GCDMP

REVIEW ARTICLE

Data Management Plan

Evaldas Lebedys*, Carolyn Famatiga-Fay†, Priyanka Bhatkar‡, Derek Johnson§, Gayathri Viswanathan‖ and Meredith Nahm Zozus¶

Every clinical study should have prospective plans for how data will be collected, processed and stored. Likewise, every study should have defined data elements and objective evidence of how data were processed. This chapter outlines the purpose of, and regulatory basis for, such documentation in the form of a Data Management Plan (DMP). Although the clinical data manager (CDM) may not personally prepare all sections of the data management plan, he or she is often responsible for assuring that comprehensive data documentation exists.

1) Learning Objectives

After reading this chapter, the reader should understand

- The purpose of and regulatory basis for the DMP
- The contents and organization of the DMP
- Creation and maintenance of the DMP

2) Introduction

Although a study protocol contains the overall clinical plan for a study, separate plans such as organizational standard operating procedures (SOPs), work instructions, and study specific documentation, are necessary to fully specify study conduct, data collection, management, and analysis. Study specific details are often covered in site operations manuals, monitoring plans, DMPs, and Statistical Analysis Plans (SAPs). Compilation of necessary documentation in discipline-specific or functional plans is an approach to organizing and maintaining the essential documentation for a clinical study. A DMP comprehensively documents data and its handling from definition, collection, and processing to final archival or disposal. A thorough DMP provides a road map for handling data under foreseeable circumstances and also establishes processes for dealing with unforeseen issues.

Plans for data management have been variously named and composed since their early uses. Common variations include DMPs, data handling plans or data handling protocols. Likewise, specific components of data management have been described in more narrowly scoped documents including research data security plans, data sharing plans, manuals of procedures, and manuals of operations, that cover some but not all aspects of the data lifecycle in a clinical study. The Society for Clinical Data Management (SCDM) defines a DMP as a compilation of, or index to, comprehensive documentation of data definition, collection and processing, archival, and disposal, sufficient to support reconstruction of the data handling portion of a clinical study. Reconstruction may involve tracing data values in result tables, listings or figures back to their origin or vice versa.

Although not focused on data management, the Greenberg report (1967) first highlighted the need for monitoring and controlling performance in large multi-center clinical studies.[1] A 1981 article was the first noted call for "a detailed procedures manual describing all aspects of data intake and processing procedures" and description of such a document serving the multiple purposes of a manual for personnel to increase consistency of work, a guide to evaluate the procedures and a tool to assess adherence to the procedures.[2] The Association for Clinical Data Management (ACDM), in 1996, published the first resource for data management planning titled *ACDM Guidelines to Facilitate Production of a Data Handling Protocol.*[3] In 1995, a collection of five review papers documenting current practice in collection and management of data for clinical studies was published as a special issue of Controlled Clinical Trials (now Clinical Trials). This compendium was the first focused attention in the literature on data collection and management methodology including procedures and documentation.[4,5,6,7,8,9] In recognition of most industry sponsors approaching comprehensive data documentation through compilation in documents called DMPs, SCDM published the initial version of the Data Management Plan chapter in the GCDMP in 2008.

* Scope International AG, DE

† Amgen Inc., US

‡ IQVIA, US

§ Inflamax Research Inc., CA

‖ Icon Plc., UK

¶ University of Texas Health Sciences Center, San Antonio, TX, US

Corresponding author: Evaldas Lebedys ()

In the past two decades, pressure to share research data and results has increased, as has focus on research reproducibility and replication. Accordingly, public and private research sponsoring of, and regulatory interest in, data management planning and associated documentation has increased. Correspondingly, review and synthesis of DMP requirements in various disciplines and sectors have appeared in the literature.[10,11,12,13] Most notably, Brand et al.,[12] reviewed DMPs from several therapeutic development companies and academic research institutions and published a guideline for writing an SOP for DMP creation and maintenance. Williams, et al.[13], on the other hand, broadly reviewed research sponsor DMP requirements with an emphasis on data definition, collection, processing and traceability. Today, most industry sponsors, and many federal and foundation funders, of clinical studies require DMPs in some form. DMPs and an associated data management report are required by the Chinese FDA.[14]

The DMP itself is not always a required document. However, the DMP, as defined by the SCDM, contains information that "individually and collectively permit evaluation of the conduct of a trial and the quality of the data produced" (ICH E6 R2 section 8.0).[15] Documents containing this information are considered essential documents (ICH E6 R2 section 8.0).[15] As such, in audits and inspections the compliance of the described procedures and the degree to which they were followed are commonly assessed.

3) Scope

ICH E6 R2 states that trial sponsors should "implement a system to manage quality throughout all stages of the trial process" and goes on to specify that quality management includes tools and procedures for data collection and processing (ICH E6 R2, section 5.0).[15] The documentation describing those tools and procedures are a main component of a DMP. This chapter presents the DMP as an approach to organizing and maintaining comprehensive data management documentation for a clinical study. Such documentation generally specifies the following: data definition and formatting; how data are collected, processed, and stored; computer systems used to collect, process, and store data; defines technical and procedural controls through which data integrity and traceability are realized; and provides documentation supporting further data use. The DMP itself may contain the aforementioned elements or serve as a central point of reference to relevant documentation.

Initiated during study planning, the DMP spans the data lifecycle from collection to archival or disposal. The documentation comprising the DMP is frequently updated during a study, and serves as documentation for the data upon completion of the study. As such, the chapter defines the contents of a DMP as well as maintenance of DMP contents as controlled documents.

The scope of this chapter does not include foundational knowledge or skills for making data management design decisions or documenting them in the DMP for a study. Operations and information engineering and design

skills, such as ensuring that all aspects of the trial are operationally feasible; avoiding unnecessary complexity, procedures, and data collection; and ensuring human subject protection and the capability of data to support trial results are the culmination of the highest level of data management practice. These processes require deep knowledge and command of the underlying theories, principles, concepts, and methods from the multiple disciplines informing clinical data management practice. Data management skills and experience may be obtained through formal education, experience in the practice of clinical data management, relevant professional development, or some combination thereof.

This chapter outlines topics currently considered necessary for a DMP or equivalent documentation. Each DMP topic is named, defined and labeled with the level of evidence supporting its necessity. Further, each topic states the data manager's level of professional responsibility. For example, is the Data Manager responsible for designing and implementing the technical and procedural controls described in the topic? Is the task accomplished collaboratively with other clinical study functions, or is it the data manager's role to ensure that procedures created by others are in place and escalate gaps? In addition, each section describes in what stage of the study the documentation should be available and the clinical study functions to which the documentation should be available.

It is difficult and often not feasible to consolidate all data documentation for a study into one document. Thus, as the comprehensive documentation of data collection and management for a study, the DMP often: (1) references, and may briefly summarize, higher level procedures such as organizational SOPs that apply to all studies; (2) may contain or reference study-specific procedures for data collection and handling; or (3) may contain or reference procedures describing how objective evidence of data collection and processing is generated and maintained. The latter indicates the degree to which data collection and handling procedures were followed during a study. The recommendations in the DMP Contents section of this chapter explicitly state which of the following is recommended.

Where **organizational procedures governing processes across all projects <u>do NOT</u> exist,** all procedures in the organization are study-specific. In this case, study-specific procedures including the DMP bear the burden of documenting all data collection and handling procedures for a study, including specification of the objective evidence documenting to what degree those procedures were followed during a study (**Figure 1**). The scenario where organizational procedures governing data collection and handling processes across all projects **<u>DO NOT</u>** exist is expected to occur in organizations that do not conduct many studies or at organizations at low Capability Maturity Model Integration (CMMI™) levels with respect to data collection and handling for clinical studies.

On the other hand, where **organizational procedures exist and govern processes across all projects**, two scenarios exist. In **scenario one**, organizational
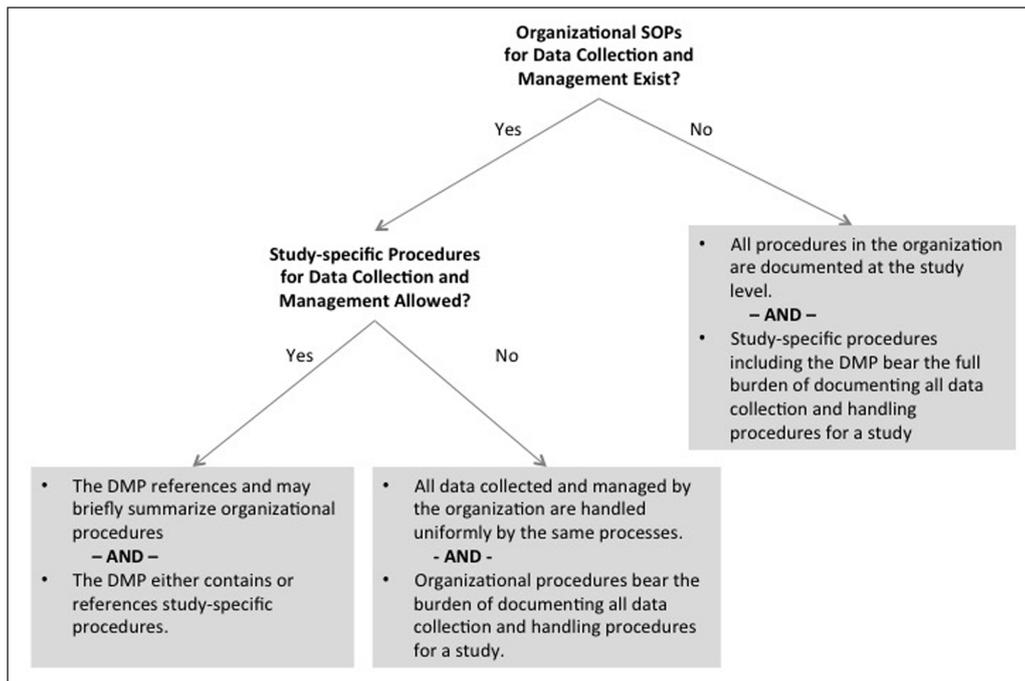
Figure 1: Documentation Scenarios for Data Collection and Management.

procedures exist and **DO NOT** allow for study-specific modifications (**Figure 1**). In this case, all data collected and managed by the organization is handled uniformly by the same processes. Organizational procedures bear the burden of documenting all data collection and handling procedures for a study including specification of the objective evidence documenting to what degree those procedures were followed during a study. Also in this case, study-specific procedures are not allowed or may require documented deviations from organizational procedures. In **scenario two**, organizational procedures also exist and **DO** allow for study-specific modification (**Figure 1**). In this situation, the DMP references and may briefly summarize in a few sentences organizational procedures **AND** the DMP either contains or references study-specific procedures including specification of the objective evidence documenting to what degree those procedures were followed during a study.

The documentation enumerated in this chapter as contained in, or referenced by, the DMP are those considered required by regulation, by guidance, or by the chapter writing group to ensure that data are capable of supporting study conclusions and that their documentation will support reconstruction of data handling (ICH E6 R2 8.0).[15] Consolidation of such documentation in, or as, a DMP is one of multiple possible approaches to meet these requirements. The documentation should be considered required; use of the DMP as an approach to such, except in countries such as China where explicitly required, is merely a recommendation. The intent of ICH E6 can certainly be met with other approaches.

4) Minimum Standards

In regions where required by regulation such as China, existence of a Data Management Plan is a minimum standard.[14] SCDM since 2008 and the DIA Clinical Data Management Community (CDMC) since 2015 have advocated compilation of (such) documentation in a DMP. While the intent and required documentation in a DMP are similar, many country regulations and regulatory guidances require the component documentation without specifying that they be compiled in a DMP. In such regions, the required component documentation are the minimum standards. In addition to local requirements, documentation should be compiled that consider the requirements of the regulatory bodies of the regions where study results will be submitted.

The International Council for Harmonisation (ICH) E6 addendum contains several passages particularly relevant to the documentation of data collection and management.[15]

Section 2.8 states that "Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective tasks."

Section 2.10 states that "All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation, and verification."

Section 4.9.0 states that "The investigator/institution should maintain adequate and accurate source documents and trial records that include all pertinent observations on each of the site's trial subjects. Source data should be attributable, legible, contemporaneous, original, accurate, and complete. Changes to source data should be traceable, should not obscure the original entry, and should be explained if necessary (e.g., via an audit trail)."

Section 4.9.2 states that "Data reported on the CRF, that are derived from source documents, should be consistent with the source documents or the discrepancies should be explained."

Section 5.0 in the following passage recommends use of quality management systems and advocates risk management.

"The sponsor should implement a system to manage quality throughout all stages of the trial process.

Sponsors should focus on trial activities essential to ensuring human subject protection and the reliability of trial results. Quality management includes the design of efficient clinical trial protocols, tools, and procedures for data collection and processing, as well as the collection of information that is essential to decision making.

The methods used to assure and control the quality of the trial should be proportionate to the risks inherent in the trial and the importance of the information collected. The sponsor should ensure that all aspects of the trial are operationally feasible and should avoid unnecessary complexity, procedures, and data collection. Protocols, case report forms (CRFs), and other operational documents should be clear, concise, and consistent.

The quality management system should use a risk-based approach."

Section 5.0.1 further advocates a process-oriented quality management system approach stating that, "During protocol development the Sponsor should identify processes and data that are critical to ensure human subject protection and the reliability of trial results."

Section 5.1.1 further states that "The sponsor is responsible for implementing and maintaining quality assurance and quality control systems with written SOPs to ensure that trials are conducted and data are generated, documented (recorded), and reported in compliance with the protocol, GCP, and the applicable regulatory requirement(s)."

Section 5.1.2 protects access to source data and documents; "The sponsor is responsible for securing agreement from all involved parties to ensure direct access (see section 1.21) to all trial-related sites, source data/documents, and reports for the purpose of monitoring and auditing by the sponsor, and inspection by domestic and foreign regulatory authorities."

Section 5.1.3 states that "Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly."

Section 5.5.1 refers to qualifications of study personnel and states that, "The sponsor should utilize appropriately qualified individuals to supervise the overall conduct of the trial, to handle the data, to verify the data to conduct the statistical analyses, and to prepare the trial reports."

Section 5.5.3 concerns validation of computerized systems and states that "When using electronic trial data handling and/or remote electronic trial data systems, the sponsor should, a) Ensure and document that the electronic data processing system(s) conforms to the sponsor's established requirements for completeness, accuracy, reliability, and consistent intended performance (i.e., validation)."

Section 5.5.3 states that validation of computer systems should be risk-based. "The sponsor should base their approach to validation of such systems on a risk assessment that takes into consideration the intended use of the system and the potential of the system to affect human subject protection and reliability of trial results." b) "Maintains SOPs for using these systems."

Section 5.5.3 The addendum introductory statement enumerates topics that should be covered in SOPs. "The SOPs should cover system setup, installation, and use. The SOPs should describe system validation and functionality testing, data collection and handling, system maintenance, system security measures, change control, data backup, recovery, contingency planning, and decommissioning."

Section 5.5.4 concerns traceability and states that "If data are transformed during processing, it should always be possible to compare the original data and observations with the processed data."

Section 8.0 states that documents that "individually and collectively permit evaluation of the conduct of a trial and the quality of the data produced" are considered essential documents (ICH E6) and shall be maintained as controlled documents.

Title 21 CFR Part 11 also states regulatory requirements for traceability, training and qualification of personnel, and validation of computer systems used in clinical trials.[16] Requirements in 21 CFR Part 11 Subpart B are stated as controls for closed systems (§ 11.30), controls for open systems (§ 11.30), Signature manifestations (§ 11.50), Signature/record linking (§ 11.70). Requirements for electronic signatures are stated in in 21 CFR Part 11 Subpart C.[16]

Medicines and Healthcare products Regulatory Agency (MHRA) 'GXP' Data Integrity Guidance and Definitions guidance provides considerations and regulatory interpretation of requirements for data integrity,[17] such as:

Section 3.4. "Organisations are expected to implement, design and operate a documented system that provides an acceptable state of control based on the data integrity risk with supporting rationale. An example of a suitable approach is to perform a data integrity risk assessment (DIRA) where the processes that produce data or where data is obtained are mapped out and each of the formats and their controls are identified and the data criticality and inherent risks documented."

Section 5.1 "Systems and processes should be designed in a way that facilitates compliance with the principles of data integrity."

Section 6.4 "Data integrity is the degree to which data are complete, consistent, accurate, trustworthy, reliable and that these characteristics of the data are maintained throughout the data life cycle. The data should be collected and maintained in a secure manner, so that they are attributable, legible, contemporaneously recorded, original (or a true copy) and accurate"

Section 6.9 "There should be adequate traceability of any user-defined parameters used within data processing activities to the raw data, including attribution to who performed the activity."

The General Principles of Software Validation; Final Guidance for Industry and FDA Staff (2002) points out a few relevant guidelines regarding proper documentation expected of software utilized in a clinical trial,[18] such as:

Section 2.4 "All production and/or quality system software, even if purchased off-the-shelf, should have documented requirements that fully define its intended use, and information against which testing results and other evidence can be compared, to show that the software is validated for its intended use."

Section 4.7 (Software Validation After a Change), "Whenever software is changed, a validation analysis should be conducted not just for validation of the individual change, but also to determine the extent and impact of that change on the entire software system"

Section 5.2.2 "Software requirement specifications should identify clearly the potential hazards that can result from a software failure in the system as well as any safety requirements to be implemented in software."

Good Manufacturing Practice Medicinal Products for Human and Veterinary Use (Volume 4, Annex 11): Computerised Systems[19] provides the following guidelines when using computerized systems in clinical trials:

Section 1.0 "Risk management should be applied throughout the lifecycle of the computerised system taking into account patient safety, data integrity and product quality. As part of a risk management system, decisions on the extent of validation and data integrity controls should be based on a justified and documented risk assessment of the computerised system."

Section 4.2 "Validation documentation should include change control records (if applicable) and reports on any deviations observed during the validation process."

Section 4.5 "The regulated user should take all reasonable steps, to ensure that the system has been developed in accordance with an appropriate quality management system."

Section 7.1 "Data should be secured by both physical and electronic means against damage. Stored data should be checked for accessibility, readability and accuracy. Access to data should be ensured throughout the retention period."

Section 7.2 "Regular back-ups of all relevant data should be done. Integrity and accuracy of backup data and the ability to restore the data should be checked during validation and monitored periodically."

Section 9.0 "Consideration should be given, based on a risk assessment, to building into the system the creation of a record of all GMP-relevant changes and deletions (a system generated "audit trail"). For change or deletion of GMP-relevant data the reason should be documented. Audit trails need to be available and convertible to a generally intelligible form and regularly reviewed."

Section 10.0 "Any changes to a computerised system including system configurations should only be made in a controlled manner in accordance with a defined procedure."

GAMP 5: A Risk-based Approach to Compliant GxP Computerized Systems[20] suggests scaling activities related to computerized systems with a focus on patient safety, product quality, and data integrity. It provides the following guidelines relevant to GxP regulated computerized systems including systems used to collect and process clinical trial data:

Section 2.1.1 states that "Efforts to ensure fitness for intended use should focus on those aspects that are critical to patient safety, product quality, and data integrity. These critical aspects should be identified, specified, and verified."

Section 4.2 states, "The rigor of traceability activities and the extent of documentation should be based on risk, complexity, and novelty, for example a non-configured product may require traceability only between requirements and testing."

Section 4.2 states, "The documentation or process used to achieve traceability should be documented and approved during the planning stage, and should be an integrated part of the complete life cycle."

Section 4.3.4.1 states, "Change management is a critical activity that is fundamental to maintaining the compliant status of systems and processes. All changes that are proposed during the operational phase of a computerized system, whether related to software (including middleware), hardware, infrastructure, or use of the system, should be subject to a formal change control process (see Appendix 07 for guidance on replacements). This process should ensure that proposed changes are appropriately reviewed to assess impact and risk of implementing the change. The process should ensure that changes are suitably evaluated, authorized, documented, tested, and approved before implementation, and subsequently closed."

Section 4.3.6.1 states, "Processes and procedures should be established to ensure that backup copies of software, records, and data are made, maintained, and retained for a defined period within safe and secure areas."

Section 4.3.6.2 states, "Critical business processes and systems supporting these processes should be identified and the risks to each assessed. Plans should be established and exercised to ensure the timely and effective resumption of these critical business processes and systems."

Section 5.3.1.1 states, "The initial risk assessment should include a decision on whether the system is GxP regulated (i.e., a GxP assessment). If so, the specific regulations should be listed, and to which parts of the system they are applicable. For similar systems, and to avoid unnecessary work, it may be appropriate to base the GxP assessment on the results of a previous assessment, provided the regulated company has an appropriate established procedure."

Section 5.3.1.2 states, "The initial risk assessment should determine the overall impact that the computerized system may have on patient safety, product quality, and data integrity due to its role within the business processes. This should take into account both the complexity of the process, and the complexity, novelty, and use of the system."

The FDA guidance, Use of Electronic Health Record Data in Clinical Investigations, emphasizes that data sources should be documented and that source data and documents be retained in compliance with 21 CFR 312.62(c) and 812.140(d).[21]

Section V.A states that "Sponsors should include in their data management plan a list of EHR systems used by each clinical investigation site in the clinical investigation" and that, "Sponsors should document the manufacturer, model number, and version number of the EHR system and whether the EHR system is certified by ONC".

Section V.I states that "Clinical investigators must retain all paper and electronic source documents (e.g., originals or certified copies) and records as required to be maintained in compliance with 21 CFR 312.62(c) and 812.140(d)".

Similarly, the FDA's guidance on electronic source data used in clinical investigations recommends that all data sources at each site be identified.[22]

Section III.A states that "A list of all authorized data originators (i.e., persons, systems, devices, and instruments) should be developed and maintained by the sponsor and made available at each clinical site. In the case of electronic, patient-reported outcome measures, the subject (e.g., unique subject identifier) should be listed as the originator."

As such, we state minimum standards for the creation, maintenance, and implementation of Data Management Plans in **Table 1**.

5) Best Practices

Best practices were identified by both the review and the writing group and are presented in **Table 2**. Best practices do not have a strong requirement based in regulation or recommended approach based in guidance, but do have supporting evidence either from the literature or consensus of the writing group. As such best practices, like all assertions in GCDMP chapters, have a literature citation where available and are always tagged with a roman numeral indicating the strength of evidence supporting the recommendation. GCDMP Levels of Evidence are outlined in **Table 3**.

6) Purpose of the DMP

The DMP serves multiple purposes. First and foremost, the DMP comprehensively documents the collection and handling of the data such that every operation performed on data from the time it is collected until finalized as

Table 1: Minimum Standards.

1	The DMP or equivalent documentation should identify all data sources for a clinical study. [I]
2	The DMP or equivalent documentation should identify risks to data integrity and evidence their evaluation, control, communication, review, and reporting. [I]
3	The DMP or equivalent documentation should outline procedures for collection, handling, and quality management of critical data, including use of computerized systems; these procedures shall exist prior to enrollment of the first subject and throughout the clinical study. [I]
4	Computer systems, software, and processes used in collection, handling, or storage of data shall maintain traceability of study data. [I]
5	The DMP or equivalent documentation shall list computerized systems or software used in the clinical study and reference or contain the following: 1) validation plans with results showing the intended use of the system or software 2) risk management plans identifying potential hazards, their impact and how risks will be managed 3) the change control process and 4) and procedures for data security, access and backup. [I]
6	The DMP or equivalent documentation shall outline responsibilities of study personnel and evidence their qualification to perform those duties through education, training, and experience. [I]
7	The DMP or equivalent documentation should be maintained as a controlled document/s. [I]

Table 2: Best Practices.

1	The DMP should support organizational compliance with applicable regulations and oversight agencies [V][12]
2	The DMP should specify all operations performed on data [V][12]
3	The DMP should be developed in collaboration with involvement from clinical and statistical operations, project management, and scientific study leadership [V] [23]; The procedure for creating a DMP should be documented in a company or institution's SOP [V][12]
4	An organizational DMP template should be used to ensure consistency and standardization across all projects [V][12]
5	The topics listed in this chapter should be documented in, or referenced by, the DMP [V]
6	As essential documents, documentation of data collection, processing, and management should be managed as controlled documents [I]
7	As a reference or job aid for study personnel, the DMP should be concise and written in plain language [VI]
8	Ensure that an approved version of the DMP is completed prior to starting on the work it describes [V][23]
9	The DMP should be reviewed at least annually to ensure that it remains current throughout the study [VI]

Table 3: GCDMP Evidence grading criteria.

Evidence Level	Criteria
I	Large controlled experiments, meta, or pooled analysis of controlled experiments, regulation or regulatory guidance
II	Small controlled experiments with unclear results
III	Reviews or synthesis of the empirical literature
IV	Observational studies with a comparison group
V	Observational studies including demonstration projects and case studies with no control
VI	Consensus of the writing group including GCDMP Executive Committee and public comment process
VII	Opinion papers

part of a dataset for analysis is attributable and can be reconstructed.[12,13] The documentation in conjunction with the SAP enables others to follow a data point in a clinical study report all the way back to the source. As comprehensive documentation of data collection and processing, the DMP may be audited to determine regulatory compliance and adherence to the documented processes. Such audits or inspections usually include review of SOPs, training records, and the DMP[24] as well as review of actual data.

Procedures for data collection and processing also serve as a reference and job aid for personnel performing data collection and management tasks to promote consistency of the work. As a reference or job aid, the DMP should be clear, concise and consistent. Further, creation, review, and approval of the DMP can help build consensus on processes as well as serve as a reference for those depending on, but not directly performing, data collection and management tasks.

Organizational procedures for creation and maintenance of the DMP help ensure that the organizational interpretation of regulation is translated into processes and daily work practices. As such, the DMP is an instrument to ensure compliance with regulations and sponsor requirements. Similarly, the DMP supplements organizational SOPs with documentation of study-specific processes.

Processes describe the method by which data are collected and managed; procedures comprising the DMP are the mechanism by which data quality is achieved. Through stated roles and responsibilities for data management tasks, the DMP establishes accountability and attribution for actions taken on the data. It is therefore critical that the DMP is agreed upon by appropriate study team members on data management processes and procedures to be followed from study initiation until the end of the study conduct.

7) Creation and Maintenance

The DMP is the documented result of decisions about data definition, collection, processing, and storage that have been made as the study workflow and data flow were designed. This section assumes that these operational design decisions have been made. Likewise, because availability of resources, (e.g., skills, technology, timeline and budget constraints, and design decisions) have impact on a number of project-related components, they must be fairly well characterized early in the design process.

a) Information needed before a DMP can be drafted
Knowledge of applicable regulations and sponsor requirements must exist to support creation of the DMP and such requirements should be referenced in the DMP. Because the DMP must address data required by the study protocol and be consistent with the overall operational plan, the data-related portions of these important inputs should be fairly mature prior to starting a DMP and complete prior to finalizing the initial version of the DMP. For work done under contract, the DMP should be consistent with the scope of work. All of these are inputs to the creation of the DMP.

The DMP is often written over time as design decisions are made and as other inputs become available or more mature. Organizations differ with respect to degree of finalization of DMP inputs prior to initiating contents of the DMP. Common information used as input to the DMP are listed in **Table 4**. When to initiate a DMP is a risk-based decision. Starting before crucial inputs are stable risks more re-work for the potential benefit of an earlier completion date. However, the DMP should be drafted during the planning phase of a study and approved prior to commencement of the work described.

b) Approach for format and structure of the DMP
DMPs have been implemented as single comprehensive documents as well as pointers to documents. The DMP may contain or reference the procedures for data collection and handing. If the latter approach is followed, procedures should be in place to ensure that the correct version or links are referenced. For example, when risk assessment is documented in other study documentation, respective reference should be included in the DMP. Documentation maintained elsewhere such as a study protocol or Trial Master File (TMF) should not be redundantly described in the DMP. A DMP that relies on and references information documented elsewhere will still be referred to here as the DMP.

The DMP should be written in a manner that can be effectively used as a job aid and work reference. Thus,

Table 4: Information Used as Input to a DMP.

Regulation, regulatory guidance and other sponsor requirements.
Database standards being considered for use including database structure, formats, or code lists, should be available.*
The final, or close-to-final, study protocol including, a study schedule of events, should be available.*
Access to, and input into, the study timeline and budget.
For data management services done under contract, a scope of work.
External data sources such as central and core labs, ePRO data, pharmacokinetic data, and data from devices or samples to be tracked and banked should be available.*
Data processing and enhancement needs such as follow-up on safety events, medical coding, or review of protocol deviations, should be decided.*
Work and data flow definition and design decisions have largely been made, or are described, in organizational SOPs.
The SAP should be drafted such that interim analyses and endpoints are specified.*

*Adapted from Brand et al.[12]

Table 5: Elements of Document Control Applicable to Data Management Plans.

· Identification and description of the DMP: title; date; version; author; study; unique document identifier;
· Format: template for the DMP; standard sections;
· Review and approval criteria and processes;
· Protection from unintended alterations;
· Distribution, access, retrieval and use;
· Storage and preservation, including preservation of human readability;
· Change or version control;
· DMP retention and disposition.

Adapted from ISO 9001.[25]

process documentation should be concise and written in clear language that is easily understood and followed.[15]

c) The DMP as a controlled document

Documentation of data collection, processing, and management are essential documents[15] and as such should be subject to document control. With respect to data management, the purpose of document control is to ensure that everyone uses the current version of procedures and that procedures used to collect and process a data value can be identified.

Components of document control (**Table 5**) include: designation of responsibility for DMP related activities; review and approval; change control; and communication of updates.[25] Because the DMP is used as a reference and job aid throughout the project, it should be up-to-date and available to study team members as needed.[25] Document control practices should adequately protect the organization from loss of confidentiality, consequences of improper use, or loss of documentation integrity through the following: constraining DMP distribution, access, retrieval and use; protection from unintended alterations; storage and preservation including preservation of human readability; change control/version control; and DMP retention and disposition.[25] Inclusion of a revision history in the DMP will aid in the recreation of changes to procedures if necessary. DMPs, or DMP components of external origin,

should be controlled in a manner equivalent to those produced internally. The DMP should be archived at the conclusion of the trial, along with the data and other study documentation.[15]

The DMP describes study data and data-related procedures that may change during the study. For example, protocol amendments may require collection of new data or changes to data collection procedures. Changes by external data providers, such as central labs, may alter the frequency or format of incoming data; thus, the DMP and associated organizational and study-specific procedures should be considered living documentation throughout the active phase of a study, should adhere to aforementioned document control practices, and should capture changes to data and data-related procedures that occur during the study.

Individuals with expertise in statistics, clinical operations, project management, or data management, commonly have review responsibility or approval authority for DMPs. The actual roles will differ from organization to organization. The work detailed in the DMP should not begin until required approvals have been obtained.

8) DMP Contents
a) Documentation of Approval
As a controlled document, all versions of the DMP or components thereof should carry an indication of approval on each version.

b) Definitions & Acronyms
Definitions of words not common in everyday use or not familiar to individuals in other disciplines, such as clinical operations or statistics, should be listed and defined in a Definitions and Acronyms section of the DMP or in component documentation.

c) Protocol Summary
A reference to where the current version of the protocol can be accessed or a brief description of the protocol should be included in the DMP. If included, a brief description might contain a statement of the therapeutic area, disease under study, type of study, the primary and secondary objectives of the study, as well as safety parameters. The DMP should reflect the current version of the protocol.

d) Scope of data management operations covered by the DMP
This section of the DMP should state which study stages, data, and activities are covered by the DMP. For example, in the case of multiple organizations, each organization may have a DMP covering the operations performed within that organization.[24] The scope section should reflect the nature and intent of the DMP as comprehensively documenting all operations performed on data.

e) Data sources
The DMP should address all data sources for a study.[2] Both data flow, the path through which the data travel, and workflow, the tasks performed by humans and machines, should be graphically represented in the DMP or component documentation. Use of standard symbols and conventions help ensure that diagrams are well-formed and professional. Using two separate diagrams ensures that both the data and workflow are completely represented.[26] Data flow and workflow diagrams should cover data sources to the final database.

f) Personnel
The personnel section of the DMP should list, or reference a list of, project personnel, duration of their association with the project, their roles, responsibilities, training and other qualifications as required in ICH E6 R2, sections 4.1.5, 5.5.1, and 5.5.3 e.[15]

g) Risk Identification and Management
Whether a risk-based approach is used or not, the DMP should include the identification of processes and data that are critical to ensure human subject protection and the reliability of trial results (ICH E6 R2 section 5.0.1).[15] Individual Project Risk is defined as "an uncertain event or condition that, if it occurs, has a positive or negative effect on one or more project objectives" such as scope, schedule, cost, or quality.[27] ICH E6 R2 section 5.0.2 states that the sponsor, "should identify risks to critical trial processes and data" at "both the system level (e.g., standard operating procedures, computerized, systems, and personnel) and clinical trial level (e.g., trial design, data collection, and informed consent process).[15]" Section 5.0.3 states that evaluation of risks should take into account the likelihood of hazards, their potential impact, and their detectability.[15] The risk identification and management section is included in the DMP because data managers are unique to each project team for their knowledge and experience to support identification and management of risk to data quality and integrity.

Figure 2 shows the high-level movement of data for the study. The central process in Figure 2 is the clinical
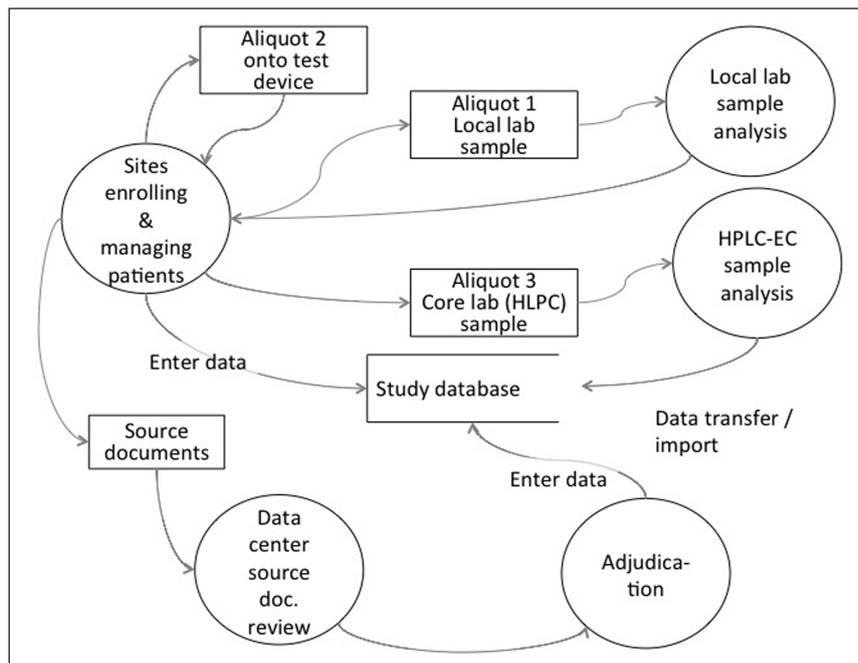
Figure 2: Data Flow Diagram (DFD) and example.[28] Circles represent processes. Rectangles represent entities. Open rectangles represent data stores. Curved arrows represent movement of data.

investigational sites enrolling and managing patients in the conduct of the study. This process generates source documents, blood samples for three separate analyses, and ultimately data. The study is testing a new bedside lab test on a cartridge. Aliquot 1 of the blood draw is sent to the local lab for immediate analysis. Aliquot 2 is pipetted directy onto the tested device, the result of which is available to the site and entered on the eCRF. Aliquot 3 is sent to a High Pressure Liquid Chromatography (HPLC) core lab for independent anlysis, the result of which is data that are imported directy into the study database. The site enters data on the eCRF, including the results from aliquot 2. The site also generates source documents that are sent to the data center for central review and independent adjudication. The data from adjudication are entered into the study database.

h) Project management

The project management section of the DMP should enumerate data-related deliverables, milestones, timelines, tasks required to meet the timelines, the responsible individuals, and other resources required. A communication and escalation plan are commonly included, as is a description of how progress and quality will be measured, tracked, and controlled. These topics are covered in the Project Management Chapter of the GCDMP.

Figure 3 illustrates a high-level workflow with respect to data collection and management for the same study as in the previous figure – testing a new cartridge-based bedside lab test. After a patient is enrolled, usually on

admission to the Emergency Room or hospital, the baseline visit (day 0) occurs at which time blood is drawn. The blood draw is split into three aliquots: the sample from Aliquot 1 is sent to the local lab for analysis, after which the site enters the data into the eCRF; the sample from Aliquot 2 is pipetted directly onto the test device which is read by the site, with the resulting data being entered by the site onto the CRF; the sample from Aliquot 3 is sent to the HPLC core lab for processing after which the data are sent electronically and imported into the Electronic Data Capture (EDC) system. The hospitalization continues and standard care is rendered. For the condition under study, hospitalization commonly lasts from 0 to 7 days. The study outcome is assessed at day 21 after which source documents required by the protocol are sent to the data center for adjudication; the adjudication data are entered by independent adjudicators directy into the EDC system. Once entered into the EDC system, data are checked for discrepancies. Discrepant values are resolved with the source of the data (either the site or the HPLC core lab). After all discrepancies are resolved, data for the patient are locked.

i) CRF

The DMP should describe or reference the CRF and form completion guideline development process (ref CRF and form guideline chapters). All study forms and form completion guidelines should be included in this section. Forms and form completion guidelines may be included as mock-ups, screen shots, data definition spreadsheets used to generate study forms, or pointers to such information.
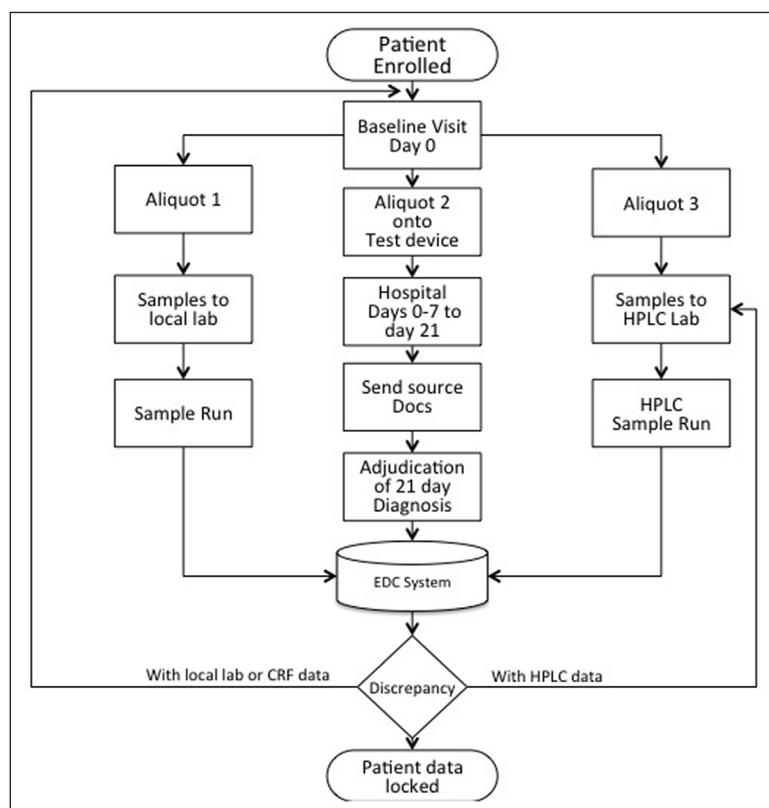
Figure 3: Work Flow Diagram (WFD) and example.[29] Rounded boxes represent the start or terminal points of a process. Rectangles represent process steps, tasks done by humans or machines. Cylinders represent databases. Diamonds represent decisions, alternate paths in the process. Arrows represent direction or sequence of process steps.

Personnel working on the study at clinical investigational sites and at the coordinating center need access to this information in the course of their daily work; thus it should be readily available. The CRFs, form completion guidelines and data dictionaries should be managed as controlled documents, per ICH E6 R2 section 8.0.[15] Post go-live changes are common and there may be a need to provide and enforce current version use as well as view-only access to previous versions.[24] In the planning stage of a study, the forms and form completion guidelines are used as a reference while other operational parts of the study are designed. During the active data collection and management phase, the forms and form completion guidelines serve as a reference for how data are collected and recorded for sites, monitors, data managers, statisticians, and others. After completion of the study, the forms and form completion guidelines document data collection and recording and provide a record of all changes to data collection and recording procedures.

j) Data definition

The DMP should document or reference complete data definitions. The data management case study from the International World Health Organization Antenatal Care Trial describes their pragmatic approach to data definition and harmonization of data from multiple countries:

> … all [form] questions are defined. Each question or study variable definition includes an identification label, data type, length, a range or list of acceptable values, and optionally, labels for defined values. In addition, study variables may be combined to define consistency rules. The complete set of variable definitions for a computer file is referred to as a data dictionary. Owing to the specificity of each site, data dictionaries and SOPs may differ in some minor points.[30]

While this is an older example, the components of data definition have not changed, and for a small number of studies, sophisticated systems for managing data definitions are not required. However, data definition

should be considered an essential document, as described in ICH E6 R2 section 8.0,[15] and as such, it should be documented, maintained over time, and made available to appropriate individuals.

Complete data definition should include complete specification of each data element.[26] A data element is a question/answer format pair and is the smallest meaningful unit of data collection, exchange, and use. The DMP should contain or reference the definition for each data element used in the study. Components of complete data element definition (**Table 6**) include conceptual and operational definitions, a unique identifier or name, a data type, some specification of valid values, and a mapping of each data element to where it is stored. Note that the format in which data are stored during the active phase may differ from that in which data are archived or shared in the inactive phase. Mapping to both formats is necessary but the latter may be done prior to archival rather than at study start.

Data elements developed through or referenced by the FDA's Therapeutic Area Data Standards Program should be considered when they match a study's needs. In fact, sponsors whose studies start after Dec. 17, 2016, must submit data in the data formats supported by FDA[31] and listed in the FDA Data Standards Catalog.[32] This applies to New Drug Applications (NDAs), Biological Licensing Applications (BLAs), Abbreviated New Drug Applications (ANDAs), and subsequent submissions to these types of applications. Clinical Data Interchange Standards Consortium (CDISC) Clinical Data Acquisition Standards Harmonization (CDASH), CDISC and Health Level Seven (HL7) data exchange standards and controlled terminology standards accepted or required by the FDA are listed in the catalog.

k) Data mapping

The mapping of each data element to the database or other structure in which the data are stored or made available for use should be provided. In clinical research, this has often been accomplished using an annotated CRF. An annotated CRF is a copy of the data collection form

Table 6: Components of Data Element Definition.

Conceptual and operational descriptions. If the operational definition does not include a prompt for questions from CRF pages or questionnaires, the prompt should be documented.

> Where standard data elements are used these definitions can be referenced rather than restated.
> Derived data elements should include in their operational definition the algorithm used for calculation or a pointer to the algorithm.

It is common to collect a few hundred or more data elements for a study. Thus, applying a unique identifier by which the data elements can be managed and organized is helpful.

Data elements should be labeled with a data type. The data type facilitates processing the data values as well as understanding additional and necessary definitional information.

Valid values for each data element should be specified.

> Discrete data elements should be accompanied by a list of the valid values with conceptual and operational definitions for each valid value. These should include References to controlled terminology standards used.
> Continuous data elements should be accompanied by their unit of measure, statement of the range or other limits on valid values, and precision.

*Adapted from The Data Book, Collection and Management of Research Data.[26]

overlaid with data element names, some components of data definition such as data type and valid values, logical storage location, and format. While there are other methods equivalent from a data perspective, the annotated CRF has the added advantage of also providing a visual representation and thus has long been used in clinical studies to document and share this mapping. Alternatively, data element definitions can be documented in a spreadsheet. Annotation in the spreadsheet format is less time consuming, while graphical annotation is easier to read for those who are not familiar with database specifications. Regardless of the form chosen for data element definition, the annotated CRF should identify the following:

- Database tables, files or datasets that will store data from a corresponding field;
- Variable or column names in which data will be will be stored;
- Metadata such as visit labels, e.g., Baseline, Day 30, Visit 2, or CRF section titles;
- If applicable, additional variables used to store pre-printed or implicit information (such as subject identification data, CRF page number, trial period, line or sequence numbers in lists, etc.).

Upon availability of the final clinical trial protocol version, final versions of CRF, subject diaries, questionnaires and other data collection forms, the clinical data manager, in collaboration with the statistician, prepares the annotated CRF or equivalent mapping documentation. This mapping serves as a resource to all study team members and external parties using or inspecting study data.

l) Data transformation
Organizations tend to transform study data to facilitate data processing and analysis at different stages of the study. Transformation of data for example to the CDISC Submission Data Tabulation Model (SDTM) and to Analysis Data Model (ADaM) based datasets is common and encouraged by regulatory agencies.[28] When data are transformed during the active phase of the study, transformations should be documented, clearly mapping the source and the destination data elements. Annotated CRFs for datasets such as SDTM and ADaM are commonly created for this purpose. The DMP should define or reference the transformations of data and respective documentation of different data structures.

Finally, to ensure that when there are changes such as data elements added, changed, or no longer collected, the documentation of data definition should be updated and the updates should be tracked. Thus, data definition is not a one-time activity but, to support traceability, must be maintained throughout the study.

m) Traceability
Traceability permits an understanding of the relationships between the analysis results (tables, listings and figures in the study report), analysis datasets, tabulation datasets, and source data.[33] For example, ICH E6 R2 sections 4.9.0, 4.9.3, and 5.5.3 clearly state that "all changes to data

should be documented," that such documentation "allows reconstruction of the course of events" and recommends use of audit trails.[15] Regulations such as 21CFR11 specify technical controls for aspects of traceability relevant to changes to data values.[16] The DMP should describe how traceability is maintained through all operations performed on data, such as transcription, database updates, and transformation of data from raw database to SDTM datasets as well as from SDTM datasets to ADaM datasets.

n) System Access and Privileges
Because traceability requires attribution, the DMP should list or reference procedures for assignment of and tracking access to and privileges in data systems including the time period for which the access and privileges are active during the study. System access and privileges within systems used to manage clinical data are typically role-based. The DMP should define or reference documentation of the actual user access and privileges for a study. Procedures for periodic review of user access, to ensure system privileges are up to date, should also be referenced or stated directly in the DMP.

o) Data systems used and their version(s) of each system used
All systems used to process data collected in the study should be listed in or referenced by the DMP. When management or hosting of information systems used for conduct of a clinical study is the responsibility of a vendor or other external party, the DMP should describe or reference organizational procedures for assuring that vendors are capable of assuring data integrity and protecting human subjects, per 21 CFR Part 11.[16] Per 21 CFR Part 312.52, the DMP shall describe or reference description of responsibilities of the vendor with respect to collection, processing, management, and ultimately integrity of clinical data and protection of human subjects or reference equivalent documents.[16] Please see the Vendor Selection and Management chapter for more information, including recommendations, minimum standards, and best practices.

p) Software Development Lifecycle procedures
Procedures for development, selection, integration, testing, installation, and change control of trial-specific systems and related activities should be described in or referenced by the DMP. Computer systems employed for study data collection, processing, and storage are considered GxP regulated systems and as such require controlled management throughout the system lifecycle. Thus, procedures may use a risk-based approach considering system impact, complexity, and novelty.[20]

The scope of validation of all systems used in the study should be stated in the DMP. Whenever details of systems' validation are defined in other documents, references to those documents should be made. For example, organizations often define the details of validation in a Validation Plan, which becomes the master document describing management of the system lifecycle. In these cases, validation summary reports are usually produced

summarizing the course and outcome of validation. References to these documents or the SOP that governs them would be appropriate in the DMP.

q) Change Control
All post-release changes in computerized systems used in clinical studies should be implemented in a controlled manner and following the lifecycle management procedure defined in the SOPs and according to regulatory requirements. Where post-release changes in the systems are required, the party responsible for system development and maintenance should evaluate the extent of required changes and should identify the elements that will, and could possibly be, affected by the requested changes. If required changes are acceptable, the change control process defined in organizational SOPs should be triggered. The CDM should make sure that the details of system modification, including evaluation of impact upon the study, are documented in, or referenced by, the DMP.

r) System interfaces
Integrations of systems used to collect and process data should be considered and planned from the early stage of the study. The need of integration and the technique used should be documented in the DMP. Systems commonly integrated include Clinical Data Management Systems, data warehouses, electronic Patient Reported Outcomes (ePRO) or Clinical Outcome Assessment (COA) systems, external randomization systems, drug supply systems, medical devices, central or site based Clinical Trial Management Systems, central laboratory, or other central analysis systems such as central ECG reading. Such interfaces may employ varied configurations, management, and data handling strategies. Please see the Integration of External Data chapter for more information, including recommendations, minimum standards, and best practices.

s) Instrumentation, Calibration and Maintenance
Instrumentation used for data collection or processing used for a clinical study should be documented in the DMP. Procedures for selection, testing, distribution, training on operation, calibration, maintenance, and acquisition of data from instrumentation and personal or medical devices should be described in or referenced by the DMP. Data acquisition from devices should appear on study data flow and workflow diagrams. Traceability of data from devices requires that data be attributable to the actual device.

t) Privacy and Confidentiality
Protection of human subjects' right to privacy and organizational procedures for maintaining confidentiality are required by multiple regulations applicable to clinical studies including the Health Insurance Portability and Accountability Act (HIPAA) security rule 45CFR160 and 45CFR164(A) and (E), The Common Rule 45CFR46, 21CFR Part 50 (Protection of Human Subjects) and Part 56 (Institutional Review Boards). By requiring de-identification for some data uses, HIPAA requires that those handling data in clinical trials understand where the rule applies and how to apply it. For example, drafting consent language around data use, handling subject withdraws according to consent, and identifying and properly handling PHI in transferred data or on source documents. HIPAA also specifies that data access be restricted to "minimum necessary." For example, access to data should be restricted to qualified and approved personnel. Safeguards such as these are needed to ensure privacy of records, regardless of the method in which the data is gathered. The DMP will describe or reference procedures for protecting privacy of human subjects.[34,35,16]

u) System security
The DMP should describe or reference procedures for the management of logical and physical security in relation to the systems used to manage and transmit clinical data. Systems security is a large concern in terms of assuring freedom from unintended and undetected adulteration of data, in terms of ensuring that information systems are accessible and available when needed for study conduct, and in terms of protecting human subjects' right to privacy and study obligations for maintaining confidentiality. Facilities where paper and electronic data are stored shall be access controlled and secure, e.g., with access, fire and flood protection.[16,34,35] Please see the Data Storage and Integration of External Data chapters of the Good Clinical Data Management Practices for more information including recommendations, minimum standards, and best practices.

While most data managers are not trained in design and implementation of information system security, as stewards of the data they are professionally responsible for assuring that organizational procedures for information system security are in place or organizational and study leadership are made aware of a lack thereof, and that information system security is deemed appropriate for the study by the organizational authority such as an Information System Security Officer. In these matters, the data manager should assure that the study Principal Investigator (PI), who is ultimately accountable for breaches and other information system related risks to human subjects, is informed of and has approved information systems security procedures.[26,35]

v) Back-up and recovery
Backup and backup testing of GxP relevant data and regular performance is required[16] Further, as a common practice and as required of trial sponsors, organizations should maintain Business Continuity Plans and Disaster Recovery Plans to address all identified risks including disaster scenarios.[36,37,38,39] The DMP shall include a brief description of data back-up and recovery or a reference should be made to the appropriate documents if documented outside the DMP. The DMP should describe or reference business resumption and disaster plans[13]

w) Data collection and recording
ICH requirements sections 4.9.0–4.9.2 and 5.0 necessitate that the quality management system extend to data origination. Clinical Data Managers may not have clinical experience or familiarity with documentation of routine

care in medical records and clinical workflow issues that impact collection of data in clinical settings.[15] The data manager will certainly have expertise of value and can significantly contribute to specification of data collection and recording and design of quality management practices for data collection and recording. Further, the impact of data origination methods and processes on the quality of study data cannot be refuted. For these reasons, the DMP should specify or reference specifications for data origination, collection, and recording.

x) Data processing

Data collection and processing determine the quality of data upon which research conclusions are based and, as such, are a main concern of regulators, sponsors, researchers, statisticians, and data managers alike.[2] ICH R2 section 8.0 further states that the procedures for a clinical study "individually and collectively permit evaluation of the conduct of a trial and the quality of the data produced." As such, procedures for data collection and processing, as well as objective evidence of the performance, are considered essential documents and shall be maintained as controlled documents.[15] For these reasons, the DMP data processing section should document or reference documentation specifying all operations performed on data and provide for maintenance of the objective evidence of such processing. Where such operations are performed by humans, the current version of these specifications should be accessible by individuals performing them. Operations performed on clinical data include:

- Key entry and verification of machine processed entry – Optical and language recognition processes. These include all manual updates to data values. Data checking that occurs during key entry and verification of machine processing should be specified or referenced by the DMP as should guidelines for manually performed entry and verification tasks. Please see the Data entry chapter of the GCDMP for more information, including recommendations, minimum standards, and best practices.
- Data cleaning – Identification and resolution of data discrepancies. These include self-evident corrections to data values and imputations as well as other methods of data cleaning such as listing review, source document verification, computational monitoring for trends, medical review, or use of specialized data cleaning or review reports to identify data discrepancies. The DMP should state or reference study-specific edit check specifications (programmed and manual). Similarly, self-evident corrections should be specified or referenced by the DMP. Because self-evident corrections are made without the additional approval of the site's data entry staff, the sites should (1) retain a list of allowed self-evident corrections and should (2) be able to access the audit trail documenting these corrections per ICH E6 R2 section 5.0.1.[15]
- Medical coding – The process by which a verbatim term or originally recorded term entered in the eCRF is translated into a standardized medical

term using a medical dictionary.[40] Medical coding can be performed manually by a human, by an auto-encoder, or as a hybrid process. The DMP should describe or reference the coding guidelines and conventions used. Please see the Medical Coding chapter for minimum standards and best practices.
- Transformations performed on data – During the active phase of a study, transformations may be performed on data. Examples of such transformations include mapping data to different coding schemes or scales, calculation of new values based on existing data, or reformatting data. All transformations to be performed should be listed in or referenced by the DMP. Such documentation should include the algorithm for each transformation, the transformation process, and how occurrence of the transformation is documented.
- Integration of externally managed data – Data that is captured outside CRFs and sent electronically, such as central lab data, should be identified in the DMP. For each external data transfer, the DMP should specify or reference the following: specification of the sender and recipient, data transfer formats, data exchange standards, media and transfer schedule, security of transfers, data importing, data transformations performed during data extraction, and data transformations performed during loading. These detailed specifications of electronic data transfer may exist as a separate document for each data source. The DMP should specify or reference specification of reconciliation of external data with data in the clinical database, including checks for logical consistency of data and handling discrepancies identified through such reconciliation. Storage and retention of data received from external parties should be specified in or referenced by the DMP. Please see the Integration of External Data chapter of Good Clinical Data Management Practices for more information, including recommendations, minimum standards, and best practices.
- Data-assisted trial operations – Study data may be used to support study operations and doing so often requires partial automation of clinical study processes through configuring data flow and workflow in information systems, programming, or reporting. ICH R2 section 8.0 states that the procedures for a clinical study "individually and collectively permit evaluation of the conduct of a trial and the quality of the data produced."[15] As a result, data-assisted (automated or partially automated) study operations such as safety event detection and reporting, blinding and un-blinding procedures, detection and handling protocol deviations, detection and handling of cases for clinical event classification or review should be listed in the DMP. For example, in the case of identification and handling safety events, alerts may be programmed to notify study team members when adverse events or serious adverse events are reported by sites. In this case, the information system may be programmed to conditionally make an SAE form avail-

able when AEs, indicated as serious, are entered. Special functionality may be added to facilitate site or coordinating center follow-up and reporting of AEs and SAEs. All of these require programming, testing and implementation. The DMP should specify or reference specifications for data-assisted trial operations. Where pertinent, the GCDMP describes minimum standards and best practices for data assisted study operations.

Comprehensively specifying all operations performed on data is not new.[2] In fact, a case study by De Leeuw, et al. describing the evolution of the data management practices used during the conduct of the Scandinavian simvastatin survival mega-trial in 4444 patients exemplifies a similarly comprehensive plan.[23] Based on the study, they concluded the following:

1. A "comprehensive definition of handling guidelines should be prepared before the study start. Since new situations will inevitably arise, however, clear procedures should be defined to determine how guidelines will be updated and what impact new guidelines will have on data already entered and reviewed. It should also be understood that it will usually be impossible to fully define final guidelines until the conclusion of a trial";

2. Specific data review responsibilities should be defined prior to the study start and that a formal data management plan would have been helpful in this regard and should be a standard component of all clinical trials and should include clear guidelines, review tools, and operating procedures communicated to all project staff at the start of the study.[23] Brand et al. conclude that a clinical trial that makes use of multiple electronic data collection systems (web based, smartphone, interactive voice recognition system) may require more DMP components than an early-phase study with few subjects that collects data on paper case report forms.[12]

The DMP shall document or reference documented procedures for data handling processes from study initiation through archival per ICH E6 R2 sections 4.9.0, 5.1.1, 5.5.3, and 5.5.4).[15] Procedures for collecting, handling, and quality management of critical data, shall be updated to reflect study changes, as described in ICH E6 R2 sections 8.3.2 and 8.3.6.[15]

y) Data Quality Control
As an important component of the Quality Management System, the DMP, in conjunction with organizational SOPs, is an important instrument of Data Quality Assurance – it comprehensively specifies how data are collected and managed. Part of data quality assurance is one or more mechanisms for data quality control. In fact, ICH E6 R2 section 5.1.3 states that, "quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly."[15] In fact, the concept that data quality control activities should be prospectively planned, conducted during the life of a

study, and reported is not new.[2] Thus, the process, tools, and reports that will be used to monitor data quality, should be described or referenced by the DMP, including aspects of data quality to be measured, the measures used, how data quality will be measured and reported, and any acceptance criteria. The process and end-of-study reports should document the quality of data at study conclusion, as well as critical data quality issues discovered during the study and actions taken to remediate those data quality issues, and prevent and monitor for future occurrence.

z) Database Lock and Unlock
Database lock signifies that no further changes to data are expected and the corresponding removal of write access to the database. Un-locking a database is the reinstatement of write privileges for a specified and approved reason. There are instances where data update is needed after database lock. Un-locking may vary between sponsors, investigative institutions, and study types. For example, in some institutions, a database may need to be unlocked in order to upload un-blinding data; this should be mentioned in the DMP. However, unlocking a database and making changes after a study has been un-blinded will understandably be heavily scrutinized and may call the objectivity of the study into question. The DMP should describe or reference procedures for database lock and unlock including study-specific criteria for locking and unlocking the database. Please see the Database Lock chapter for more information, including recommendations, minimum standards, and best practices.

aa) Data Archival
The procedures for data archival should be described or referenced by the DMP. Such procedures include responsibilities for data archival, enumeration of the data to be archived, the data format for archival, how and when data will be transferred for archival as well as how receipt will be acknowledged, and how long data should remain in archival prior to disposal.[41] Please refer to the Clinical Data Archiving chapter for more information, including recommendations, minimum standards and best practices. The DMP should describe or reference procedures for assuring that clinical investigational sites have and retain a copy of source data, data submitted to the sponsor, and changes to such data. ICH E6 states that, the sponsor and investigator/institution should maintain a record of the location(s) of their respective essential documents including source documents.[15] The storage system used during the trial and for archiving (irrespective of the type of media used) should provide for document identification, version history, search, and retrieval.

bb) Issue or incident reports
Unexpected events occurring during the trial that may impact the overall analysis of the study that cannot be documented or recorded in the database or not covered by any operational plans should be documented. These issues impact a sponsor or regulator's ability to reconstruct a study. Thus, the DMP should provide the method of how these issues or incident reports will be handled and documented.

Table 7: DMP Table of Contents.

	Reference to Supporting SOP expected	SOP generates study-specific procedures	Objective evidence/study-specific documentation
Study context and management			
Protocol Summary	X	X	
Scope of data management operations			X
Data sources			X
Personnel	X		X
Risk Identification and Management			X
Project Management	X		
Data Definition and Processing			
CRF	X	X	
Data definition	X		X
Data collection and recording	X	X	X
Data processing	—	—	—
Instrumentation, calibration, maintenance	X	X	X
Key entry and verification of machine entry	X	X	X
Data cleaning	X	X	X
Medical coding	X	X	X
Transformations performed on data	X	X	X
Integration of external data	X	X	X
Data-assisted trial operations	X	X	X
Data quality control	X	X	X
Privacy & confidentiality	X	X	
Database Lock and Unlock	X	X	X
Data Archival	X	X	X
Issue or incident reports	X		X
Data Systems			
Data systems used and version(s) used	X		X
System access and privileges	X	X	X
SDLC procedures	X		X
System interfaces			X
System security	X		X
Back-up and recovery	X		X

9) Role of the DMP in Audits

DMP as documentation that "individually and collectively permit evaluation of the conduct of a trial and the quality of the data produced" ICH E6 R2 section 8.0 is commonly in the focus of audits and inspections. Organizations are expected to demonstrate compliance and the degree to which procedures defined or referenced in the DMP are followed (see **Table 7**). Deviations of the defined procedures may be reported as non-compliant and incur audit or inspection findings.

10) Recommended SOPs

ICH E6 R2 section 5.0.1 states that, "During protocol development, the sponsor should identify processes and data that are critical to ensure human subject protection and the reliability of trial results."[15] This implies that organizations should map out the processes involved in study design, start-up, conduct, and closeout and make explicit decisions about which are considered to impact human subject protection and the reliability of trial results. Organizational processes may be partitioned differently leading to different scope and titles for SOPs. Though organizations may differ in how the processes are covered in their SOPs, below is a list of processes commonly considered to impact human subject protection and the reliability of trial results:

- Data Management Plan Creation and Maintenance
- Document Control (ICH E6 R2 section 8.0)[15]

· System validation and functionality testing (ICH E6 R2 section 5.5.3 b)[15]
· Data collection (ICH E6 R2 section 5.0)[15]
· Data processing (ICH E6 R2 section 5.0)[15]
· System maintenance (ICH E6 R2 section 5.5.3 b)[15]
· System change control (ICH E6 R2 section 5.5.3 b)[15]
· System security measures (ICH E6 R2 section 5.5.3 b)[15]
· Data backup and recovery (ICH E6 R2 section 5.5.3 b)[15]
· Contingency planning (ICH E6 R2 section 5.5.3 b)[15]
· System decommissioning (ICH E6 R2 section 5.5.3 b)[15]

11) Literature Review details and References

This revision is based on a systematic review of the peer-reviewed literature indexed for retrieval. The goals of this literature review were to (1) identify published research results and reports of evaluation of new methods regarding Data Management Planning and to (2) identify, evaluate, and summarize evidence capable of informing the practice of data management plan creation and maintenance.

The following PubMed query was used:

"data management plan" OR "data management plans" OR "data management procedures" OR "data quality assurance") AND ("clinical trial" OR "clinical study" OR registry OR "observational study" OR "interventional" OR "phase 1" OR "phase 2" OR "phase 3" OR "phase 4" OR "phase I" OR "phase II" OR "phase III" OR "phase IV" OR "first in man" OR "clinical research" OR "device study" OR "interventional trial" OR "phase 1" OR "phase 2" OR "phase 3"

OR "phase 4" OR "phase I" OR "phase II" OR "phase III" OR "phase IV " OR RCT OR "randomized clinical trial" OR "non-interventional" OR "post authorization" OR "adaptive trials" OR "feasibility study" OR "phase 2/3" OR "phase II/III" OR "phase 2a" OR "phase 2b" OR "phase IIa" OR "phase IIb"

The search query was customized for, and executed on, the following databases: PubMed (34 results); CINAHL (7 results); EMBASE (88 results); Science Citation Index/Web of Science (27 results); PsychINFO (1 result); Association for Computing Machinery (ACM) Guide to the Computing Literature (267 results); the Institute of Electrical and Electronics Engineers (IEEE) (109 results). A total of 532 works were identified through the searches. The searches were conducted on January 13, 2017. Search results were consolidated to obtain a list of 482 distinct articles. Because this was the first review for this chapter, the searches were not restricted to any time range. Literature review and screening details are included in the PRISMA diagram for the chapter.

Two reviewers used inclusion criteria to screen all abstracts. Disagreements were adjudicated by the writing groups. Twenty articles meeting inclusion criteria were selected for review. Two individuals reviewed each of the twenty selected articles and the eight additional sources identified through the review. Each was read for mention of explicit practice recommendations or research results informing practice. Relevant findings have been included in the chapter and graded according to the GCDMP
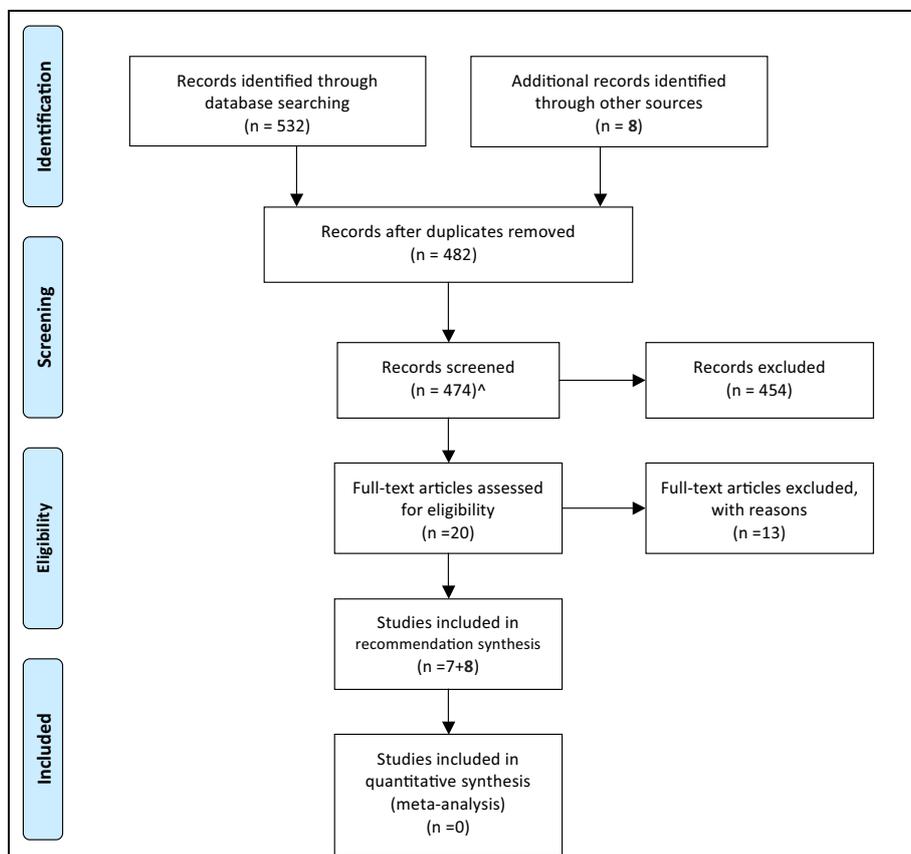
Figure 4: PRISMA** Diagram for the Data Management Plan Chapter.

evidence grading criteria in the table below. This synthesis of the literature relevant to data management planning and support transition of this chapter to an evidence-based guideline.

12) Revision History

Date	Revision description
December 2008	Initial version of the DMP chapter
December 2019	Complete revision based on systematic literature review

Competing Interests

The authors have no competing interests to declare.

References

1. **Organization.** review, and administration of cooperative studies (Greenberg Report): a report from the Heart Special Project Committee to the National Advisory Heart Council, May 1967. *Control Clin Trials.* 1988; 9(2): 137–48. DOI: https://doi.org/10.1016/0197-2456(88)90034-7

2. **Knatterud GL.** Methods of quality control and continuous audit procedures for controlled clinical trials. *Control Clin Trials.* 1981; 1(4): 327–332. DOI: https://doi.org/10.1016/0197-2456(81)90036-2

3. **Association for Clinical Data Management (ACDM).** Data Handling Protocol Working Party, Guidelines to facilitate Production of a Data Handing Protocol. 1996. Accessed March 11, 2017, Available from http://www.acdm.org.uk/assets/DHP_Guidelines.pdf

4. **Blumenstein BA, James KE, Lind BK, Mitchell HE.** Functions and organization of coordinating centers for multicenter studies. *Control Clin Trials.* 1995; 16: 1s–3s. DOI: https://doi.org/10.1016/0197-2456(95)00092-U

5. **Gassman JJ, Owen WW, Kuntz TE, Martin JP, Amoroso WP.** Data quality assurance, monitoring, and reporting. *Control Clin Trials.* 1995; 16: 1s–3s. DOI: https://doi.org/10.1016/0197-2456(94)00095-K

6. **Hosking JD, Newhouse MM, Bagniewska A, Hawkins BS.** Data collection and transcription. *Control Clin Trials.* 1995; 16: 1s–3s. DOI: https://doi.org/10.1016/0197-2456(94)00094-J

7. **McBride R, Singer SW.** Introduction to the 1995 clinical data management special issue of Controlled Clinical Trials. *Control Clin Trials.* 1995; 16: 1s–3s. DOI: https://doi.org/10.1016/0197-2456(95)90409-7

8. **McBride R, Singer SW.** Interim reports, participant closeout, and study archives. *Control Clin Trials.* 1995; 16: 1s–3s. DOI: https://doi.org/10.1016/0197-2456(94)00096-L

9. **McFadden ET, LoPresti F, Bailey LR, Clarke E, Wilkins PC.** Approaches to data management. *Control Clin Trials.* 1995; 16: 1s–3s. DOI: https://doi.org/10.1016/0197-2456(94)00093-I

10. **Jones C, Bicarregui J, Singleton P.** JISC/MRC Data Management Planning: Synthesis Report. Sciences and Facilities Technology Council (STFC). Open archive for STFC research publications. 2011. Accessed April 30, 2016. Available from http://purl.org/net/epubs/work/62544

11. **Knight G.** Funder Requirements for Data Management and Sharing. Project Report: London School of Hygiene and Tropical Medicine; 2012; London, UK. Accessed April 30, 2016. Available at http://researchonline.lshtm.ac.uk/208596/

12. **Brand S, Bartlett D, Farley M,** et al. A model data management plan standard operating procedure: results from the DIA clinical data management community, committee on clinical data management plan. *Ther Innov Regul Sci.* 2015; 49(5): 720–729. DOI: https://doi.org/10.1177/2168479015579520

13. **Williams M, Bagwell J, Nahm-Zozus M.** Data management plans: the missing perspective. *J Biomed Inform.* 2017; 71: 130–142. DOI: https://doi.org/10.1016/j.jbi.2017.05.004

14. China cFDA registration application documents for class 5.1 imported innovative drugs (2016 version) accessed March 10, 2018. Available from http://www.sfdachina.com/info/200-1.htm

15. **Food and Drug Administration.** US Department of Health and Human Services. ICH E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R1), March 2018. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e6r2-good-clinical-practice-integrated-addendum-ich-e6r1.

16. **Food and Drug Administration.** US Department of Health and Human Services. Code of Federal Regulations Title 21, available at https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm

17. **Medicines & Healthcare products Regulatory Agency (MHRA).** 'GXP' Data Integrity Guidance and Definitions. Revision 1: March 2018. Accessed June 2, 2018. Available at https://www.gov.uk/government/publications/guidance-on-gxp-data-integrity.

18. **U.S. Department of Health and Human Services Food and Drug Administration.** General Principles of Software Validation Guidance for Industry and FDA Staff, January 2002, available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-principles-software-validation.

19. **European Commission Health and Consumers Directorate-general.** EudraLex The Rules Governing Medicinal Products in the European Union Volume 4 Good Manufacturing Practice Medicinal Products for Human and Veterinary Use (2011) Chapter 4: Documentation. Commission Européenne, B-1049 Bruxelles/Europese Commissie, B-1049 Brussel – Belgium. Available at https://ec.europa.eu/health/documents/eudralex/vol-4_ga.

20. *GAMP® 5 A Risk-based Approach to Compliant GxP Computerized Systems.* North Bethesda, MD: International Society for Pharmaceutical Engineering (ISPE). 2008.

21. **Food and Drug Administration.** US Department of Health and Human Services. Guidance for industry: Use of Electronic Health Record Data in Clinical Investigations. July 2018. Accessed August 8, 2018. Available from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM501068.pdf

22. **Food and Drug Administration.** US Department of Health and Human Services. Guidance for Industry: Electronic Source Data in Clinical Investigations. September 2013. Accessed August 8, 2018. Available from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf

23. **De Leeuw N, Wright A, Cummings SW, Amos JC.** Data management techniques applied to the Scandinavian simvastatin survival (4S) mega-trial. *Drug Inf J.* 1999; 33: 225–236. DOI: https://doi.org/10.1177/009286159903300125

24. **Prokscha S.** The data management plan. *Practical guide to clinical data management.* 3rd ed. Boca Raton, FL: 2012: 3–8. DOI: https://doi.org/10.1201/b12832

25. *Quality management systems − Requirements* (ISO 9001:2015(en)) 5th ed. Geneva: International Organization for Standardization; 2015. Available from https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:en.

26. **Zozus MN.** Data management planning. *The data book: collection and management of research data.* 2017; 67–82. Boca Raton, FL: Taylor and Francis. DOI: https://doi.org/10.1201/9781315151694

27. A guide to the project management body of knowledge, 6th edition. (PMBOK guide). Newtown Square, PA: Project Management Institute, Inc.; 2017.

28. **Yourdon E.** Just Enough Structured Analysis. 2006. Available from https://docs.google.com/file/d/0B42Cu1mD9Z7seVVHLUdqb1Q1SlU/preview.

29. Information processing − Documentation symbols and conventions for data, program and system flowcharts, program network charts and system resources charts,last reviewed and confirmed in 2019. Available at https://www.iso.org/standard/11955.html

30. **Pinol A, Bergel E, Chaisiri K, Diaz E, Gandeh M.** Managing data for a randomised controlled clinical trial: experience from the WHO Antenatal Care Trial. *Paediatr Perinat Epidemiol.* 1998; 12(Suppl 2): 142–55. DOI: https://doi.org/10.1046/j.1365-3016.12.s2.2.x

31. **Food and Drug Administration.** US Department of Health and Human Services, Center for Drug Evaluation and Research (CDER), CDER Data Standards Program. Accessed March 11, 2018. Available from the U.S. FDA at https://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm249979.htm.

32. **Food and Drug Administration.** US Department of Health and Human Services, Center for Drug Evaluation and Research (CDER), CDER Data Standards Program. Accessed March 11, 2018. Available from the U.S. FDA at FDA Data Standards Catalog https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm.

33. **Food and Drug Administration.** US Department of Health and Human Services, Study Data Technical Conformance Guide, 2017 available from https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm384744.pdf.

34. **US Department of Health and Human Services.** Code of Federal Regulations, Title 45 CFR Health Insurance Portability and Accountability Act (HIPAA) security rule: Part 160: General Administrative Requirements, 2005, available at https://www.govinfo.gov/content/pkg/CFR-2005-title45-vol1/pdf/CFR-2005-title45-vol1-part160.pdf, Part 162: Administrative Requirements, 2002, available at https://www.govinfo.gov/content/pkg/CFR-2016-title45-vol1/xml/CFR-2016-title45-vol1-part162.xml, and Part 164: Security and Privacy, 2004, available at https://www.govinfo.gov/content/pkg/CFR-2004-title45-vol1/pdf/CFR-2004-title45-vol1-part164.pdf.

35. U.S. Department of Health and Human Services Title 45 Public Welfare (the Common Rule) CFR Part 46, 2009. available at https://www.hhs.gov/ohrp/sites/default/files/ohrp/humansubjects/regbook2013.pdf

36. **Calvert WS, Ma JM.** *Introduction to research data management. Concepts and case studies in data management.* Cary, NC; 1996.

37. **Calvert WS, Ma JM.** *The importance of planning. Concepts and case studies in data management.* Cary, NC; 1996.

38. **Calvert WS, Ma JM.** *Establishing the RDM system. Concepts and case studies in data management.* Cary, NC: 1996.

39. **Calvert WS, Ma JM.** *Basic RDM system management. Concepts and case studies in data management.* Cary, NC: 1996.

40. **Shen T, Xu LD, Fu HJ,** et al. Infrastructure and contents of clinical data management plan. *Yao Xue Xue Bao.* 2015; 50(11): 1388–92.

41. **Stiles T, Lawrence J, Gow N, Rammell E, Johnston G, Joyce R.** *A Guide to Archiving Electronic Records.* Shenfield, Essex UK: Scientific Archivists Group Ltd.; 2014. Accessed April 12, 2018. Available from https://the-hsraa.org/wp-content/uploads/2017/12/AGuidetoArchivingElectronicRecordsv1.pdf