UKCRC Registered Clinical Trials Units

Archiving electronic data - points for consideration

UK CLINICAL RESEARCH COLLABORATION

igniting our potential

Archiving electronic data points for consideration

Contents

1. Background

Clinical trials units (CTUs) are involved in the running and oversight of clinical trials and other research studies. Documentation and data are produced for each study, and it is important to retain this for a period after the study ends to demonstrate good clinical practice. Wherever possible these documents and data should be stored in a format to facilitate more efficient reproduction of results.

2. Purpose

The purpose of this document is to provide guidance and an aide memoire for Clinical Trials Units (CTUs) of points to consider when creating an electronic study data archive. It may also be useful as a means of assessing the ongoing suitability of a CTU's archiving policies and procedures.

3. Scope

The scope of this document includes archiving study-related data, metadata and associated documentation. As such, it relates to the archival of studies rather than CTU-level documentation. It does not relate to routine database backups.

This document is intended for use by project managers, data managers, Information Systems (IS) staff or dedicated archivists, and to be used in conjunction with local Standard Operating Procedures (SOPs) and guidance and study-specific requirements.

Archiving physical content, such as paper records and printed documentation, is out of scope (cf. *Good Clinical Practice Guide, ch.10* for guidance on physical archival).

The content is not definitive and provides a suggested approach. Local requirements and overarching legislation must always be considered.

A glossary of relevant terms is provided in Appendix 1.

4. Context

The study sponsor should appoint individuals responsible for archiving. The nominated archivist(s) should ensure that they are following the appropriate regulatory, legal and industry standards. Relevant regulatory frameworks may include the Medicines for Human Use (Clinical Trials) Regulations, Good Clinical Practice (GCP), Medicines and Healthcare products Regulatory Agency (MHRA) and European Medicines Agency (EMA) guidance; legal considerations may include the Data Protection Act 2018 and General Data Protection Regulation (GDPR); industry standards may include ISO27001 and Cyber Essentials Plus. Links to these and other relevant documents can be found at Appendix 2. The archivist should also consider relevant local or third-party operating procedures and policies that relate to archival or retention.

The required retention period for archived clinical trials documentation is dependent on a variety of factors (see Appendix 3 for further information). Notwithstanding this, the retention period should be agreed during study set-up, documented in the trial protocol and included

in trial agreements, and considered prior to authorising deletion or destruction of records in case there is a reason for extending the retention period further.

The archive should include a true copy of electronic data, as defined in MHRA 'GXP' Data Integrity Guidance, including relevant metadata and associated information (e.g. context, layout, electronic signatures and authorisations), and the full GXP audit trail. Consideration should also be given to the ability to maintain or restore the study database within its original operating environment, enabling dynamic functionality where this is critical to its integrity or later verification.

5. Considerations for archiving

CTUs should have an SOP or similar local documentation that outlines the processes for electronic archiving and subsequent access or retrieval, and addresses both CTU-level and study-level actions. Suggestions for these are detailed in Table 1, as follows.

Table 1: CTU-level considerations and study-level actions for archiving

	CTU considerations	**Study-level actions**

Archivists	• Nominate archivist(s)
	  ○ Role-based (DM, IS, other) or named individuals
	  ○ Consider qualifications, i.e. education, training and experience
		• Assign archivist, independent from the study where possible

Archive management	• Define process for identifying studies that should be archived, including triggers and timescales • Ensure oversight of archiving process and periodic archive review ○ Consider establishing oversight committee • Define process for identifying archived studies or supporting data with elapsed retention periods that require deletion • Risk assessment to support design and validation of electronic archiving arrangements	• Define time constraints on when the study should be archived, e.g. a defined period after publication of the primary outcome • Establish and document prevailing retention period for the archive, as per protocol / PIS / consent form • Establish legal requirements relating to the type(s) of study data • Establish whether there is a requirement to retain some data for different periods, e.g. longer / indefinite retention of anonymised data for secondary analysis.

	CTU considerations	Study-level actions
Archival request process	• Which role and/or process step can initiate archival (CI, sponsor, senior TM)? ○ This may depend on the type of study • How should requests be submitted (form, email)? • How are requests logged and acknowledged?	• Receive, record and acknowledge request

Archive infrastructure	• Offsite or onsite
	• Locally hosted or third-party cloud-based storage
	  ○ Consider any contractual conditions and constraints, service level agreements, and explicit statements regarding data ownership.
	• Consider security
	  ○ Firewalls
	  ○ Physical access restrictions
	  ○ Environmental controls
	  ○ Backups
	• Consider data integrity
	  ○ Ensure archives can only be accessed by appropriate individuals
	  ○ Direct access to the archive should be read-only
	• Storage capacity
	• Virtual environments/emulators
	• Archive infrastructure maintenance
	• Availability
	• Cost
	• Appropriate media
	• Ability to log file access/operations to detect possible changes to archived content
	• Compression for efficient storage and/or to minimise hosting costs.
	  ○ If used, how will content be decompressed?	

	CTU considerations	Study-level actions
Archiving	• Archiving processes should be validated, operated and maintained in ways that are compliant with GCP • Establish standard for what constitutes study-data, metadata and related documentation for the CTU. See Appendix 4 for more details and additional suggestions • Establish suitable archive file format standards, including risk-based assessment of non-proprietary/open standard formats, to guard against technology obsolescence (consider CSV, ODM-XML formats)	• Establish appropriate folder structure for the content to be archived • If study contains routine data from other providers (e.g. NHS Digital), action specific retention / deletion requirements • Remove identifiable data from study data if required (document data deleted/anonymised) • Archive relevant content, in appropriate format(s), into appropriate location • Log what has been archived, where, when and in what format • Record document characteristics for archived content to allow future integrity checking, e.g. checksums or file size and record count • Remove any live links or references to archived source material • Test ability to access, and restore, archived material • Erasure of original electronic source material if required, e.g. deleting original study data • Acknowledge/confirm archival to the requester • Final sign-off of study archive
Restoring an archive	• Which role and/or process step can initiate access to or restoration from an archive (CI, sponsor, senior member of CTU, archive owner, inspector/auditor)? • How should restore requests be submitted (form, email)?	• Receive, record and acknowledge request ○ Establish scope and purpose of restore • Perform restore of relevant content to active file storage

	CTU considerations	Study-level actions
	• How are restore requests logged and acknowledged? • What contingency measures are in place for full or partial restore failure?	• Validate restored content, e.g. compare checksum(s) of restored content against original checksum(s) • Once purpose of restore is met, delete restored files from active file storage

Acknowledgements

Martin Pond, Norwich Clinical Trials Unit, University of East Anglia

Tim Chater, Liverpool Clinical Trials Centre, University of Liverpool

Claire Johnson, Centre for Trials Research, Cardiff University

Adam Barrett, Leeds Clinical Trials Research Unit, University of Leeds

Charlotte Potter, Centre for Trials Research, Cardiff University

David Murray, NPEU CTU, University of Oxford

John Turgoose, Hull Health Trials Unit, University of Hull

Jonny Wilks, Peninsula Clinical Trials Unit, University of Plymouth

Paul Maslowski, Leicester Clinical Trials Unit, University of Leicester

Ruth Wood, Newcastle Clinical Trials Unit

Sarah Sumpter, Hull Health Trials Unit, University of Hull

Simon Lumley, Papworth Trials Unit Collaboration

Steph Foster, Institute of Cancer Research, Clinical Trials & Statistics Unit

Tola Erinle, Priment Clinical Trials Unit, University College London

Appendix 1: Glossary

Archive	The facilities and supporting resources necessary for the secure retention, maintenance and retrieval of materials accumulated by an organisation
Archivist	An individual designated to be accountable for the management of the archive, i.e. for the operations and procedures for archiving
Audit trail	A formal record or log of all transactions (creations, updates, deletions) of the study database
CRF	Case Report Form – data collection forms which together capture all participant data
CTIMP	Clinical Trial of Investigational Medicinal Product
CTU	Clinical Trials Unit
Electronic archive	The designated repository in which electronic records are retained for their long-term preservation
Cyber Essentials	Cyber Essentials is a Government-backed and industry-supported scheme that helps businesses protect themselves against the growing threat of cyberattacks and provides a clear statement of the basic controls organisations should have in place to protect them
Cyber Essentials Plus	As Cyber Essentials but independently verified rather than self-declared
DMP	Data Management Plan
DSP Toolkit	The Data Security and Protection Toolkit (DSP Toolkit) is an online self-assessment tool that allows organisations to measure their performance against the National Data Guardian's 10 data security standards
EMA	European Medicines Agency
GCP	Good Clinical Practice (GCP) ICH E6(R2) is the international ethical, scientific, and practical standard to which all clinical research is conducted. It was established under the auspice of the International Conference on Harmonisation (ICH) of technical requirements for registration of pharmaceuticals for human use.
GDPR	General data protection regulation (GDPR) https://gdpr-info.eu/
IS	Information Systems
ISO27001	ISO/IEC 27001:2013 (also known as ISO27001) is the international standard that sets out the specification for an information security management system (ISMS)
Metadata	Information associated with data that provides context and understanding, i.e. data about data. Most commonly this is data that describes the structure, data elements, inter-relationships and other characteristics of electronic records. Metadata also permit data to be attributable
MHRA	The Medicines and Healthcare products Regulatory Agency (MHRA) is an executive agency of the Department of Health and Social Care in the United Kingdom which is responsible for ensuring that medicines and medical devices work and are acceptably safe.

MoU	Memorandum of Understanding
ODM XML	Vendor-neutral, platform-independent format for exchanging and archiving clinical and translational research data, along with their associated metadata, administrative data, reference data, and audit information
PIS	Patient Information Sheet
SLA	A service-level agreement (SLA) is a contract that establishes a set of deliverables that one party has agreed to provide another. This agreement can exist between a business and its customers, or one department that delivers a recurring service to another department within that business
SOP	Standard Operating Procedure
TM	Trial Manager
UKCRC	UK Clinical Research Collaboration

Appendix 2: Key legislation and industry standards

Legislation	• The Medicines for Human Use (Clinical Trials) Regulations 2004 (https://www.legislation.gov.uk/uksi/2004/1031/contents/made) • The Medicines for Human Use (Clinical Trials) Amendment Regulations 2006 (https://www.legislation.gov.uk/uksi/2006/1928/made) • Data Protection Act 2018 (https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted) • Regulation (EU) 2016/679 of the European Parliament - GDPR (https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:02016R0679-20160504) • Regulation (EU) 536/2014 of the European Parliament - clinical trials on medicinal products for human use (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014R0536) • Electronic Records, Electronic Signatures, US Code of Federal Regulations - Title 21 Part 11 (21CFR11) (https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11)
Standards	• ISO27001 Information Security Management (https://www.iso.org/isoiec-27001-information-security.html) • Cyber Essentials / Cyber Essentials Plus (https://www.ncsc.gov.uk/cyberessentials/overview) • Data Security & Protection (DSP) Toolkit (https://www.dsptoolkit.nhs.uk/) • EMA ICH E6 (R2) Good clinical practice (https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice)
Guidance	• EMA guideline on the content, management and archiving of the clinical trial master file (paper and/or electronic) 2018 (https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-content-management-archiving-clinical-trial-master-file-paper/electronic_en.pdf) • MHRA Clinical investigations of medical devices – guidance for manufacturers (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/989420/Guidance_for_mfrs_on_clinical_investigations-May_2021.pdf) • MHRA 'GXP' Data Integrity Guidance and Definitions – March 2018 (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf)

Appendix 3: Estimating retention period

Type of Trial	Location of trial data	Archive retention time	Referenced regulations / guidance
CTIMP - data used for marketing authorisation / licensing of drug	ICH GCP region	At least 2 years after the last approval of a marketing application	ICH GCP (as amended)
CTIMP – data not used for marketing authorisation / licensing of drug	UK only	Medical files and the TMF are retained for at least 5 years after the conclusion of the trial	The Medicines for Human Use (Clinical Trials) Regulations (SI 2004/1031 as amended)
	EU	TMF for minimum of 25 years. Medical files in accordance with National Law	REGULATION (EU) No 536/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC
	3rd country	Refer to applicable laws	Relevant national law
ATIMP	UK & EU	30 years after the expiry date of the product or longer if required by the clinical trial authorisation	Guidelines on Good Clinical practice specific to Advanced Therapy Medicinal products (2019
Clinical Investigation of a medical device	Great Britain only	Minimum of 5 years	Medical Devices Regulations (SI 2002 No 618, as amended)
	EU and Northern Ireland	At least 10 years after the clinical investigation with the device in question has ended, or in the event that the device is subsequently placed on the market, at least 10 years after the last device has been placed on the market. In the case of implantable devices, the period shall be at least 15 years	Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC

Type of Trial	Location of trial data	Archive retention time	Referenced regulations / guidance
			Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU
	3rd country	Refer to applicable laws	Relevant national law
Non-clinical and non-public health research projects	Any	Follow funder requirements. If there are no funder requirements retain for a minimum period of 5 years after the end of the project* or after publication of any findings based upon the data (whichever is later)	Local retention policies
Research records and data from clinical or public health research projects not including clinical trials of investigational medicinal products	Any	Follow any funder requirements. If there are no funder requirements retain for a minimum period of 15 years after the end of the project* or after publication of any findings based upon the data (whichever is later)	Local retention policies

** End of project is defined as completion of project closure report or publishing of final articles.*

Please note: if there are multiple rows that would apply to a study then typically the longest retention period would apply. For example, an international trial that is run in multiple countries would require the maximum retention period, rather than minimum. Also, the retention periods in the prevailing approved clinical trial protocol must be adhered to (if they are longer than that stated above for that type of trial or study), since this is the period that has been decided by the sponsor and approved by regulatory authorities and REC for that particular trial.

Appendix 4: Example archive content

	Notes
Study data	Multiple formats, open/non-proprietary formats to mitigate technical obsolescence. Alongside data from the main clinical database, this may also include data from multiple other sources, such as ancillary datasets, pharmacovigilance and randomisation systems, IVRS/IRT systems, ePRO, eConsent, and so on
Derived data	Analysis datasets or queries to reproduce them
Metadata	Open standards-based definition, e.g. ODM XML
Audit trail	• Include whether separate or integrated • Readable, reproducible
Documents to include	• DMP (or SLA/MoU if appropriate) • CRFs, protocol requirements, supplementary design documents • Testing and approval documents • Training materials (internal and external) • Database specification / data dictionary • Data validation specification • Summary of queries issued throughout the study (and their status/resolution) • Delegation log and evidence of training • Database revision/change log – what, why, when, how, and who authorised • Issues and related action log • Data sharing agreements, inc. anonymisation, transfers and encryption requirements • Data imports and log thereof • Self-evident correction rules, log of instances where applied • Data exports • Relevant messages from email accounts, and possibly messenger application messages (e.g. SMS, WhatsApp, Teams) and collaboration software recordings, transcripts or conversations (e.g. Teams, Zoom)
Possible considerations	• Application software where it is not possible to export data in an open standard • Virtual environments where these offer critical bespoke/custom functionality