, that recycle bins are also properly emptied.

3.11. PHASE 5: EVALUATE AND ARCHIVE: TOOLS

No formal tools are required as part of this process. Refer to the concepts within the Phase 5: evaluate and archive section for details on storing data upon completion of research.

Destruction of data is dependent on the data classification and the mode by which it is being stored. If the data destruction guidance within the concepts section above does not address a specific scenario, reach out to the IT Help Desk.

3.12. PHASE 6: SHARE AND DISSEMINATE: CONCEPTS

It has been common and, in some cases, essential for researchers to make their data available to others when a study is complete. Data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve society. Data sharing is not only growing due to the increased demand from researchers to study other researchers' work, but also due to compliance requirements from journals and funding organizations.

Benefits of data sharing include:

- Ability to find data years after finishing a project
- Enable others to replicate work
- Enable others to conduct new analysis using the research data
- Establish data citations when data is added to data repositories

The main data repository used by MU is https://epublications.marquette.edu/. In addition to the MU repository, the university is aware that the number of available resources for data sharing and data publication has increased substantially, making it difficult for individual researchers to evaluate the advantages and limitations of each resource. The global registry of academic data repositories, re3data.org, can be used to locate repositories for specific disciplines. When data is required to be stored in another location due to publication of funding source requirements or data lease agreements for category 4 data, MU asks that at least the journal and citation of where data is stored be located within the MU repository for reference purposes.

> Note: Data being published may sometimes be sensitive and unable to be published without being de-identifiable, such as category 3 data. Additionally, data sharing protocols and procedures should be reviewed and approved by institutional compliance areas such as the IRB or IACUC. Refer to section 3.8.1 Data Security for additional details on data security guidelines.

3.13. PHASE 6: SHARE AND DISSEMINATE: TOOLS

All final product deliverables no matter the data classification, should be stored within the MU repository. Category 3 or 4 data that is controlled can have access restricted accordingly. If data is required to be stored in another location, such as those found through the re3data.org global registry of academic data repositories, MU asks that at least the journal and citation of where data is stored be located within the MU repository for reference purposes.

Benefits of the MU repository are:

- Ability to cite work and reference in variety of ways
- Ability to house the journal publication, as well as critical support (e.g., data code, materials, formulas, etc.)
- Permissions can be granted to other researchers to allow data storage of critical research that is secured and non-viewable to the public.
- Allow easy knowledge sharing as data is categorized by discipline
- Google search results show individual research activity instead of requiring searching within the repository
- Reports can be generated identifying who is accessing research

- Data resides within MU environment, but the researcher has access to the data even if they leave MU

PHASE 7: ACCESS AND REUSE

3.14. PHASE 7: ACCESS AND REUSE: CONCEPTS

Access to data and how a person is permitted to use the data should be addressed as part of this phase, especially for data that is category 2 or 3 data. When publishing or sharing data, determine how and to whom access will be provided. Additionally, determine what are deemed acceptable uses of the data and if provisions need to be established for reuse, redistribution, and creation of derivatives, especially with category 2, 3, or 4 data.

It is important that the data produced should have sustainability both in a researcher's active project work and as a researcher progresses in their career. Critical steps to ensure clarity of research data are:

1. Verify that the DMP is clear and representative of research during and upon completion of research. Please refer to Phase 1: plan and design for DMP guidance. A DMP should be clear enough that another researcher can reproduce findings using the corresponding research data, including clear steps on actions taken or assumptions used when completing the analysis. If the DMP is not clear, additional researchers may not be able to reproduce or validate the research.
2. Publish the data within a data repository upon completion as identified within Phase 6: share and disseminate.
3. Offboarding from the university should result in the completion of a checklist that clearly identifies the documentation being stored at the university, why it is being stored, and how the data can be interpreted as detailed in section 12 Research Offboarding Checklist.

3.14.1. OFFBOARDING CHECKLIST

Research data management provides an opportunity for researchers to create a plan ensuring data will be organized and shared with others. When students, faculty or staff leave a research project, they take their skills and institutional knowledge with them. It is important to record essential information related to projects and datasets to ensure the success of future researchers and users.

Section12 Research Offboarding Checklist outlines important steps that should be followed when an employee leaves.

3.14.2. KNOWLEDGE TRANSFER FILE

Similar to the offboarding checklist, the Knowledge Transfer file should be completed by a researcher that will be leaving the university. The knowledge transfer file allows the university to gain a holistic picture of research/publications completed, and research/publications in progress. While the knowledge transfer is used as part of offboarding, the file can be completed throughout the tenure of a researcher and can be introduced at the start of a project or hire date for tracking research activities. Refer to section 13.2.8 Knowledge transfer form for the template.

3.15. PHASE 7: ACCESS AND REUSE: TOOLS

When researchers start a new project or join a new project/lab the researcher should review section: 7 Research Data Management Onboarding Checklist to understand the standards of the project/lab.

When researchers leave the university, it is critical to ensure the research and knowledge is not lost. There are no formal tools that are used as part of the transition, but as described in Phase 7: access and reuse there are two critical documents that should be completed, ensuring projects can continue following a researcher's departure and enhance research consistency and efficiency. The two forms are:

- Research offboarding checklist: Provides clarity on specific project status/data
- Knowledge transfer form: Provides a holistic picture of research/publications completed, and research/publications in progress

If electronic files are stored and shared with users, as identified in 3.8.1 Data Security, an access review should be done at least quarterly to ensure users have appropriate access to data based on their job role and responsibility.

4. DEFINITIONS

The definitions apply to any specialized terms used in this document.

Term	Definition
API	Application programming interface
Confidentiality breach	Use or disclosure of a patient's private information without their consent
DMP	Data Management Plan
HIPAA	Health Insurance Portability and Accountability Act of 1996
IACUC	Institutional Animal Care and Use Committee
IRB	Institutional Review Board
MU	Marquette University
ORC	Office of Research Compliance
PHI	Protected Health Information
PI	Principal Investigator
PII	Personally Identifiable Information
Teams	Microsoft Teams
UPP	University Policy and Procedure

5. OWNERSHIP, REVISION HISTORY, & EFFECTIVE DATE

Document Owner

- IT Compliance Manager

Authorization

- Reviewer 1:
- Reviewer 2:

Revision History

A record of changes made to this document.

Effective Date	Reason for Revision	Author
TBD	New Procedure	IT Compliance Manager

6. FUTURE TOPICS

These are topics that will be discussed and potentially added to a future draft of this document. If a student, faculty, or staff has any context or best practices related to these topics, please email the IT Compliance Manager within IT Services.

- Export controls
- Formal guidance on lab notebooks and preferred electronic lab notebook tools
- More comprehensive guidance about data encryption and standard tools to be used
- Additional tools and tool categories that should be added to the applicable research data life cycle, such as NVivo and GIS
- Establish tools specific to the level of risk that accompanies the data availability and data classification for research data

7. RESEARCH DATA MANAGEMENT ONBOARDING CHECKLIST

Purpose: To provide a general, research data management-focused guide for researcher onboarding as they join a new lab or begin a new project. This section provides two checklists that can be used to assist in navigating critical steps for the research:

- Join a new project
- Join a new lab

Scope: All student, faculty, and staff that support or directly work on a new research project or lab

Join a New Project

Stage: Planning

Focus areas	Key steps	Reference points
Transfer prior data and related records to the university (if relevant)	☐ Contact the Office of Research and Sponsored Programs	Office of Research and Sponsored Programs
Review project and granting institution requirements	☐ Most government funding agencies have published guidelines. If the researcher has non-federal funding, check with the granting agency/award agreement.	NSF Data sharing Policy NIH Data Sharing Policy NIH Public Access Policy
	☐ For projects involving human subjects research, review any project-specific requirements stipulated by Marquette's Institutional Review Board (IRB).	Marquette IRB
	☐ For projects involving animal research, review any project-specific requirements stipulated by Marquette's Institutional Animal Care and Use Committee (IACUC).	Marquette IACUC
Write a DMP or review existing DMP	☐ Construct a data management Plan (DMP)	Raynor Memorial Libraries DMP Tool
Create a data workflow or review existing data workflow	☐ If applicable, review existing project workflows and directory structures	
	☐ Develop project specific workflows and directory structures	Phase 2: collect and create
	☐ Assess using an electronic lab notebook	3.6.2 Lab Notebooks
	☐ Establish controlled vocabulary / metadata or adapt existing controlled vocabulary	
Create preliminary ReadMe file(s) for each project dataset	☐ Document the data workflow(s) in a ReadMe file. The metadata will help ensure the data is understandable, usable, discoverable, and reproducible.	3.4.4 ReadMe files

Stage: Storage

Focus areas	Key steps	Reference points
Review storage options	☐ Review storage resources in place for the existing project or choose relevant storage options for a new project.	Phase 4: store and manage

Stage: Sharing

Focus areas	Key steps	Reference points
Review intellectual property policy	☐ Review Marquette intellectual property policy	Intellectual Property Policy
Review publisher and funder requirements	☐ Understand funder and/or publisher data sharing requirements, if applicable	NIH Public Access Policy NSF Public Access Policy
Review available collaborative tools	☐ Review collaborative tools already in use for an existing project or choose relevant tools for a new project	Phase 3: analyze and collaborate
Review potential data repositories	☐ Review public data repositories already established for an existing project or choose a relevant data repository for a new project	
	☐ Develop a new data repository for the project, if applicable	Phase 6: share and disseminate

Join a New Lab

Stage: Planning

Focus areas	Key steps	Reference points
Review data management plan (DMP) policies	☐ Review DMP guidance	Raynor Library DMP Tool
	☐ Understand data retention and destruction of records guidance	3.10.1 Data retention types
	☐ Review information security guidelines	MU IT Services page
Create a preliminary data workflow	☐ Review existing lab workflows and directory structures	
	☐ Develop a preliminary organizational workflow for the research, including a file (or directory) structure	Phase 2: collect and create
Create preliminary ReadMe file(s) for each dataset	☐ Document the data workflow(s) in a ReadMe file. The metadata will help ensure the data are understandable, usable, discoverable, and reproducible.	3.4.4 ReadMe files

Stage: Storage

Focus areas	Key steps	Reference points
Review storage options	☐ Review storage resources in place for the existing project or choose relevant storage options for a new project	Phase 4: store and manage

Stage: Sharing

Focus areas	Key steps	Reference points
Review intellectual property policy	☐ Review Marquette intellectual property policy	Intellectual Property Policy
Review available collaborative tools	☐ Review collaborative tools already in use for an existing project or choose relevant tools for a new project	Phase 3: analyze and collaborate
Review potential data repositories	☐ Review public data repositories already established for an existing project or choose a relevant data repository for a new project.	Phase 6: share and disseminate

8. RESEARCH DATA MANAGEMENT CHECKLIST

Purpose: To provide a high-level checklist to ensure completion of critical components of the research data life cycle.

Scope: All students, faculty, or staff that conduct a research project

Phase 1: plan and design	
☐	Write a data management plan (DMP)
☐	Define roles & responsibilities
☐	Onboard project members
☐	Choose documentation standards

Phase 2: collect and create	
☐	Choose file naming convention
☐	Establish file directory structures
☐	Establish version control
☐	Develop ReadMe files

Phase 3: analyze and collaborate	
☐	Establish digital notebook
☐	Use collaborative file sharing
☐	Establish backup process for data
☐	Clean data (validate and tidy)

Phase 4: store and manage	
☐	Determine data type and size
☐	Assess data storage retention
☐	Manage paper and physical samples
☐	Evaluate data security level

Phase 5: evaluate and archive	
☐	Establish data ownership
☐	Evaluate data retention
☐	Anonymize sensitive data
☐	Curate and archive data

Phase 6: share and disseminate	
☐	Obtain compliance approval for sensitive data
☐	Publish data in MU repository
☐	Publish data in other repository, if applicable
☐	Share with collaborators

Phase 7: access and reuse	
☐	Transfer access & knowledge
☐	Evaluate data security level

9. FILE NAMING CHECKLIST

Purpose: To provide general guidance for naming files and support for research purposes.

Scope: All students, faculty, or staff that conduct a research project

FILE NAMES SHOULD BE MACHINE READABLE, HUMAN READABLE, AND EASILY SORTABLE (i.e., DEFAULT ORDERING)	
☐	DO NOT put patient information or names within naming conventions
☐	Check for established file naming conventions in the researcher's discipline or department. Naming conventions should be documented so that project team members or department members can follow this standard.
☐	File names should be descriptive and provide just enough contextual information
☐	Try not to make file names too long. Operating systems have different limits to the number of characters. Generally, try to aim for a 40-50 character limit.
☐	Put the most important information first. The computer arranges files by name, character by character. *Note: If a researcher anticipates wanting to find a file by date, then put the date first*.
☐	Use ISO 8601 standard for: - Date (YYYYMMDD) - Date and timestamp (YYYYMMDDThhmm). Use 24-hour military time to avoid any confusion over a.m./p.m. o Example: 202111240834 is November 24, 2021 at 8:34 a.m.
☐	When using a sequential numbering system, use leading zeros to make sure files sort in sequential order. - Example: 001, 002, …., 010, 011 …100, 101
☐	Use versioning to indicate the most current version of a file - Example: filename_v2.xxx
☐	Avoid special characters, such as ~! @ # $ % ^ & * ( ) ` ; : < > ? . , [ ] { } ' " |
☐	Do not use spaces as some software will not recognize file names with spaces. Other options include underscores, dashes no separation, or camel case (first letter of each section of text is capitalized)

File Naming Convention Examples
[investigator]_[method]_[subject]_[YYYYMMDD]_[version].[ext]
[project #]_[method]_[version]_[YYYYMMDD].[ext]
[YYYYMMDD]_[version]_[subject]_[datacollector].[ext]
[type of file]_[author]_[date].[ext]

File Naming Examples		
Examples of good naming convention		***Examples of poor naming convention***
20190102_AC_smithlab_utra_exp01_gel_003.tiff		Test data 2016.xlsx
20190501_exp123_analysis_v01.pdf		Meeting notes Jan 17.doc
20190811_bioassay_toxicity_v1.sps		Notes Eric.txt
2020-plos-manuscript-v26.pdf		Final last version.docx

10. README FILE CHECKLIST

Purpose: To provide a high-level checklist to ensure key questions are clarified to allow others to better understand the project research and research evidence within a ReadMe file.

Scope: All students, faculty, or staff that conduct a research project

Who will write the ReadMe files?	
Need a clear task assignment	
☐	Researcher
☐	Fellow researcher
☐	Data manager
☐	Other:

What information will be required?	
Know what should be included	
☐	Specific format to use
☐	All categories provided
☐	Data type requirements
☐	Other:

When will the researcher update and review?	
Make a realistic plan	
☐	Write it once and leave
☐	Update monthly
☐	Review by the researcher and others
☐	Other:

Where will ReadMe files be placed?	
Keep track of their locations	
☐	Each file
☐	Each folder
☐	Group folder
☐	Other:

Why does the target audience need the ReadMe file?	
Define the target audience	
☐	Required by funder
☐	Future review by researcher (i.e., self)
☐	Explain the study again
☐	Other:

BONUS: How will the researcher make ReadMe files?	
Try a few approaches	
☐	Use a template
☐	Write individually
☐	Automated program
☐	Other:

11. DATA SECURITY AND PRIVACY CHECKLIST

Purpose: To provide a planning tool primarily for use by researchers to think through the research plan and prepare for submitting an application or grant proposal. The checklist is intended to strengthen project plans, alerting researchers to potential vulnerabilities, and to prompt additional planning to reduce information risks, to the extent necessary and feasible.

Research involves increasingly complex arrangements for the storage and transmission of research data. Robust data privacy and security planning is necessary to protect the privacy of research subjects and to secure sensitive, personally identifiable information (PII).

After completing the checklist, researchers are encouraged to contact their institutional compliance areas such as the IRB or IACUC, and/or IT departments as appropriate. The checklist is not intended as an audit tool; it does not certify compliance and expresses no opinion as to the adequacy of any given plan.

Scope: All students, faculty, or staff that conduct a research project

Project Information	
1.1 Project Details	Project Name: Single Site Study: ☐ Yes ☐ No Will there be a coordinating center? ☐ Yes ☐ No Will data be shared between centers? ☐ Yes ☐ No
1.2 Principal Investigator (PI) Information	PI Name: PI Institutional Affiliation:
1.3 Data Manager/Data Custodian *(Individual responsible for data, other than PI)*	
1.4 Study Coordinator Name:	
1.5 Other Persons at the Institution with Access to the Data (indicate role/title)	

Receiving and Collecting Data	
2.1 Will data be obtained from a source outside the study? *(e.g., a vendor, a company, a collaborator from a different institution or department, a government agency)*	☐ Yes ☐ No If yes, please specify:
2.2 Will data be produced by the study?	☐ Yes ☐ No If yes, please describe the datasets:
2.3 Where, and in what format, will data be stored?	Data will be stored: ☐ Yes ☐ No Format of data:

2.4 Will this project involve secondary use of data? *(i.e., reuse of data from another project)*	☐ Yes ☐ No If yes, list the project name and investigator who originally obtained the data:
2.5 Is there an approval letter from the original data owner for this reuse?	☐ Yes ☐ No
2.6 Is this research funded by an outside sponsor?	☐ Yes ☐ No If yes, please specify:
2.7 Do the terms of the award or the research agreement limit how the data may be used, maintained, or shared?	☐ Yes ☐ No If yes, please specify:

Data Information	
3.1 Data type	☐ Lab data ☐ Survey data ☐ Imaging data ☐ Claims and enrollment ☐ Service ☐ Clinical data ☐ Genetic information ☐ Media (Video ☐, photo ☐, audio ☐) ☐ Other, please specify:
3.2 Will the data contain any HIPAA identifiers? *(Refer to the Additional Checklist Guidance section for examples of HIPAA identifiers)*	☐ Names ☐ Geographic subdivisions smaller than a state (except the first three digits of a zip code) ☐ Elements of dates (except year) directly related to an individual, including birth date, admission date, discharge date, date of death ☐ Ages over 89 ☐ Telephone number ☐ Fax numbers ☐ Email addresses ☐ Medical record numbers ☐ Health plan beneficiary numbers ☐ Account numbers

	☐ Certificate/license numbers ☐ Vehicle identifiers and serial numbers ☐ Device identifiers and serial numbers ☐ Web Universal Resource Locators (URLs) ☐ Internet Protocol (IP) address numbers ☐ Biometric identifiers, including finger and voice prints ☐ Full face photographic images or any comparable images ☐ Any other unique identifying number
3.3 Does this research involve identifiable human subject data?	☐ Yes ☐ No If yes, has an IRB reviewed this study? ☐ Yes ☐ No ☐ Pending IRB Name: IRB Approval date (if any): IRB approval number (if any):
3.4 Does this research involve animal research?	☐ Yes ☐ No If yes, has an IACUC reviewed this study? ☐ Yes ☐ No ☐ Pending IACUC Name: IACUC Approval date (if any): IACUC approval number (if any):
3.5 Will datasets received or created be "limited datasets"? *(Refer to the Additional Checklist Guidance section for the definition of limited datasets.)*	☐ Yes ☐ No Coded? ☐ Yes ☐ No If yes, who has the link?
3.6 Will data be de-identified …	De-identified? ☐ Yes ☐ No If yes, will a third-party de-identification service be used?
3.7 For what purpose will data be used?	For student research ☐ Yes ☐ No For post-doctoral research ☐ Yes ☐ No For publication? ☐ Yes ☐ No For external collaboration? ☐ Yes ☐ No Other, please describe:

Data Storage, Access, Collection, and Security	
4.1 Does this study have a Data Management Plan (DMP) or Data Security Plan?	☐ Yes ☐ No If yes, who approved the plan? (i.e., IRB, IACUC, IT Services)
4.2 Describe how data will be stored while the study is active:	Data storage: If data will be collected, transmitted, and/or analyzed via an internet application or cloud service, include the security plan for this data, if any:
4.3 Where will the data be accessed from?	Data will be accessed from: ☐ Marquette network (hard drive, local server, SharePoint) ☐ Internet/web application ☐ Cloud service ☐ Other If accessed from a source not within the Marquette Network, please specify:
4.4 Will the data be accessed from a remote device *(e.g., e-tablet, smart-phone, non-Marquette managed personal/home computer)*?	☐ Yes ☐ No If yes, please specify:
4.5 Will data from this study be stored electronically?	☐ Yes ☐ No If yes, please specify: ☐ Local server (SharePoint or Teams) ☐ Third party servers ☐ Hard drives ☐ Portable devices ☐ Other, please describe:
4.6 Does the researcher have a reporting plan in the event of intentional or unintentional loss, alteration, or destruction of data?	☐ Yes ☐ No If yes, please specify:
4.7 Will the researcher keep paper-based records?	☐ Yes ☐ No If yes, please specify:

4.8 Does the researcher have a plan for maintaining backup copies of the data?	☐ Yes ☐ No If yes, please specify:
4.9 Does the researcher have means to notify institutional departments or data vendors about material changes to the data plan?	☐ Yes ☐ No If yes, please specify:
Data Sharing and Data Transport	
5.1 Will this data be shared with individuals outside of MU staff, faculty, or students within the research group (e.g., external collaborators)?	☐ Yes ☐ No If yes, please specify:
5.2 Will data be submitted to publicly accessible repositories during the research?	☐ Yes ☐ No If yes, please specify:
5.3 If the project involves Protected Health Information, are appropriate agreements in place?	☐ Yes ☐ No If yes, please specify: Has the IRB approved the plan to share the data? ☐ Yes ☐ No
5.4 Is there a plan for encryption of data when transferred electronically from site to site or safeguarding of data if physically transported?	☐ Yes ☐ No If yes, please specify: Has IT Services approved the transfer plan: ☐ Yes ☐ No
5.5 Will data be collected, analyzed, stored on an internet application or remote third party service?	☐ Yes ☐ No If yes, please describe the security protocols for the application:
5.6 For data that is being shared, in what format is the data being sent?	☐ Identifiable ☐ Coded ☐ De-identifiable ☐ Other, please describe:

Data Retention and Destruction	
6.1 How long will the data be stored? Additionally, if data is stored beyond 7 years, what value will the data have after 7 years?	Storage duration: Value of data stored beyond 7 years, if applicable:
6.2 Is there a plan for post-study disposal / destruction of data?	☐ Yes ☐ No If yes, please specify:

6.3 How will data be returned to the original owner (either the funding institution, principal investigator (PI), or the research participant), if applicable?	Describe process:
6.4 Will the data be populated within the MU data repository upon research being completed?	☐ Yes ☐ No If no, please specify where the data is being stored:

Additional Checklist Guidance

Definition of PHI
Any individually identifiable health information, whether oral or recorded in any form or medium that

- Is created or received by a health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse, and
- Relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual

HIPAA Identifiers

- Names
- Geographic subdivisions smaller than a state (except the first three digits of a zip code if the geographic unit formed by combining all zip codes with the same three digits contains no more than 20,000 people and the three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000).
- Elements of dates (except year) directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89
- Telephone numbers
- Fax numbers
- Email addresses
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers
- Device identifiers and serial numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images or any comparable images
- Any other unique identifying number, characteristic, or code

Limited Datasets
DO NOT include direct identifiers (see above), but datasets may include the following indirect identifiers:

- Town or city, state, zip code
- Ages in years up to 90 years (must aggregate all ages 90 or older)
- Dates directly related to an individual—such as birth date, date of death, admission date, discharge date, visit date, diagnosis date, etc. (Month/Year is preferred [no exact day]).

Sometimes vendors or agencies provide a study number with the data. To be labeled as a limited dataset these study numbers CANNOT be an encoded identifier such as a scrambled birth date, patient initials, last four digits of the social security number, etc.

12. RESEARCH OFFBOARDING CHECKLIST

Purpose: To provide a general guide for employee/trainee offboarding and should be reviewed as an employee or trainee leaves a research group.

Scope: All students, faculty, or staff that conduct a research project

Stage: Planning

Focus areas	Key steps	Reference points
Create, refer to, or update a knowledge transfer file	☐ Create a descriptive *Knowledge Transfer File* with relevant metadata. - Include in the *Knowledge Transfer File* the entity responsible for future maintenance of the data	○ 13.2.8 Knowledge transfer form ○ 10 ReadMe File Checklist ○ 9 File Naming Checklist ○ 13.2.5 Data management plan tool (DPM tool)
Comply with institutional, departmental, and lab policies and procedures related to data retention	☐ Determine the length of time the data produced must be retained	○ 3.10.1 Data retention types

Stage: Storage

Focus areas	Key steps	Reference points
Review and organize the data	☐ Review existing lab and departmental data storage proposals.	○ Phase 4: store and manage
	☐ Review and organize data in collaborative folders so they are easily accessible to colleagues.	
	☐ Create and/or update the ReadMe documents for each distinct dataset.	○ 10 ReadMe File Checklist
	☐ Document in the *Knowledge Transfer File* the location of data related to the research.	○ 3.10.1 Data retention types
Transfer access permissions for collaborative data to appropriate lab members or colleagues	☐ Transfer file folder and webpage/website ownership, as appropriate.	○ Phase 7: access and reuse
Identify data for migration to long term storage	☐ Review data and data storage guidelines to determine if data needs to be stored in long term storage and store applicable support accordingly.	○ Phase 4: store and manage ○ Phase 5: evaluate and archive
	☐ Ensure data not moved to long term storage is properly backed-up to prevent data loss.	○ Phase 5: evaluate and archive
	☐ Document in the *Knowledge Transfer File* the location of data related to the research.	○ 13.2.8 Knowledge transfer form
Identify data for potential deletion	☐ Determine and discuss with the PI the deletion of duplicate or dispensable data to help reduce lab or departmental lob storage.	○ 3.10.1 Data retention types ○ Phase 5: evaluate and archive

Ensure proper storage of and access to lab notebooks (physical or electronic)	☐ Store the researcher's lab notebook and other lab records according to lab protocol.	○ 3.6.2 Lab Notebooks
	☐ Confirm the lab notebooks are accessible to appropriate team members and collaborators.	

Stage: Sharing

Focus areas	Key steps	Reference points
Review security and confidential guidelines	☐ Review storage and dissemination guidance to ensure compliance with university guidelines.	○ 13.2.5 Data management plan tool (DPM tool) ○ Phase 4: store and manage ○ Phase 6: share and disseminate
Identify publisher, funder, or institutional requirements for data sharing	☐ Identify publisher, funder, and/or institutional requirements for data sharing and long-term maintenance.	○ Phase 4: store and manage ○ Phase 6: share and disseminate
Identify which of the datasets should be deposited and shared in repositories	☐ Identify datasets that should be deposited and shared in public or non-public repositories.	○ Phase 6: share and disseminate
	☐ Confirm that data in proprietary repositories is accessible to other team members.	
	☐ Document shared datasets in the *Knowledge Transfer File*.	○ 13.2.8 Knowledge transfer form
Transferring data to other institutions	☐ Consult with lead researcher/ principal investigator (PI) or department head prior to data transfer.	
	☐ If transferring data to another institution prior to the researcher's departure, ensure that sensitive data is securely stored, and appropriate approvals are sought/obtained from the institution the data is being transferred to.	

13. REFERENCES

13.1. INTERNAL REFERENCES

- UPP 1-05: Acceptable use of Electronic Resources
- UPP 1-12: Records Management
- UPP 1-28: Information Sensitivity
- UPP 1-40: Account Creation, Deletion and Retention
- UPP 1-41: Cloud Computing Policy
- UPP 2-04: University Intellectual Property policy

13.2. FORMS/TEMPLATES/KEY DOCUMENTS

13.2.1. RESEARCH DATA MANAGEMENT ONBOARDING CHECKLIST

- Refer to section: 7 Research Data Management Onboarding Checklist

13.2.2. HIGH LEVEL CHECKLIST

- Refer to section: 8 Research Data Management Checklist

13.2.3. FILE NAMING CHECKLIST

- Refer to section: 9 File Naming Checklist

13.2.4. README FILE CHECKLIST

- Refer to section: 10 ReadMe File Checklist

13.2.5. DATA MANAGEMENT PLAN TOOL (DPM TOOL)

- https://www.marquette.edu/research-sponsored-programs/data-management-planning-tool.php

13.2.6. DATA SECURITY AND PRIVACY CHECKLIST

- Refer to section: 11 Data Security and Privacy Checklist

13.2.7. RESEARCH OFFBOARDING CHECKLIST

- Refer to section: 12 Research Offboarding Checklist

13.2.8. KNOWLEDGE TRANSFER FORM

Knowledge%20Transfer%20Form_Final.x