A Risk Based Approach To Long Term Data Integrity of Clinical Data And The Role of CDISC Standards

Matthew Addis, matthew.addis@arkivum.com

www.arkivum.com

Clinical Data: Long Term Data Integrity

- Risk Based Approach
- ALCOA++ Data Integrity
- Long Term Digital Preservation
- Qualification and Validation

Regulations and Guidelines

- **GxP Data Integrity** applies to all stages of the data lifecycle, including **archiving** [2].
- Retention periods can be 25 years or more, for example eTMF under GCP [1].
- The **ALCOA++ principles** need to be followed using a **risk-based approach** [12].
- Systems used for archiving need to be validated and from qualified suppliers.

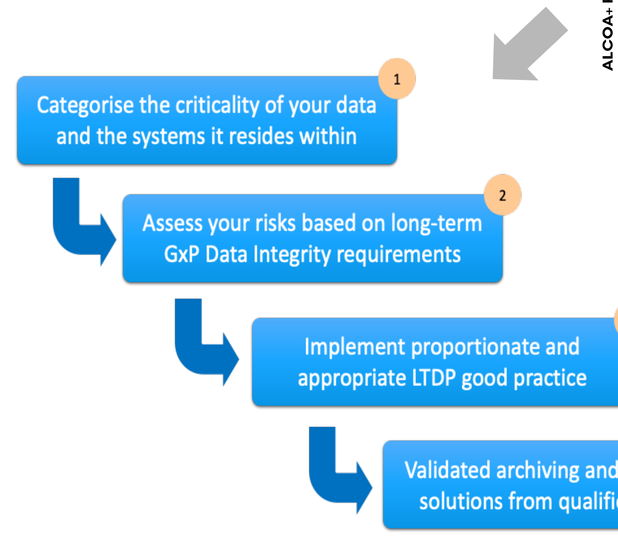
Digital Preservation

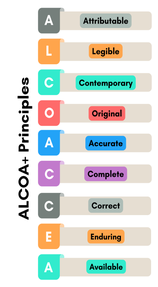
- The **digital preservation community** has developed **good practice** and **guidelines** for ensuring digital content remains accessible and usable for decade timescales.
- Good practice covers data, metadata, safe storage, long term access and more.
- **Long Term Digital Preservation good practice** directly supports the **ALCOA++** principles and can be used to **reduce the long-term risk** of loss of Data Integrity [10]
- LTDP requirements and good practice is ideal for including in a URS, forming part of Computerised Systems Validation, and selecting and qualifying suppliers [11].

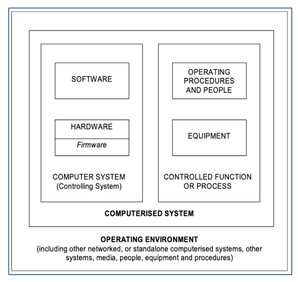
CDISC Standards

- Digital Preservation good practice recommends the use of open standards and specifications for data that are widely adopted and freely available [9]
- Using **CDISC standards** [8] helps ensure data will retain **ALCOA++ Data Integrity**.

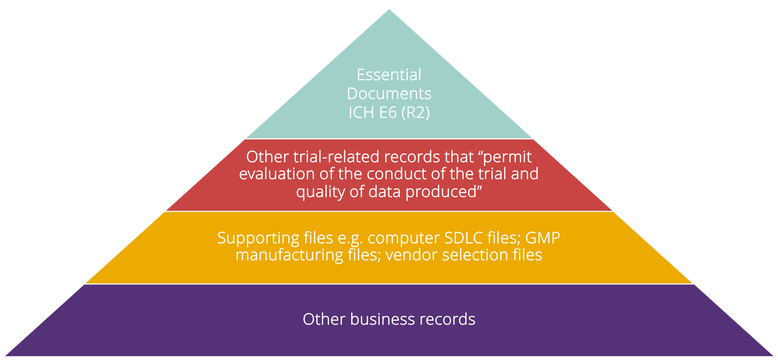
Probability	Harm severity			
	Minor	Marginal	Critical	Catastrophic
Certain	High	High	Very high	Very high
Likely	Medium	High	High	Very high
Possible	Low	Medium	High	Very high
Unlikely	Low	Medium	Medium	High
Rare	Low	Low	Medium	Medium
Eliminated	Eliminated			

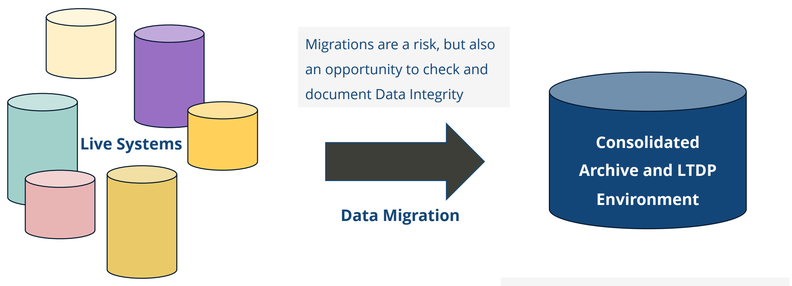






Archive Data and Systems





Live systems become end of life systems
Live systems grow in number
Leaving data in live systems is a risk multiplier

Manage risks in a consistent way
Use specialist staff, systems and services
Active Long Term Digital Preservation

Risks: Challenges and Causes

Legible and Traceable
- Will documents and data still be readable after 25 years?
- Can the audit trail be used to recreate events from 25 years ago?
- Is there documentation so someone can still understand the data?

Enduring
- Could data become corrupted or lost when it is being stored?
- Will data become distributed across and locked into many EoL systems?
- Is the data immutable and can attempts to change it be detected?
- Are there tight controls over who can remove or delete data?

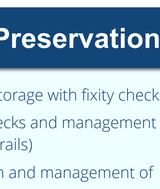
Attributable, Accurate, Contemporaneous
- Can timestamps be altered?
- Is the audit trail permanent?
- Will signatures always validate?

Complete and Correct
- Can any deletions or changes go unnoticed or unapproved?
- Can data integrity issues go undetected during transfers or archiving?
- Can you prove the entire TMF (structure, files, metadata) is complete?
- Can you prove migrations (formats, systems, people) were successful?

Available
- Can data be discovered easily (metadata)?
- Can data be retrieved quickly (ready access)?
- Is everything documented?
- Will data become spread across legacy systems and be impossible to find?
- Does BCDR cover cyberattacks, vendors going bust, disasters in the cloud?
- Are there sufficient budget, staff and skills to sustain the archive?

Data Integrity Risks

Probability	Harm severity			
	Minor	Marginal	Critical	Catastrophic
Certain	High	High	Very high	Very high
Likely	Medium	High	High	Very high
Possible	Low	Medium	High	Very high
Unlikely	Low	Medium	Medium	High
Rare	Low	Low	Medium	Medium
Eliminated	Eliminated			



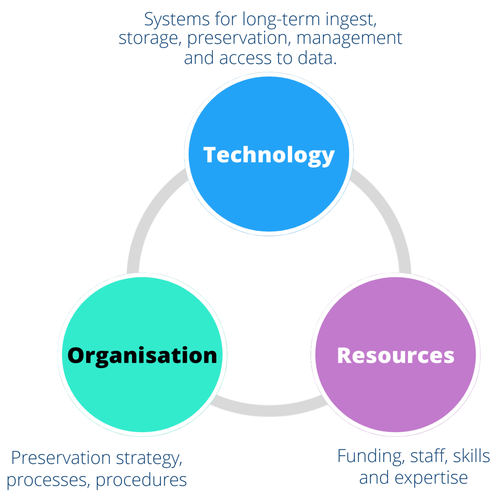
The sponsor should evaluate potential risks by considering:
(a) the likelihood of harm/hazard occurring;
(b) the extent to which such harm/hazard would be detectable;
(c) the impact of such harm/hazard on trial participant protection and the reliability of trial results.

ICH E6 (R3) Risk Management

Risks: Impact and Consequences

- Health and safety of study participants
- Trial Results are no longer reliable
- Rejection or delay to marketing application
- Removal of drug from market
- Financial penalties such as fines
- Quality issues with products
- Cost of repeat or additional additional work
- Reputational damage
- Delayed sale or MNA
- Ethical issues where quality can't be proven

Long Term Digital Preservation (LTDP)



"The series of managed activities necessary to ensure continued access to digital materials for as long as necessary"

Digital Preservation Coalition

LTDP Good Practice

- **The digital preservation community has 30+ years of experience in ensuring digital content remains accessible and usable over time.**
- Communities and organisations include the Digital Preservation Coalition (DPC) [3] and Open Preservation Foundation (OPF) [4].
- Good Practice Models and Guidelines include the **DPC Rapid Assessment Model (DPC RAM)** [6]. Core Trust Seal [5] and the NDSA Levels of Preservation [7]
- These guidelines cover on how data is should be stored (multiple copies, different locations, regular fixity checks), how content can be preserved (format migrations, software emulation), how metadata (descriptive, technical, structural, preservation) can be managed (standards and reuse), how content can be accessed (discovery, delivery, rendering, viewing), and more.
- **LTDP Good Practice such as DPC RAM, can be mapped to the ALCOA++ principles and provides a way to reduce risks to long term Data Integrity** [10]
- CDISC standards fit well with LTDP guidelines such as the risk assessment work done by the National Archives and Records Administration (NARA) [9]
- **Arkivum has over 10 years experience of applying LTDP to Life Sciences data**, we are a DPC Supporter and Digital Preservation Award Winner.

Data Archiving

- Place where data is held for safe keeping
- Data is typically read-only
- Backed up
- Restricted access
- Kept 'as-is' with no changes or updates
- Sometimes held within a live system, e.g. after data is 'locked'
- Often treated as the digital equivalent of 'boxes of paper in a storage facility'
- Not a viable solution for data that needs to be readable and usable for 25 years!

vs.

Digital Preservation

- Long-term safe storage with fixity checks
- Data Integrity checks and management (files, metadata, audit trails)
- Technology watch and management of technical obsolescence
- Preservation actions so content maintains its meaning and remains usable
- Metadata ensures content is documented, discoverable and usable
- Evidence of ongoing data integrity through application of digital preservation
- All the processes, techniques and systems for indefinite retention and use

References and Further Reading

[1] EU CTR 536/2014: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32014R0536
[2] EMA/INS/GCP/112288/2023: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-and-electronic-data-clinical-trials_en.pdf
[3] Digital Preservation Coalition: https://dpconline.org
[4] Open Preservation Foundation: https://openpreservation.org/
[5] Core Trust Seal: https://www.coretrustseal.org/
[6] DPC RAM: https://www.dpconline.org/digipres/dpc-ram
[7] NDSA Preservation Levels: https://ndsa.org/publications/levels-of-digital-preservation/
[8] CDISC Standards: https://www.cdisc.org/standards
[9] NARA format risk assessment: https://www.archives.gov/preservation/digital-preservation/risk
[10] Arkivum eBooks (LTDP, Digital Preservation, Assessing Suppliers): https://arkivum.com/blog/
[11] DPC Procurement Toolkit: https://www.dpconline.org/digipres/implement-digipres/procurement-toolkit
[12] ICH E6 (R3): https://database.ich.org/sites/default/files/ICH_E6%28R3%29_DraftGuideline_2023_0519.pdf

matthew.addis@arkivum.com