Clinical Data Archiving

June 2008

Abstract

In order to meet the requirements of industry guidelines and regulations, clinical data managers must ensure that data captured during a clinical trial are retained correctly. This chapter provides an overview of the regulations that must be followed and discusses approaches to satisfying the requirements. Consideration is given to proper handling of electronic data that are collected in a clinical trial. The components that constitute a clinical data archive are reviewed, and technical requirements for the correct use of open electronic data formats, such as XML (Extensible Markup Language) and SAS®, are discussed with an emphasis on ensuring long-term accessibility.

Introduction

Clinical data archiving includes planning, implementing and maintaining a repository of documents and/or electronic records containing clinical information, supporting documentation, and any interpretations from a clinical trial.

Scope

This section provides an outline to help clinical data managers develop an archiving strategy, working in conjunction with the project team and/or other appropriate department(s). Included are details of the regulatory requirements surrounding clinical data archives, a description of the components of an archive and information about data formats that can be used to support an archive. This document focuses on the components of the study archive that are the responsibility of data management.

Minimum Standards

- The clinical data archive should include a centralized table of contents.
- Accessibility of the clinical data archive electronic records should be tested following every major upgrade of the active clinical data management system.
- For paper case report form (CRF) studies, the original signed, dated, and completed CRF and original documentation of CRF corrections should be kept in the sponsor's files or offsite archive facility.
- The clinical data archive should be retrievable within a reasonable timeframe.
- For each study, the documentation should identify the hardware and software used, as well as specific version of the software or hardware.

Best Practices

- All clinical data, metadata, administrative data, and reference data should be maintained in an industry standard, open system format, such as CDISC's Operational Data Model (ODM).
- An electronic repository should link all study components, including the clinical data, CRF images in Portable Document Format (PDF) format, program files, validation records, and regulatory documentation.
- The audit trail should be stored in open format files in a secure system location.
- Copies of all user and system documentation for any applications used to collect or manage clinical data should be retained in the corporate library or archive facility.
- Reports describing the metadata and validation of study metadata, including data structures, edit check descriptions, and electronic data-loading specifications should be stored in the clinical data archive.
- System security reports, including user listings, access rights and the dates of authorization, should be printed and filed or scanned.

- The archive should include all program code for edit checks, functions, and sub-procedures, together with a copy of the version control information and validation documentation.

- Paper CRFs should be scanned and indexed. If an electronic data capture (EDC) system is used, all entry screens should be archived in an easily accessible format, such as a PDF file.

- Address archive responsibility for external data management providers. The sponsor should ensure that any signed contract with a vendor includes a section on archiving.

Background

The International Conference on Harmonisation Good Clinical Practice[1] (ICH GCP) requirements stipulate that data collected in a clinical trial must be maintained for a period of two years, following either the last regulatory submission or a decision to discontinue development of a compound, biologic, or medical device. To meet this requirement, as well as to ensure that the sponsor is able to answer questions related to clinical trial data that may emerge many years after the trial is conducted, it is important to archive clinical data, as well as the accompanying trial processing documentation.

Historically, the most common mechanism for long-term clinical data storage has been to extract the final data from the clinical data management system into SAS® datasets. The extracted SAS® datasets are still an important component of the clinical data archive; however, with the increasing importance of electronic regulatory submissions in recent years, requirements for clinical data archives are changing. As a result, clinical records that are part of an electronic submission must now comply with the 21 Code of Federal Regulations[2] (CFR) Part 11 ruling, which was originally published in 1997. Part 11 defines specific requirements with respect to authentication and auditing of electronic records. In addition, the Food and Drug Administration's (FDA) *Guidance for Industry: Computerized Systems Used in Clinical Investigations*[3, 4] defines requirements for data archiving. This guidance was published in 1999 and updated in 2007. To fully meet the requirements of these regulations and guidelines, a comprehensive archiving strategy is needed.

Regulations and Guidance

The tenets of 21 CFR Part 11 include no specific requirements for data retention or data archiving capabilities. However, the FDA has made it clear that the intent of the guidance is to supplement the predicate rules and ICH GCP requirements for those cases where electronic records are either directly or indirectly part of an electronic submission.

Guidance documents with specific mention of archive and record retention requirements include:

- *Guidance for Industry: Computerized Systems Used in Clinical Investigations*[3, 4] (CSUCI) published by the FDA in 1999 and updated in May 2007. This document describes requirements surrounding the need to preserve the systems environment in which electronic records are captured and managed.

- *ICH Good Clinical Practice*[1] (Section 5 Sponsor requirements) provides information about record retention requirements.

Regulatory guidance is being actively developed in the area of electronic records handling. Before finalizing your clinical data archive design, it is necessary to consult with the regulatory affairs specialists within your organization to ensure your design approach is consistent with the organization's regulatory policies.

Archive Contents

To successfully reconstruct a clinical trial, an auditor must be able to view not only the clinical data, but also the manner in which the data are obtained and managed. A summary of the types of information that should be included in a clinical data archive is provided in Table 1.

Table 1

Archive Component	Requirement
Clinical data	All data collected in the trial. This includes both CRF data and data that is collected externally (e.g., electronically submitted laboratory or patient diary data).
External data*	For data that are collected externally and loaded into a Clinical Data Management System (CDMS), the archive should include all loaded files, loading documentation, and quality control documentation.
Database Metadata	Information about the structure of clinical data, such as an annotated CRF. The annotated CRF will document the tables, variable item names, forms, visits and any other objects. It also includes codelists. This should also contain images of the entry screens (provided in PDF).
Coding Dictionaries	If data have been auto-encoded using a company dictionary or synonym table, a copy of the appropriate dictionary version should be included.
Laboratory Reference Ranges	If more than one version exists for reference ranges used during the course of the trial, each version should be retained in the archive. Documentation of the handling and processing of the laboratory ranges should also be present.
Audit trail	It is essential that the entire study's audit trail be included in the archive in a tamper-proof format.
Listings of edit checks, derived data, change controls	Edit check definitions and derived data change controls may be provided either as program listing files or as a report from the study definition application.
Discrepancy management logs, data handling guidelines	Listings of records that failed edit checks together with information on how the discrepancies were managed during the course of the study should be maintained.
Queries	Retain copies of all queries, query correspondence and query resolutions. Paper queries may be scanned and indexed.
Program code	Program code from data quality checking programs, data derivations and statistical analyses performed with clinical data and program documentation should be stored. Ideally, these documents should be stored online and indexed or hyperlinked.
CRF images in PDF format	For paper-based trials, CRF images are typically obtained by scanning the forms and converting them to PDF format. For trials using EDC, PDF images of the electronic forms may be created by the EDC application.
Data Management Plan (DMP)	PDF or paper version of MS Word and Power Point documents containing the study data management plan. The DMP may include sections or documents listed above.
Study Validation	Contents are described in the GCDMP chapter on systems

Archive Component	Requirement
Documentation	validation. This document may be in paper or electronic form.
Clinical Documents/Memos	Maintain copies of quality control documentation, database lock/freeze, SAE reconciliation, Personnel listing documents, etc.

*For data managed externally and then loaded into an in-house system for reconciliation, reviews, or other purposes, it is generally sufficient to limit the archive to actual data and any information pertaining to how the data are managed internally. When using an external vendor, the vendor is responsible for archiving any records that reflect how the data are managed in the vendor's system. The trial sponsor is ultimately responsible for ensuring that any vendor who provides trial data works in accordance with regulatory requirements. Therefore, the sponsor should ensure that any signed contract with a vendor includes a section on archiving. The information in this section should comply with both sponsor and regulatory requirements.

Technical Requirements

Designing a clinical data archive for long-term accessibility presents a challenge in the face of proprietary applications, tools, and platforms. This design should include input from all team members to ensure that the archive will meet department, corporate and regulatory requirements. A well-designed clinical study archive can facilitate compliance with the long-term data access requirements of the regulations for both paper based and electronic clinical trials. For this reason, the ideal clinical data archive should be based on standards and open systems.

The open formats that are typically used for clinical study archives are described in Table 2. No single format is ideal in all circumstances. Due to the fact that a study archive will usually include many different types of information, it will most likely include multiple formats. The format chosen for each type of information should be based on the likely future use of the information.

Table 2

Format	Description	Pro	Cons
Comma Separated Values (CSV)	Plain ASCII text with commas used as field delimiters. CSV files can be edited with text editors, word processors, and spreadsheet programs such as Microsoft® Excel.	Conceptually straightforward readily imported into almost any database.	Requires separate handling of metadata, administrative data and audit trails.
XML	Extensible Markup Language. Vendor independent, ASCII based technology for transfer of structured information between dissimilar systems. Used as the basis for the CDISC Operational Data Model.	Open standard ideally suited for clinical trial data. XML can include structural metadata, administrative data, and clinical data within a single file.	Still unfamiliar to many data managers and IT staff.
SAS® Transport files	Open source format provided by SAS® Institute Inc. Commonly used for submitting clinical data to the FDA. Can be read by the SAS Viewer that is distributed free of charge on the SAS Web site.	Familiar to clinical data managers and regulators. Works well with SAS data analysis tools.	Proprietary format. Variable naming restrictions. Requires separate handling of metadata, audit trails, and administrative data.
Adobe® PDF	Product provided by Adobe Systems Incorporated. Widely used standard for transmission of text documents. Default format for transmission of information to the FDA. Can be read by the Acrobat Reader, which is available free of charge from the Adobe® Web site.	Many applications output PDF files as an optional output format. Reader is available free of charge.	Predefined PDF output from EDC applications may not comply with or have the flexibility to produce standardized Sponsor formats.

Long-term data access requirements suggest that the choice of data format is limited to ASCII based formats, or formats based on an open standard, such as SAS® Transport files. The choice may be further influenced by the format used in the original data management or data collection system.

Archives for Clinical Sites

The CFR predicate rules and the ICH Good Clinical Practice (GCP) guidelines specify that a copy of clinical data must be retained at the investigator site throughout the records retention period. For paper based studies, this can be achieved by keeping a copy of the paper records at the site. For EDC studies it is important to have a strategy in place for ensuring that these guidelines are met appropriately. Many EDC vendors will provide PDF files for all of the electronic Case Report Forms (eCRFs) collected from the site during the trial. The Clinical Data Manager (CDM) may provide assistance and/or coordination with this procedure. If your company builds EDC studies in-house, the data manager will be responsible for ensuring the quality of the PDF outputs prior to sending the files back to the clinical sites.

Recommended Standard Operating Procedures

- Study Archiving Procedures
- Document Archiving Procedures

References

1. International Conference on Harmonisation. Good Clinical Practice. *Federal Register*: 1997.
2. *Code of Federal Regulations*, Title 21, Volume 1, Part 11. Food and Drug Administration. US Government Printing Office, 1998.
3. *Guidance for Industry: Computerized Systems Used in Clinical Trials*. Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; April 1999.
4. *Guidance for Industry: Computerized Systems Used in Clinical Investigations.* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration; May 2007.

Further Reading

The Data Archiving Challenge. Clinical Data Interchange Standards Consortium; May 2000.

Paul Bleicher. Diamonds May Be Forever, But Data? *Applied Clinical Trials*; February 2002

CDISC (Clinical Data Interchange Standards Consortium) Web site, Available at: http://www.cdisc.org/. Accessed on November 28, 2007.

Chapter Revision History

Publication Date	**Comments**
September 2003	Initial publication.
May 2007	Revised for style, grammar, and clarity. Substance of chapter content unchanged.
June 2008	Revised for content, style, grammar, and clarity.

This page is intentionally blank.