

HBS Research Data Management Checklist

This document serves as a reference checklist to keep track of the elements that make up good research data management (RDM) in the lifecycle. Remember that it is not linear and you may find yourself jumping around this lifecycle throughout your project.

Planning and Data Acquisition/Collection

<input type="checkbox"/> Project description (background/rationale)	<ul style="list-style-type: none"> * What research question(s) are you addressing? * What study methods and design will you use?
<input type="checkbox"/> Data description and collection	<ul style="list-style-type: none"> * What instruments/questionnaires will you use? * What format will your data be in (audio, text, database, images)? * How large do you anticipate your files will be? * How will you collect or acquire your data?
<input type="checkbox"/> Participating researchers	<ul style="list-style-type: none"> * Who will be a part of the research study? * Are all researchers HBS affiliated, or will there be individuals from outside organizations?
<input type="checkbox"/> Ethical review and legal compliance	<ul style="list-style-type: none"> * Does your data include human subjects and need to go through the Institutional Review Board (IRB)? * Does your data include sensitive information (e.g., birthdates, social security numbers)? * Are you collecting or obtaining data from the European Union that is subject to General Data Protection Regulation (GDPR)? * If you are obtaining existing data, what agreements need to be put in place (e.g., Data Use Agreement; Memorandum of Understanding; Nondisclosure Agreement) and reviewed by Harvard?
<input type="checkbox"/> Planning for data sharing	<ul style="list-style-type: none"> * Might you have any requirements later on to share the data you collect (e.g., from a journal in which you may want to publish)? If so, that will impact how you collect the data (e.g., license provisions, consent forms, etc.).

Storage, Security, and Analysis

<input type="checkbox"/> Security and storage	<ul style="list-style-type: none"> * What security level is your data? * How will sensitive data be stored and transferred? * Can your data be stored on local machines (e.g., laptop/desktop), or will you require a more secure location such as the HBS computing environment? * Is your project data so large that you need to consider cloud storage? * Are you working with non-HBS collaborators that will require access to your project data? How will you securely share data with them?
<input type="checkbox"/> Software	<ul style="list-style-type: none"> * What research software will you and the project team need (e.g., Stata, R)? * Will you need any research software beyond what HBS licenses?



<input type="checkbox"/> Project organization	<ul style="list-style-type: none"> *How will you document project methods/procedures? *How will you document the contents of your data files (both at a high level and at the detailed variable level) so that others can understand them in the future? *What file organization and naming conventions will you use? *What type of version control system will you employ?
---	---

Dissemination and Preservation

<input type="checkbox"/> Storing your data post-project	<ul style="list-style-type: none"> *Where will you store your data post-project (until you figure out what you will do with it long-term)? *What data must be retained/destroyed for contractual, policy, legal, or regulatory purposes? For what period of time?
<input type="checkbox"/> Data sharing and repositories	<ul style="list-style-type: none"> *Might data sharing be a requirement of your funder or journal? Do you want to consider data sharing in order to promote your work and expand the impact of your research? *Do you have permission to share the data? *How will you de-identify identifiable and/or sensitive data for sharing? *What repositories might be a good place to share your data?
<input type="checkbox"/> Archiving	<ul style="list-style-type: none"> *How long will the data be retained/preserved? *What aspects of the data may have long-term value and should be archived permanently?

Partially adapted from the Harvard Medical School Research Data Management checklist:
<https://datamanagement.hms.harvard.edu/files/bdata/files/2018-rdm-checklist.pdf>