WOMEN'S HEALTH RESEARCH INSTITUTE AT BC WOMEN'S

AI tools were applied in this document to improve clarity and grammar. The content, structure, and analytical framework were developed independently by the author.

Data principles and governance

This document is intended for research investigators, data managers, and study teams involved in the design and conduct of health-related studies. It supports protocol development, ethics submissions, data management training, and institutional planning. While relevant across disciplines, it is particularly aligned with projects that involve sensitive health data, administrative data linkage, or multi-site coordination.

Introduction

High-quality data are the foundation of any successful project, whether large or small. Without reliable and well-managed data, even the most sophisticated analyses can lead to misleading or inconclusive results. The integrity, consistency, and completeness of data are not merely technical considerations—they are central to the credibility and utility of research findings.

Large-scale projects often involve complex data ecosystems. These may include multiple data sources (such as surveys, administrative records, clinical data, or electronic health systems) collected at different time points and across different settings. These projects typically support multiple research questions and hypotheses and require coordination among multidisciplinary teams, institutions, and even jurisdictions. Managing such complexity demands a clear and standardized approach to data documentation, cleaning, validation, and governance.

Equally important are small-scale studies. These projects are often more focused, addressing highly specific research questions with greater granularity. Many serve as pilot studies or feasibility assessments, generating critical insights that inform the design of future larger-scale investigations. Despite their scale, the rigor applied to data management in smaller studies should be no less than in large ones, as errors or inconsistencies can be amplified in smaller datasets and lead to biased or non-replicable findings.

This document outlines the principles and procedures for working with data at all stages of a project lifecycle—from planning and collection through cleaning, analysis, and archiving. It is designed to ensure transparency, reproducibility, and scientific integrity, regardless of the project's size or scope.

Key Terms

Causal Diagram / Causal Map**:** A visual representation (often a **Directed Acyclic Graph** or **DAG**) that outlines hypothesized relationships among variables in a study. It helps identify exposure(s), outcome(s), confounders, mediators, and potential sources of bias. Causal diagrams guide decisions about variable selection, model structure, and the appropriate interpretation of associations in observational research.

Data Management Plan (DMP): A document outlining how data will be managed throughout a research project. It typically covers data collection methods, storage, cleaning, quality control, documentation, sharing, and archiving. A DMP ensures data integrity, supports reproducibility, and aligns project practices with governance and ethical standards. It can be brief for small projects or highly detailed for complex or multi-site studies.

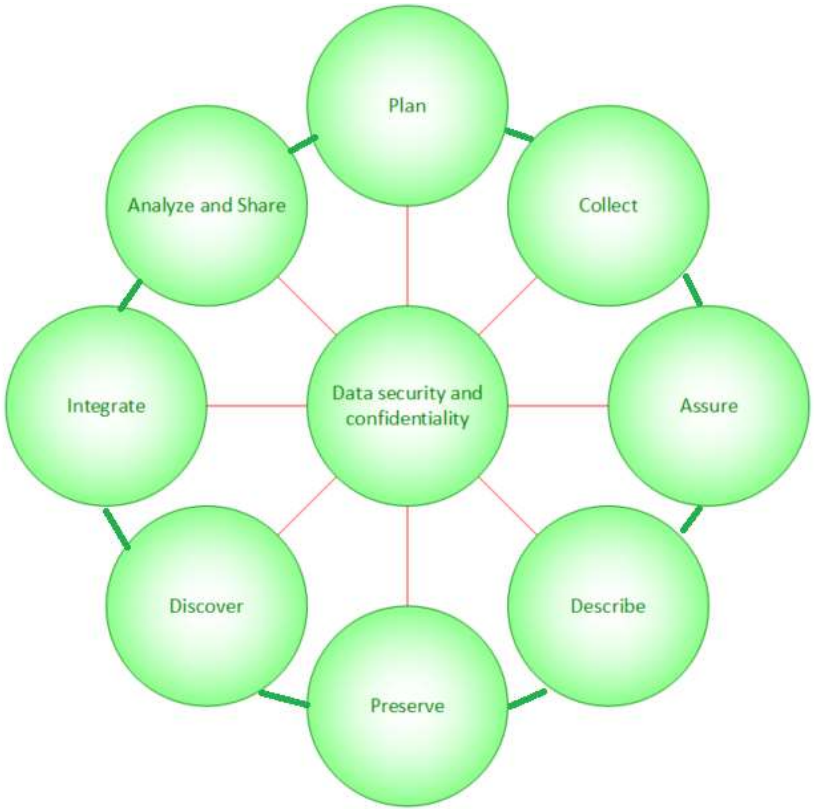
Data Governance: A framework of policies, procedures, and oversight that ensures responsible, secure, and ethical handling of data. It addresses issues such as data access, security, privacy, ownership, retention, and compliance with institutional, legal, or regulatory requirements. Governance is embedded throughout the data cycle and often operationalized through a DMP, access controls, and audit trails.

Data Dictionary: A structured document or table that defines each variable in a dataset. It includes variable names, descriptions, formats (e.g., numeric, categorical), allowable values, coding schemes, and—when applicable—reference to validated instruments. A good data dictionary ensures consistency, facilitates analysis, and supports collaboration and data reuse.

Data cycle: This diagram illustrates the data lifecycle—a structured approach to managing data across its entire journey in a research or project setting. At the core is Data Security and Confidentiality, emphasizing the protection of sensitive information at all stages. Surrounding this central principle are eight interconnected phases:

Data lifecycle

The data lifecycle provides a structured framework for managing research data from initial planning through to analysis and sharing. It promotes transparency, efficiency, and data integrity across all stages of a project. Below is an overview of each key phase, highlighting its role in supporting high-quality, ethical, and reusable data.



Plan: Define data needs, sources, formats, and governance structures before data collection begins.

Collect: Acquire data using standardized tools and ethical protocols, ensuring accuracy and completeness.

Assure: Implement quality control procedures to validate, clean, and verify data integrity.

Describe: Document metadata, variable definitions, and data structures to support understanding and reuse.

Preserve: Store data securely, with attention to backup, access control, and long-term archiving.

***Discover**:* Make data findable and accessible for authorized users, often through indexing or cataloging.

***Integrate**:* Combine datasets across sources or timepoints to support richer analyses and broader insights.

***Analyze and Share**:* Conduct statistical or qualitative analysis and disseminate findings, ensuring proper attribution and ethical use.

Each phase is interconnected, reinforcing the need for thoughtful design, rigorous data handling, and strong governance.

Data governance

Data governance refers to the overarching framework of policies, procedures, standards, and roles that ensure data is managed responsibly, securely, and effectively throughout its lifecycle. In the context of the data cycle, data governance acts as the backbone that supports integrity, accountability, and compliance at each stage—from planning through sharing. Data governance also ensures that appropriate safeguards (such as access controls, encryption, de-identification, and data use agreements) are implemented and monitored across all stages of data handling.

Governance Tied to Each Stage

***Plan**:* Begin with establishing data management plans, clarifying ownership, ethical approvals, and defining roles and responsibilities. It sets the foundation for consent, access control, and data classification.

***Collect**:* Ensure adherence to protocols for informed consent, data accuracy, and regulatory compliance (e.g., HIPAA, PIPPA). It mandates standardized formats and procedures for consistency and auditability.

***Assure**:* Enforce quality assurance standards, including validation checks, documentation of cleaning processes, and audit trails to ensure reproducibility and transparency.

***Describe**:* Develop robust metadata practices—ensuring datasets are accompanied by comprehensive descriptions, dictionaries, and coding schemas that facilitate responsible reuse and interpretation.

***Preserve**:* Provide guidelines for secure storage, retention periods, backup protocols, and disaster recovery. Identify roles for data stewardship and administration to ensure ongoing custodianship.

***Discover**:* Develop framework to support discoverability by regulating indexing and access rights, through data catalogs or repositories, while balancing openness with ethical constraints.

***Integrate**:* Establish rigorous protocols for data linkage and integration to maintain data lineage and interoperability. Develop procedures for harmonization, provenance, and consent compatibility across datasets and systems.

***Analyze and Share**:* Ensure appropriate use of data in analysis and dissemination. This includes data use agreements, review of outputs for disclosure risk, and defining who has authority to publish or distribute findings.

Data governance is not a separate function—it is embedded in every step of the data lifecycle. It fosters accountability, protects individuals, ensures compliance, and enables trustworthy and ethical data use. When implemented effectively, governance turns data from a liability into a valuable, secure, and reusable asset.

Data management plan (DMP)

DMP serves as a practical tool to connect data governance principles with the operational steps of the data lifecycle. While not all projects require a formal, detailed DMP, especially smaller-scale studies, every project should outline at least the basic steps of how data will be managed—from collection to sharing. A well-structured DMP helps translate governance policies into day-to-day practices, ensuring that responsibilities, access rights, quality checks, metadata standards, and preservation strategies are clearly defined and consistently applied. By embedding these considerations early in the planning stage, a DMP supports transparency, efficiency, and ethical compliance throughout the data lifecycle. It can include but not limited for the following steps:

Data collection

Clearly define what data will be collected. Begin by specifying which variables are necessary to answer the primary research questions or evaluate key hypotheses. This process should be guided by the study's causal framework. The causal diagram helps to identify which variables are:

- Exposures, outcomes, and mediators
- Potential confounders that must be adjusted for

- Effect modifiers or stratification variables
- Proxies for constructs that cannot be measured directly

Variables should be grouped into logical domains to ensure a comprehensive and organized data structure. Common domains include:

- Demographics: age, sex at birth, gender identity, race/ethnicity, education, income, immigration status.
- Clinical characteristics: diagnoses, physiological measurements, lab results, medication use, reproductive history.
- Psychosocial constructs: depression, anxiety, social support, perceived discrimination—often measured through validated scales (e.g., PHQ-9, EPDS, GAD-7).
- Behavioral and lifestyle factors: smoking, alcohol use, physical activity, diet, sleep, substance use.
- Health system use: primary care visits, hospitalizations, medication adherence, specialist referrals.
- Environmental/contextual variables: postal code (linked to neighborhood-level data), housing stability, access to services.

In some cases, proxy variables will be used to represent constructs that are difficult to measure directly (e.g., "trust in the healthcare system" may be approximated by continuity of care or number of providers seen). It is important to justify proxy selection within the causal model and clearly document the assumptions.

This step is foundational for developing data collection tools, preparing ethics submissions, and building analysis-ready datasets. All variables should be defined in a preliminary data dictionary—including names, formats, coding schemes, and reference instruments where applicable.

Specify data sources

Identifying the source of each data element is essential for planning, documentation, validation, and governance. It ensures that data is collected from appropriate, reliable channels and that its origin is transparent and traceable throughout the research lifecycle.

Data sources should be aligned with the study's design, ethical requirements, and data governance structures. Each variable in the data dictionary should be linked to one or more of the following source types:

- Participant-Reported Data: Collected directly from participants through self-report tools. This includes data captured via interviews, paper or electronic surveys, mobile apps, diaries, or verbal responses. These sources are commonly used to gather subjective experiences, perceptions, or behaviors.
- Clinician-Reported or Study Staff Data: Collected and entered by trained personnel such as clinicians, nurses, or research assistants. This includes data based on observation, clinical judgment, or structured assessments conducted during study visits. Standardized forms or case report forms (CRFs) are typically used.
- Electronic Health Records (EHRs): Data passively recorded as part of routine clinical care and stored within institutional or provincial EHR systems. These may include diagnoses, procedures, medications, vital signs, laboratory test results, and visit summaries. Extraction may be manual or automated.
- Administrative Health Data: Data collected by health authorities or insurance systems for purposes such as billing, reporting, and surveillance. These datasets include hospitalizations, physician billing claims, prescription dispensing, and birth/death records. Access typically requires data-sharing agreements.
- Registry Data: Structured datasets collected for disease surveillance, quality monitoring, or health system evaluation. These registries often include specific clinical or diagnostic criteria and are maintained at the institutional, provincial, or national level.
- External Linked Data: Supplementary data from non-health sectors that may be linked to research data for contextual analysis. This may include census data, geographic indicators, education or social services records, and environmental exposures.

Outline collection methods

The method by which data is collected directly impacts its validity, reliability, and suitability for analysis. Choosing appropriate collection methods requires consideration of the type of variable, the population

involved, logistical feasibility, and potential for bias. Each method should be clearly documented to support transparency, reproducibility, and ethical oversight.

Below are common categories of data collection methods:

- Self-Report: participants provide information directly, often through questionnaires, surveys, or diaries. This method is particularly useful for collecting subjective data such as symptoms, attitudes, behaviors, or perceptions. It can be administered in paper form, electronically (e.g., online, or tablet-based), or verbally. This is best suited for psychosocial variables, quality of life, lifestyle behaviors, knowledge, and beliefs.

Please consider this limitation when collecting self-reported data. It is susceptible to recall bias, social desirability bias, and literacy limitations.

- Observer-Rated: a trained individual (e.g., research assistant, clinician) collects or scores data based on observation or structured interviews. Observer-rated methods are useful for collecting data that participants may not be able to report reliably or that require clinical judgment or behavioral assessment. This is best suited for the functional assessments, behavioral observations, standardized clinical tools.

Please acknowledge the following limitations. It requires training and standardization to ensure inter-rater reliability.

- Automated or Systematic Extraction: data are retrieved from existing digital systems such as electronic health records (EHRs), administrative databases, or monitoring devices. This method can be manual (e.g., chart abstraction) or automated (e.g., SQL queries, data warehouse pulls). It best suited for clinical measurements, utilization data, laboratory values, medication records.

Additional considerations when using this collection method. It will depend on data system structure, coding accuracy, and technical access.

- Hybrid Methods: in some cases, the same variable may be collected through multiple methods to improve completeness or validation. For example, a self-reported medication list may be verified against prescription records or clinician-entered forms.

- Passive Monitoring: data collected through wearable devices, sensors, or system logs without active input from participants. Increasingly used in digital health and behavioral research. It best suited for physical activity, sleep, geolocation, screen time.

Some limitations: requires consent for continuous monitoring, data volume management, and privacy safeguards.

For each variable in the data dictionary or collection instrument, the collection method should be specified along with any associated tools (e.g., software, devices, interview guides) and the personnel involved. This ensures consistency and allows for proper training, data auditing, and methodologic transparency.

Indicate timing and frequency of data collection

Clearly specifying when each data point will be collected is critical for ensuring consistency, aligning with the study design, and enabling meaningful longitudinal or time-based comparisons. The timing and frequency of data collection should correspond to the conceptual framework (e.g., causal diagram), study objectives, intervention phases, and expected changes in outcomes.

- Baseline: data collected prior to any exposure, intervention, or follow-up, serving as the reference point for each participant. It ensures comparability across study groups and allows for adjustment of pre-existing differences. Typically includes demographics, pre-existing conditions, initial outcome measures. This is usually collected at enrollment or pre-intervention visit.
- Intervention or Exposure Period: for interventional studies, data may be collected during or immediately after an intervention to monitor change, adherence, or immediate effects. It may include mid-intervention check-ins, process measures, short-term outcomes. The timing is aligned with protocol (e.g., weekly during a 6-week program, and etc.).
- Follow-Up: collected after the intervention or exposure to assess changes in outcomes over time. Multiple follow-up time points may be used to evaluate short-, medium-, and long-term effects. For examples: 3-month, 6-month, and 12-month follow-up visits or surveys or repeated measures for longitudinal modeling.
- Event-Based or Triggered Collection: some data are collected in response to specific events, such as hospital admissions, adverse outcomes, or reaching a certain gestational age. For example, collecting delivery data upon birth, or symptom escalation reports in real-time.

- Continuous or Routine Monitoring is used in surveillance studies or when working with administrative or digital data sources. For example, daily symptom tracking apps, monthly health service usage reports, real-time device data.

Specify Units of Time

For all the above, be explicit about the units of measurement (e.g., days, weeks, months, gestational age in weeks) and align data collection windows with clinical relevance and participant burden.

Each variable should be associated with one or more collection time points in the data dictionary or protocol. Timelines should be visualized in a schedule or table (e.g., SPIRIT diagram) to assist with coordination, ethics review, and adherence monitoring.

Identify responsible personnel

Assigning clear roles for data collection and management is essential for ensuring accountability, consistency, and data quality. The personnel involved may vary by setting, data type, and study phase. Their responsibilities should be clearly documented in the data management plan, protocol, or standard operating procedures, and all individuals should receive appropriate training and oversight.

Research Assistants (RAs)

Often responsible for direct data collection from participants, especially for:

- Administering surveys or structured interviews (in-person, phone, or online)
- Entering data into platforms such as REDCap
- Verifying completeness of forms and flagging inconsistencies
- Assisting with recruitment and informed consent procedures

RAs should be trained on data collection protocols, confidentiality, use of tools, and standardized instruments.

Clinicians (e.g., Nurses, Physicians, Midwives)

Responsible for collecting or recording clinically relevant data, often as part of routine care. This may include:

- Documenting diagnoses, adverse events, or clinical measurements
- Completing delivery forms or postnatal assessments

- Providing expert judgment in observer-rated tools
- Ensure alignment between clinical documentation and study CRFs; clinical personnel may need separate training for research procedures.

Site Coordinators

Typically oversee site-level data operations and are responsible for:

- Supervising RAs and clinicians in data collection
- Conducting quality checks or audits
- Coordinating data transfer and reporting schedules
- Ensuring protocol compliance and resolving discrepancies

Site coordinators act as the liaison between field teams and central data managers or investigators.

Data Managers / Analysts

Although not directly involved in collection, they play a key role in:

- Designing data collection instruments and workflows
- Validating data structure and formatting
- Managing secure storage, cleaning, and version control

They may also build audit logs, dashboards, or real-time monitoring tools for field staff.

Principal Investigators or Project Leads

Provide overall oversight and are responsible for the integrity of data collection. Their role includes:

- Ensuring ethical and regulatory compliance
- Approving training plans and data workflows
- Making final decisions on handling missing or problematic data

Describe standardization protocols

Standardization protocols are essential for ensuring that data is collected consistently across participants, sites, and time points. They minimize variability due to human error, interpretation differences, or equipment inconsistencies, thereby enhancing data quality and comparability. These

protocols should be clearly defined and implemented through training, documentation, and regular monitoring.

Interviewer Scripts and Guides

Structured scripts ensure that all participants receive the same instructions and wording, reducing interviewer bias and improving reproducibility. These include:

- Introduction and consent language
- Standardized question phrasing
- Neutral probing techniques
- Closing scripts and participant reminders

Use of interviewer scripts is particularly important for psychosocial or self-reported data collection (e.g., mental health surveys).

Calibration Procedures

Equipment used for measurements (e.g., blood pressure cuffs, scales, thermometers) must be regularly calibrated to maintain accuracy.

- Set calibration schedules (e.g., weekly, monthly)
- Maintain calibration logs and tracking forms
- Ensure standard operating procedures are followed for equipment handling and positioning

Use standardized equipment across sites and train staff using mock scenarios or reference manuals.

Form Templates and Case Report Forms (CRFs)

Predefined forms help standardize:

- Variable names and coding (e.g., Yes = 1, No = 0)
- Field formats (e.g., dropdowns, date formats, units of measurement)
- Skip logic and validation rules
- Data completeness checks

Please note that CRFs should be piloted before full rollout to identify any issues with clarity or usability.

Training Manuals and Checklists

Develop manuals that describe how to administer tools, conduct measurements, and resolve discrepancies. Training should include:

- Protocol walkthroughs
- Role-playing or simulation exercises
- Inter-rater reliability assessments (for observer-rated tools)
- Provide job aids or quick-reference guides for field staff.

Electronic Standardization Features

Use of digital tools (e.g., REDCap, ODK) enables built-in controls. These features reduce error at the point of entry and support audit trails.

- Range checks and required fields
- Branching logic for skip patterns
- Automated timestamps and user IDs

All standardization protocols should be included in the study's operations manual and referenced during training and audits. Deviations from standard procedures should be logged and reviewed as part of quality assurance.

Document any data entry tools or platforms

Clearly specifying the tools and platforms used for data entry ensures transparency, supports reproducibility, and facilitates data security and quality control. The choice of data entry method should align with the study design, resource availability, participant population, and data governance policies. Each method has implications for training, error prevention, accessibility, and downstream data handling.

Paper-Based Case Report Forms (CRFs)

Used when digital infrastructure is limited or when field conditions make electronic data collection impractical. Those usually are required double data entry or verification to reduce transcription errors.

The documents must be stored securely in locked cabinets with restricted access. Scanned copies can be archived as backups or for audit purposes.

Electronic Data Capture (EDC) Systems

Platforms such as REDCap, ODK, Qualtrics, or Castor EDC are widely used for structured, secure, and scalable digital data collection. They offer field validation (e.g., required fields, range checks), support skip patterns, timestamps, and user-level access control. Those also allow for real-time monitoring and centralized data management. Those platforms are Ideal for clinical trials, large cohort studies, or projects with remote data collection.

Tablets, Laptops, or Mobile Devices

Used for point-of-care data entry or in-person survey administration and can be configured to work offline and sync later. Those devices enable quick validation and reduce time between collection and entry. They may include secure device locking, encryption, and device tracking. Usually, those are well suited for community-based or household surveys and multi-site studies.

Secure Web-Based Surveys

Useful for remote or self-administered data collection from participants. Those are accessed via emailed links or participant portals and may include IP tracking, captcha validation, and multi-language options. Those typically hosted on encrypted, GDPR- or HIPAA-compliant servers. Online surveys are commonly used for longitudinal follow-up, symptom tracking, or experience-of-care surveys.

Custom-Built Tools or APIs

For advanced integration with clinical systems or registries, custom interfaces may be used.

They allow automation of EHR or administrative data extraction and can be set to pull structured fields on a predefined schedule. Usually are require collaboration with IT and compliance with institutional data security policies.

Each tool should be documented in the data management plan along with:

system name and version, access controls and user permissions, backup, and recovery protocols,

audit trail functionality. Proper documentation supports ethical compliance, platform validation, and eventual archiving or data sharing.

In summary, a well-structured Data Management Plan (DMP) operationalizes data governance principles across each stage of the data lifecycle—from planning and collection through to preservation and sharing. While not all studies require a detailed DMP, every project benefits from outlining how data will be collected, stored, processed, and secured. The DMP serves as a living document that translates abstract governance requirements—such as confidentiality, accountability, and access control—into practical, enforceable procedures tied to real workflows. When anchored to a project's causal diagram, the DMP ensures that all essential variables, including exposures, confounders, and proxies, are identified and consistently managed across time points and sources.

Each component of the DMP reflects a corresponding phase of the data cycle and reinforces the need for quality assurance, ethical compliance, and methodological transparency. From specifying data sources and collection methods to defining timing, assigning roles, and standardizing tools, the DMP creates alignment between governance frameworks and research operations. This structure not only safeguards data integrity but also facilitates reproducibility, supports regulatory compliance, and enables responsible data sharing. The integration of the DMP into the data lifecycle supports a culture of accountability and scientific rigor throughout the project lifespan.

Difference between Data Governance and Data Management

In healthcare research and clinical practice, data governance is a critical component within the broader framework of data management. While data management refers to the overarching practice of collecting, storing, processing, and using data effectively and securely, data governance focuses specifically on the policies, standards, and oversight that ensure data is managed ethically, legally, and responsibly.

Data management encompasses the full data lifecycle, including areas such as data collection methods, cleaning procedures, system architecture, data quality assurance, and data security. These functions are deeply interconnected. For example, the decisions made in one area (e.g., data access) directly impact others (e.g., privacy compliance). Because of this interconnectedness, data governance cannot function

in isolation. Collaboration across teams—researchers, data managers, clinicians, IT specialists, and ethics boards—is essential to developing and maintaining an effective governance strategy.

For example, a data governance team in a hospital might develop policies on the use of electronic health records (EHRs) for secondary research, including rules for de-identifying patient data and defining who may access what type of information. The data management team then implements these rules by configuring systems (e.g., REDCap, EPIC, or provincial repositories), enabling secure access through role-based controls, and ensuring technical compliance. Similarly, when a study seeks to link survey data with administrative datasets (e.g., MSP or PharmaNet), governance teams ensure that consent and privacy requirements are met, while data managers design and execute the technical processes for secure data integration.

Summary

This document provides a foundational overview of the principles and practices that support high-quality data management in research, regardless of project size. Central to this framework is the data lifecycle—a structured sequence of phases from planning and collection to analysis and sharing. At its core lies data governance, which ensures that ethical, legal, and institutional responsibilities are upheld at every stage. A strong governance structure not only protects participants but also ensures the reproducibility, transparency, and long-term value of research data.

The Data Management Plan (DMP) serves as a critical bridge between governance policies and operational workflows. It translates abstract principles such as confidentiality, access control, and stewardship into specific procedures for data collection, standardization, documentation, and oversight. Aligned with the causal diagram and research questions, the DMP details which variables are needed, how they will be measured, who is responsible for collecting them, and when. These specifications support accuracy, consistency, and auditability while aligning with broader data integrity and compliance goals.

This initial guidance focuses on the early phases of the data lifecycle—planning, collecting, assuring, and describing—laying the groundwork for high-quality, ethically governed datasets.

Although this document focuses on early phases of the data lifecycle, projects should also plan for potential data sharing and secondary use. This includes decisions about data de-identification, repository

submission, and the terms under which future investigators may access the data. These considerations are central to responsible and ethical data governance.

Future documents will expand on later stages, including preservation, discovery, integration, and sharing, to ensure that data remains secure, accessible, and reusable long after the study concludes. Together, these materials will support a unified, lifecycle-based approach to data stewardship across all projects.

References

Arts, D. G. T., De Keizer, N. F., & Scheffer, G. J. (2002). Defining and improving data quality in medical registries: A literature review, case study, and generic framework. Journal of the American Medical Informatics Association, 9(6), 600–611. https://doi.org/10.1197/jamia.M1087

DAMA International. (2017). The DAMA guide to the data management body of knowledge (DAMA-DMBOK®) (2nd ed.). Technics Publications. https://www.dama.org/cpages/body-of-knowledge

European Open Science Cloud. (n.d.). Data life cycle. RDMkit. https://rdmkit.elixir-europe.org/data_life_cycle

Government of British Columbia. (1996). Freedom of information and protection of privacy act. https://www.bclaws.gov.bc.ca/civix/document/id/complete/statreg/96165_00

Government of British Columbia. (2003). Personal information protection act. https://www.bclaws.gov.bc.ca/civix/document/id/complete/statreg/03063_01

Government of Canada. (2021). Tri-agency research data management policy. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada.

https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-research-data-management-policy

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. Epidemiology, 10(1), 37–48. https://doi.org/10.1097/00001648-199901000-00008

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. Journal of Biomedical Informatics, 95, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing

translational research informatics support. Journal of Biomedical Informatics, 42(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Kahn, M. G., Brown, J. S., Chun, A. T., Davidson, B. N., Meeker, D., Ryan, P. B., Schilling, L. M., Weiskopf, N. G., Williams, A. E., & Zozus, M. N. (2016). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. eGEMs (Generating Evidence & Methods to Improve Patient Outcomes), 4(1), 1244. https://doi.org/10.13063/2327-9214.1244

National Institutes of Health. (2023). NIH data management and sharing policy. https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/dms

Partridge, E. F., & Bardyn, T. P. (2018). Research electronic data capture (REDCap). Journal of the Medical Library Association, 106(1), 142–144. https://doi.org/10.5195/jmla.2018.319

Peng, R. D. (2011). Reproducible research in computational science. Science, 334(6060), 1226–1227. https://doi.org/10.1126/science.1213847 https://www.science.org/doi/10.1126/science.1213847

Provincial Health Services Authority. (n.d.). Data access & privacy. https://www.phsa.ca/researcher/data-access-privacy

Provincial Health Services Authority. (n.d.). Research privacy. https://www.phsa.ca/researcher/data-access-privacy/research-privacy

University of British Columbia. (n.d.). Data governance. UBC Office of the Chief Information Officer. https://cio.ubc.ca/data-governance

University of British Columbia. (n.d.). Research data management. UBC Library. https://researchdata.library.ubc.ca/

University of British Columbia. (n.d.). Data management plans. UBC Library Research Data Management. https://researchdata.library.ubc.ca/plan/

Van den Broeck, J., Argeseanu Cunningham, S., Eeckels, R., & Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. PLOS Medicine, 2(10), e267. https://doi.org/10.1371/journal.pmed.0020267

Walther, B., Hossin, S., Townend, J., Abernethy, N., Parker, D., & Jeffries, D. (2011). Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. PLOS ONE, 6(9), e25348. https://doi.org/10.1371/journal.pone.0025348

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18