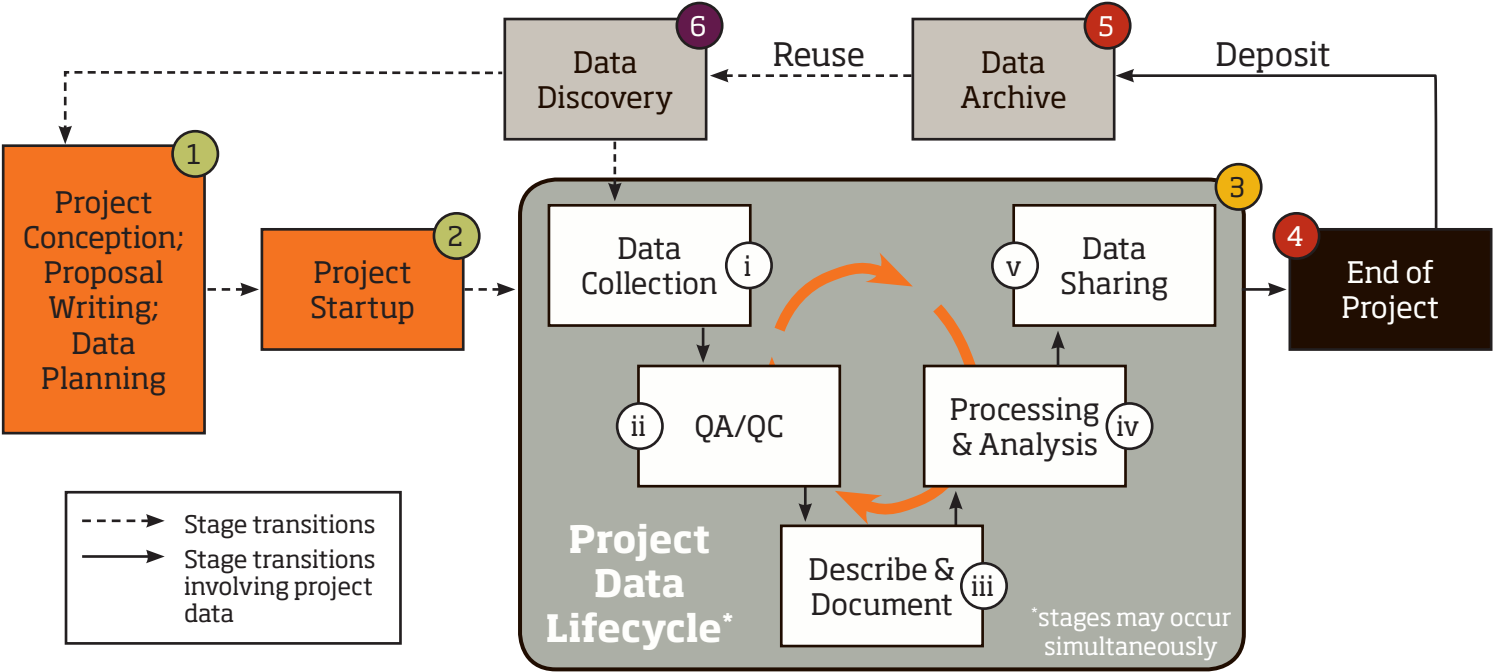


Data Management During the Research Lifecycle



Each of the boxes above represents a stage in the research lifecycle for a given project. Transitions between stages involving project data (solid black arrows) are vulnerable to data loss. These phases are often coincident with shifts in responsibility for the data (e.g. from data collector to data analyst, or from graduate student to PI), which makes transitions a critical time for active data management. Data loss may occur in many other ways, for various reasons: data are unusable due to lack of necessary contextual details (lack of sufficient documentation/metadata); failure to move data to the next stage, reaching a dead-end prior to systematic preservation; bit rot or other digital degradation; or data are physically misplaced. The lifecycle model is used to provide context in describing data stewardship activities that should take place over the course of a research project.

Research Lifecycle Stages & Data Management Actions

Project Planning

- 1 Project ideas are developed and fleshed out in a proposal. Review existing datasets. Determine where you will archive the data (your funder may have a required location), and consult with the archive about metadata and data formatting requirements. Plan consent for data sharing, if needed. Determine costs related to data storage & archiving. Data management plan (DMP) is created.
- 2 Finalize plans for data documentation procedures. Revise DMP if necessary. Communicate data responsibilities to project members. Establish, document and communicate protocols and methods.

Project Data Lifecycle (note: all 'stages' may occur simultaneously)

- 3
 - i. Organize files; use consistent, thoughtful file-naming conventions; carry out regular backups; consider access control and security.
 - ii. Assign clear roles for QA/QC; check instrument precision, bias, and scale; replicates where possible; use controlled vocabulary on log/data sheets; check for out-of-range values, empty cells, etc.
 - iii. **Document the context of data collection:** project history, aim, objectives and hypotheses; data collection methods: sampling, data collection, process, instruments used, hardware and software used, scale and resolution, temporal and geographic coverage and secondary data sources used; dataset structure of data files, study cases, relationships between files; data validation, checking, proofing, cleaning and quality assurance procedures carried out; changes made to data over time since their original creation and identification of different versions of data files; information on access and use conditions or data confidentiality.
Data-level documentation: names, labels, units and descriptions for variables, records and their values; explanation or definition of codes and classification schemes used; definitions of specialist terminology or acronyms used; codes of, and reasons for, missing values; derived data created after collection, with code, algorithm or command file; weighting and grossing variables created; data listing of annotations for cases, individuals or items.
 - iv. Document and manage file versions; document file manipulations; document analysis procedures; write software codes with future sharing in mind.
 - v. Determine appropriate file formats; finalize data documentation; share data within project team for analysis and interpretation.

Project Wrap-up

- 4 Write paper(s); contact archive for up-to-date requirements; deposit data in archive or repository; establish links between dataset(s) and paper(s) via citation and the use of identifiers (DOI, ORCID, etc.); perform data management audit against DMP; submit final report to funder.
- 5
- 6 Data are available for discovery and reuse, by you and others.

