Research Data Management Planning Checklist

This document serves as a reference checklist for keeping track of the various elements related to managing data and any research outputs during the course of a research project and once the project ends. The checklist follows the sequence of the Research Data Management Life Cycle, but research projects are often not linear: you may address certain stages before or after others or enter the lifecycle at different stages, depending on the nature of your project. The elements on this checklist are often included in a README for a dataset or for a research project, as a way to describe the project concisely and comprehensively.

Resarch Data Management Life Cycle

☐	**ID**	
		Determined by the funder and/or institution
☐	**Funders**	
		Read and comply with your funder's:
		• Data retention policy
		• Data sharing policy
		• Data retention policy
☐	**Grant Number**	
		Post-award DMPs only
☐	**Project Name**	
		As it appears exactly on the grant. Append to grant proposal
☐	**Project description (background and rationale)**	
		What research questions are you addressing?
		Summarize the study methods and design including data collection methods and purpose of collection
		If creating or collecting data in the field, how will you ensure its safe transfer into your main secured systems?
	Data description	
☐	*Describe the content (brief):*	
		• Define data values
		• Describe questionnaires
		• List instruments
		• Include analysis procedures
☐	*Describe the data type:*	
		• E.g., genomic, spatial, observational, experimental, simulations, models, etc.

☐	*List the data formats:*	
		Tabular data with extensive metadata:
		SPSS portable format (.por); Delimited text and command file (SPSS, Stata, SAS, etc.);Structured text or mark-up file of metadata information (.xml)
		Tabular data with minimal metadata:
		Comma-separated values (.csv); Tab-delimited file (.tab); delimited text with SQL data definition statements
		Geospatial data:
		ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn); Geo-referenced TIFF (.tif, .tfw); CAD data (.dwg); Tabular GIS attribute data; Geography Markup Language (.gml)
		Textual data:
		Rich Text Format (.rtf); Plain Text, ASCII (.txt); eXtensible Markup Language (.xml)
		Image data:
		TIFF 6.0 uncompressed (.tif)
		Audio data:
		Free Lossless Audio Codec (FLAC) (.flac)
		Video data:
		MPEG-4 (.mp4); OGG video (.ogv, .ogg); Motion JPEG 2000 (.mj2)
		Documentation and scripts:
		Rich Text Format (.rtf); PDF/UA, PDF/A, PDF (.pdf); XHTML or HTML (.xhtml, .htm); OpenDocument Text (.odt)
	If you need to convert or migrate your data files from one format to another, be aware of the potential risk of the loss or corruption of your data and take appropriate steps to avoid/minimize.	
		Briefly justify the use of format–is your chosen format open, non-proprietary and in widespread use?
		Estimated volume of data?
		Describe any existing data being used (citations, link, and DOI)
☐	**PI**	
		Name of the Principle Investigator(s) or main researchers on the project
☐	**PI ORCID ID**	
		ORCID (http://orcid.org)
☐	**Administrative data**	
		Contacts/addresses/email details
		Date of first DMP
		Date and details for subsequent revisions of DMP
☐	**Additional Institutions**	
		Read and comply with your funder's:
		• Data retention policy

		• Data sharing policy
		• Data retention policy
☐	**Storage and Backup**	
		For information on storage and backup options, visit researchdata.wisc.edu/data-storage-backup/

Data Storage for Active Data

Where will each of your datasets be stored? Where will any subsets of those data be stored?

Storage is a consideration at every stage of the lifecycle

☐	**Organization**	
		Will someone new to the project be able to follow the workflow easily? Is the process and organization consistent throughout?
		Controlled vocabularies used as appropriate (e.g., MeSH, Getty Thesaurus of Geographic Names, ICPSR Thesaurus, etc.)
		Describe your **file-naming/folder structure**. Research data files and folders need to be labeled and organized in a systematic way agreed upon by the entire research team, so they're both identifiable and accessible for current and future users
		Version control: manually or with a system (GitHub or GitLab)
		Resource: researchdata.wisc.edu/file-naming-and-versioning
		Do you have a master version of your raw data?
		Are the **raw data** stored in a location where they will not be modified or deleted? Raw data should have a master version where no changes are made. Any changes to the raw data in subsequent versions should be well documented.
		Quality assurance processes (calibration, repeat samples or measurements, standardized data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies)
		Team concensus/agreement to use standard file-naming conventions and versioning plans
		Record scripts for every stage of data processing and/or have a plan to document every manual action/change
☐	**Active data**	
		Where is the data stored? Electronic computing systems maintained by the University?
		Resource: researchdata.wisc.edu/data-storage-and-backup
☐	**Long-term (retention)**	
		How long will the data need to be retained and preserved according to the relevant policies?
☐	**Metadata**	
		What information is needed for the data to be read and interpreted in the future?
		Metadata standards: Follow standards for documenting your data based on what is used in your discipline, or within your research group. At minimum, provide the following information.
		• Who created or contributed to the data
		• Title of dataset and research project
		• Date of creation
		• Access location and restrictions
		• Methodology
		• Analytical information and tools

		• Variable definitions (codebooks, data dictionaries)
		• Standard vocabularies/units of measurement
		• Data format
		• Data file type
		• Data file size
		Resource: researchdata.wisc.edu/metadata
☐	**Cost**	
		Do you have sufficient storage or will you need to include charges for additional services?
		Depending on the organization funding your research project, the costs for storing and backing up your data can be included in the budget for your research proposal

Storage and Maintenance: Security

How secure is your data and how can it be assessed?

Storage is a consideration at each stage of the research data life cycle.

University IT teams provide robust managed storage and backup solutions for campus members. Consult UW DoIT or your department's IT staff to determine the best storage and backup options for your research project. http://researchdata.wisc.edu/data-storage-backup/

☐	**General Security**	
		What are the risks to data security and how will these be managed?
		• How will you control accesss to keep the data secure?
		• How will you ensure that collaborators can access your data securely?
		• Where will you store your data?
		• Will external media related to your research, such as paper lab notebooks, be kept secure in locked cabinets with access logs and a list of authorized users?
		• How will you protect the integrity of your data? *(i.e., data transferred over the network will be encrypted, access to data related to my research is accessible only by those who are authorized to access it, a plan for validating the integrity of the data)*
		• How will you protect the identity of participants (i.e., honest broker, anonymized data) according to the Common Rule, FERPA, and HIPAA
☐	**Software**	
		How will you protect your hardware and software systems? (e.g., Anti-virus software, systematic plan for updating and patching all applications and operating systems, firewall, anti-intrusion software, restricted physical access)
☐	**Hardware**	
		Does the physical location where your computers, servers, and data storage reside have appropriate security controls?
☐	**Backups**	
		• How will the data be backed up?
		• Where are the backups stored? (network drives, remote/cloud storage)

Data Sharing

What data sharing policies and data use agreements do you need to consider?

☐	**Who**	
		Identify potential reusers of the project data
☐	**Privacy/Confidentiality**	
		How will you protect the identity of participants during data sharing (i.e., honest broker, anonymized data) according to the Common Rule, FERPA, and HIPAA?
		How will sensitive data be handled to ensure it is stored and transferred/shared securely?
		How might managing identifiers negatively affect the usability of the data set for secondary analysis?
☐	**Availability**	
		Desribe how others might find your data (i.e., discipline-specific repository, proprietary repository)
		If you have questions about identifying a repository, contact an RDS consultant: http://researchdata.wisc.edu/contact-us/
☐	**Access**	
		Submit data (and relevant code) to a reputable DOI-issuing repository
		Registry of Research Data Repositories: https://www.re3data.org/
☐	**Restrictions & Use Conditions**	
		Will data sharing be postponed/restricted? (to publish or seek patents)
		What are the circumstances of the contract termination/data destruction for the requester using your data?
		Do you have a Data Use Agreement (DUA)?
		A DUA is an agreement between the data producer and secondary data user, and may impose rules for reuse, storage, re-dissemination and disposal/termination
☐	**Citations and acknowledgement**	
		"Data citation helps promote the reproduce-ability of research results. It allows us to track the usage and impact of data and it provides a structure by which we can recognize and reward data creators." – DataCite
		Details to include when citing data:
		• Is there a persistent ID? (DOI, ORCID, etc.)
		• What is being cited? (dataset, map, sound file, website)
		• Creator/author
		• Title
		• Version

		• Geography or origin
		• Database name and accession number (sequence data)
		• Date of download/access
		If the data are unpublished, the citation principles still apply. For example, if someone shared data with you via an email attachment, you can reference it as a private communication. ***Always provide information when you are citing data to help users find it.***
Legal and Ethical Issues: To effectively share data, researchers should first resolve any data ownership issues		
☐	**Ownership**	
		Who owns the data (PI/institution/funder/other)
		If you move to a new institution, what records are you allowed to take?
☐	**Copyright/Intellectual Property Rights (IPR)**	
		If used, are there any restrictions on the use of third-party data?
		Who will own the copyright and IPR of any data that you will collect or create, along with the license(s) for its use and reuse?
		For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR)
☐	**Grant or contract**	
		Does the sponsor or contract have any requirements?
☐	**License for reuse**	
		Scientific and factual data: many researchers apply a **Creative Commons 0 license**, which allows for free distribution of the data
		Creative works: there are various levels of Creative Commons licenses
		• Choose a CC License: https://creativecommons.org/choose/
		Software: MIT License, GNU GPL, or Apache Licenses
		For more information about license options, see this lesson from the Libraries' micro-course on Ethical Considerations for Responsible Research:
		License Options

Data Repositories

What repository is best for your data? Some repositories charge users for depositing or curation services. If cost becomes a factor in selecting an appropriate repository, this can be included in the budget portion of your research proposal.

Re3Data: Registry of Research Data Repositories: https://www.re3data.org/

☐	**Selecting a repository**	
	When selecting a repository, think about *how long the repository will preserve your data*, and *what will happen to your data if the repository goes away*	
		• In which repository or archive will the data be held?
		• What costs, if any, will your selected repository or archive charge? *(This can potentially be included in any budget for your project.)*
		• Does the repository support the creation of unique citations/DOIs?
		• Does the repository support your file format?
		• Is there a size limit per file?
		• Is there a size limit for the full dataset?
☐	If there isn't an appropriate repository for your data or if your data is a, you'll need to consider the following questions:	
		• How will you provide access to your data?
		• How long will your data retention period be?
☐	**Access**	
		Who can find and access deposited data?
☐	**User**	
		Is there journal-integrated, anonymous access (for pre-publication peer review)
		Are there tiered access roles and settings?
		Is there an optional embargo for data release post-publication?
☐	**Data**	
		• Is there data access via direct download? API?
		• Are there built-in tools to read proprietary file formats?
		• Are there integrated data analysis tools?
		• Are there comprehensive data and metadata search tools available?
☐	**Depositing Data**	
		• Have you planned for cost, time, and effort to prepare the data for sharing/preservation?
		• What fees are involved in deposit and maintenance?

Data Retention

Appraisal for long-term storage, permanent archival retention, and destruction

Retention requirements may depent on a variety of factors, including the type of data, the purpose for which the data has been collected, the policies of funding institutions, and UW-Madison's policies. UW-Madison has specific requirements for research data, including an interest in permanently keeping some of these records as a part of its institutional history or intellectual property

☐	**Data appraisal**	
		What data must be retained/destroyed for contractual, legal, or regulatory purposes?
		How long will the data be retained and preserved?
☐	**Storage**	
		Refer to our recommendations about data storage and backup, and it is also encouraged to contact UW-Madison's DoIT team or your departmental IT for more specific data storage and backup recommendations and options. http://researchdata.wisc.edu/data-storage-backup/
☐	**Archiving data**	
		What are the foreseeable research uses for the data?
		What are the essential records required to understand this research data and project?
		Is the research data replicable?
		Has the research been published?
	A small percentage of data and related records might be identified for permanent storage as a part of the historical record of a discipline or institution, or as intellectual property. Records eligible for permanent retention may be those that: document a breakthrough, are generated by a lab or individual who had a great impact on the field, or are highly reusable in a particular area of research	
	Permanent retention, or archiving, is often a significant investment for an institution, as it implies ongoing migration of electronic formats and storage costs, as well as care, maintenance, and access services for the records in perpetuity. This is not the same as ensuring long-term storage or preservation of research data.	
☐	**Data Disposal**	
		How will you permanently remove sensitive data/project data?