A risk-based approach to eTMF data integrity across the whole trial lifecycle

Jim Markley, Associated Director of Consulting, Just In Time GCP
Matthew Addis, Chief Technology Officer, Arkivum

Meet the Speakers

Jim Markley

Title: Associate Director of Consulting

Organization: Just in Time GCP

Jim has been in the industry for over 10 years starting at a clinical research site, moving to a large CRO, and then moving into consulting at JiT. He is responsible for helping clients optimize and improve business processes to ensure TMF Completeness and Inspection Readiness. He is a member of the Controlled Terminology and Risk Initiatives within CDISC.

Matthew Addis

Title: Chief Technology Officer

Organization: Arkivum

Matthew is CTO and co-founder of Arkivum, where he responsible for technical strategy and is Arkivum's subject matter expert on data archiving and long term digital preservation, including how these can be applied in regulated environments to GxP data.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *{Please disclose any financial relationship or conflict of interest relevant to this presentation here OR}*
- *The author(s) have no real or apparent conflicts of interest to report.*

Agenda

1. Introduction
2. Defining a risk-based approach for the whole trial lifecycle
3. Archiving, Digital Preservation and long-term Data Integrity
4. Questions

Introduction

Maintaining data integrity of your TMF throughout the trial lifecycle is challenging.

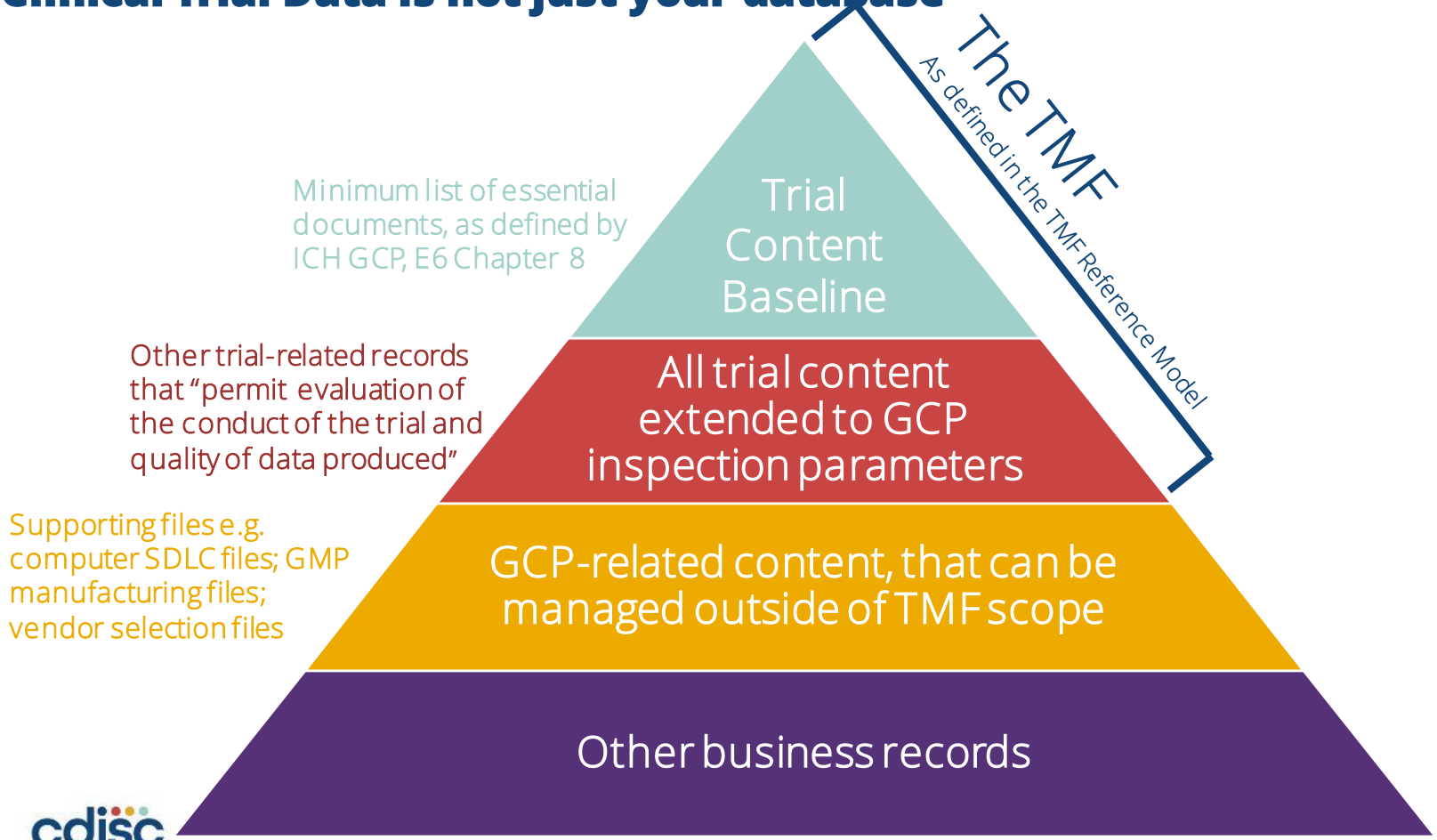
We believe the key to success is to:

a. prepare for archiving throughout the whole data lifecycle
b. employ recognised standards such as the eTMF Reference Model
c. take a quality and risk-based approach at all stages including archiving
d. apply recognised digital preservation good practice
e. embody this in a well thought out Data Management Plan / eTMF plan

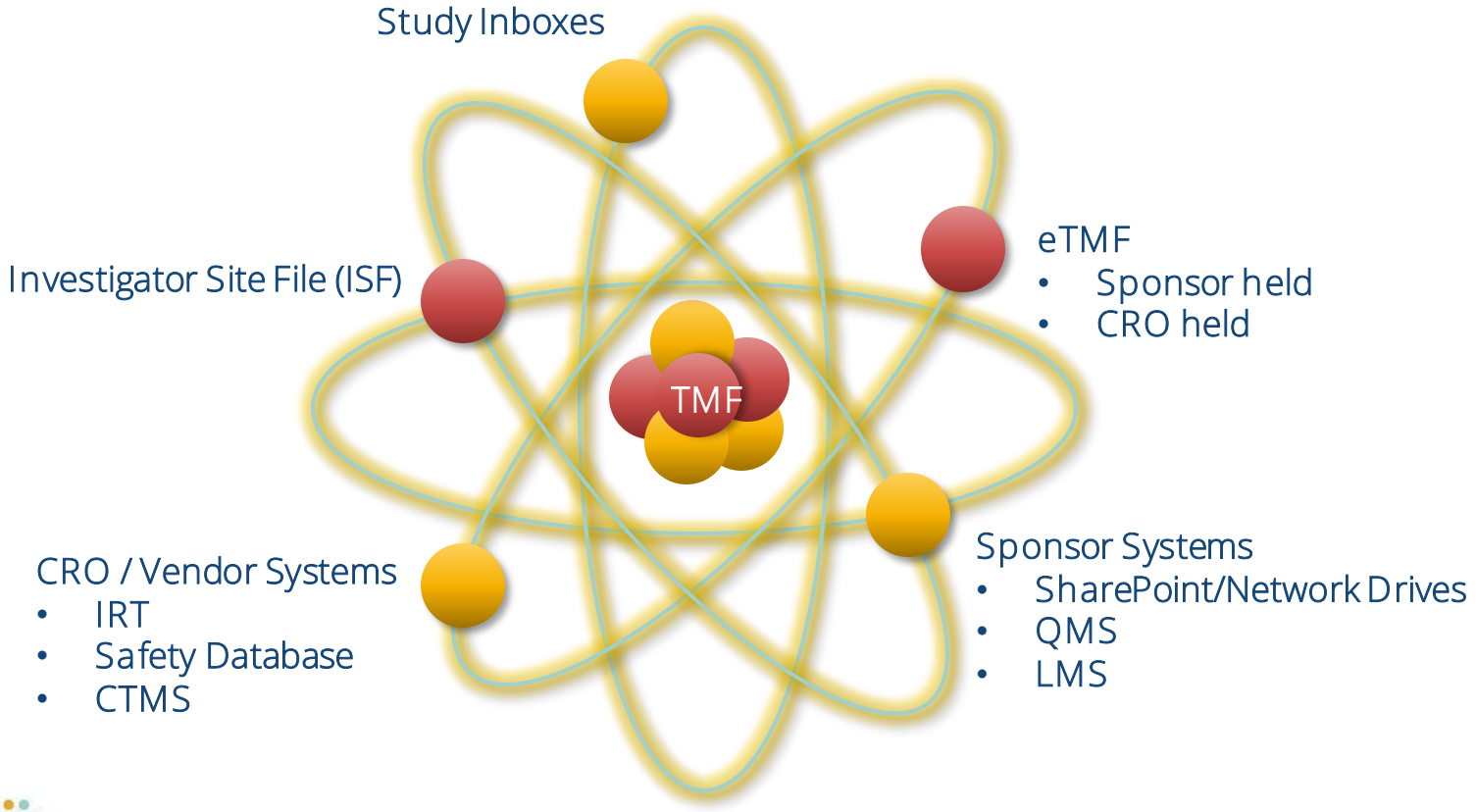
Defining a risk-based approach for the whole trial lifecycle

Jim Markley, Just In Time GCP

Clinical Trial Data is not just your database

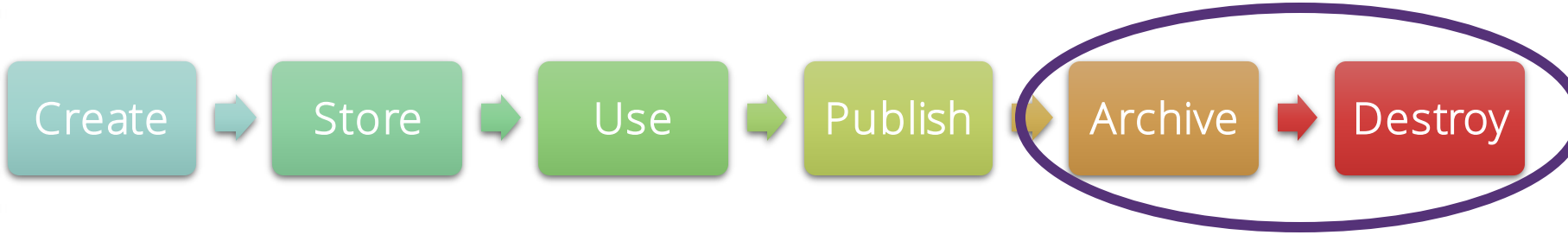


Where is your Clinical Trial Data Stored?



What is the Clinical Trial Data Lifecycle?

The end of the trial is not the time to figure out what you are going to do with your data!



PI signs a Financial Disclosure Form

FDF gets filed in the eTMF

FDF is used to determine the need for any disclosure

FDFs provided to Regulatory Authority as needed

FDFs held in an archive location for required time frame

Records destroyed

cdisc

The end of the trial is not the time to figure out what you are going to do with your data!

- ICH E6 (R3) emphasizes the need to incorporate **"Quality by Design"** into all elements of clinical trials. This means that Critical Processes and Data should be identified and there should be a plan for:
  - What is produced as evidence of Critical Processes and Data?
  - Where will it be stored?
  - How will it be used?
  - What does oversight look like?
  - When is it time to archive your data?
  - How do you archive your data?

We recommend performing a Risk Assessment to help determine what is critical, which can then be leveraged to develop your Risk Based Approach.

cdisc

The Basis of a Robust TMF Strategy

- **Company Level**
  - **Business Rules** should be developed to set quality standards across programs
- **At the beginning of the study:**
  - A **Risk Assessment** should be performed to identify Critical Processes, Data, and Documents.
  - A **TMF Plan** should be developed for managing the TMF and defining the risk based oversight strategy
  - A **TMF Index** should be developed that outlines where all relevant content will be held
- **Throughout the life of the study:**
  - **Periodic reviews** should be performed to help identify any trends in what is being observed and help ensure missing content is collected in a timely manner
- **At the end of the study:**
  - A final review should be completed to serve as a quality oversight sign off on the finalized data

cdisc

Risk Based Approach Example

Risk Assessment

Primary Endpoint requires assessment by trained raters and therefore evidence of their training is critical.

TMF Plan

Scope of Oversight Review always includes crosscheck of individuals performing this assessment against their training documentation.

TMF Index

Clearly delineates where training documents are being filed and who is responsible for filing them

cdisc

Oversight Reviews

- An Oversight Review should be a risk based approach determined by your risk assessment.
- It may include a check for expected content that should be present by a specific milestone focused around higher risk content.
- It is done in a cyclical manner throughout the life of the study to ensure expected content is filed correctly
- Issues identified during the review and discussed with content owners

Leveraging Technology to Oversee Risk

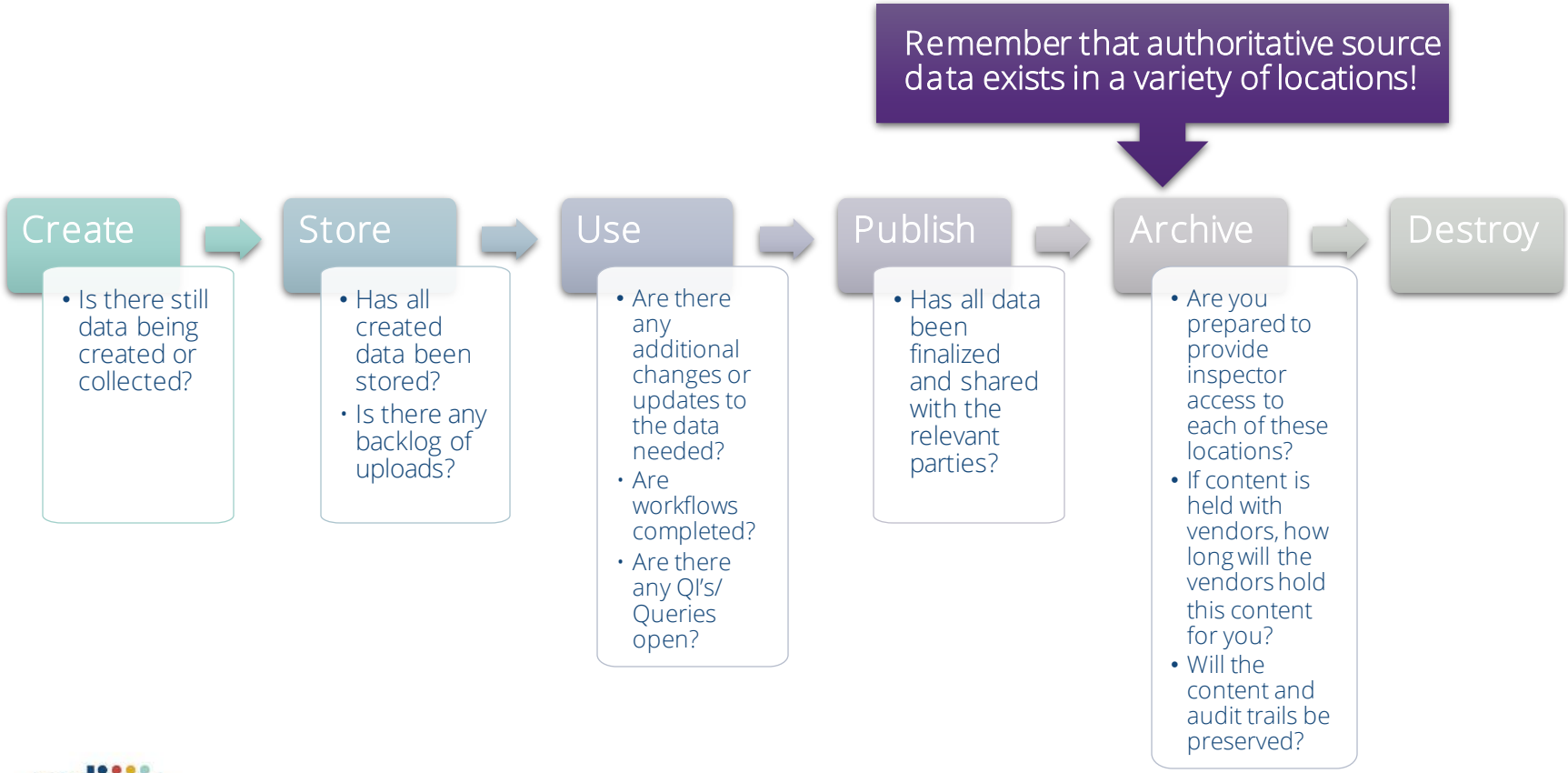
Reporting & Dashboards

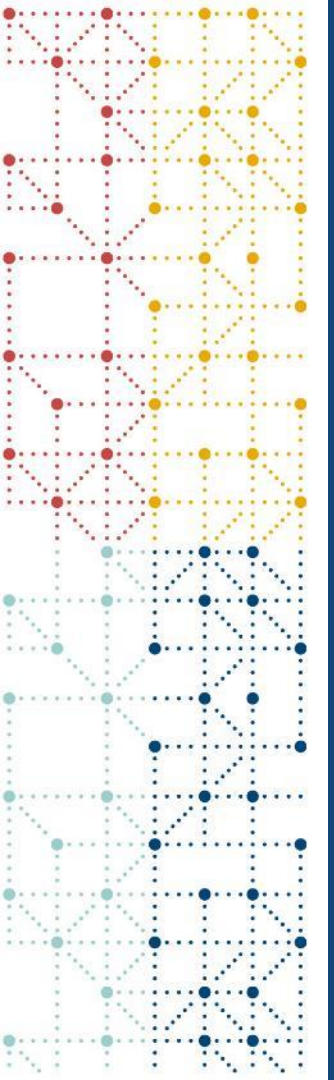
- Leverage **reporting & dashboard** to perform high level logic checks to identify areas where additional review may be needed
- Leverage **focused reports** to pull only the documents/scope of review

Features & Functionality

- Create and maintain **placeholders** to ensure missing or expected documents get resolved
- Management of **Expected Document Lists** can help inform sites or artifacts that require deeper review
- Triggering of **Milestones & Events** can be customized to ensure that your high risk content is being actively tracked in the system

Preparing for Archive





Archiving, Digital Preservation, and long-term Data Integrity

Matthew Addis, Arkivum

Regulations and Guidelines

- GxP Data Integrity applies to all stages of the data lifecycle, including archiving
- Archived data is subject to:
  - Long term retention, e.g. 25 years for TMF (can be 30 years)
  - ALCOA+ principles
  - Risk based approaches and QbD
  - Computerised Systems Validation
  - Supplier Qualification

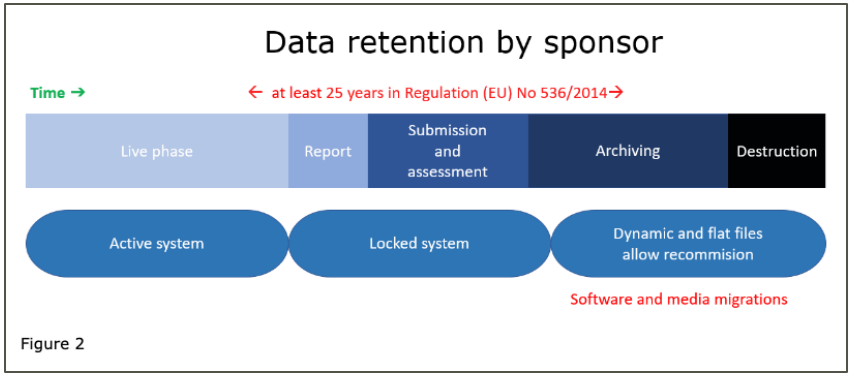


Figure 2

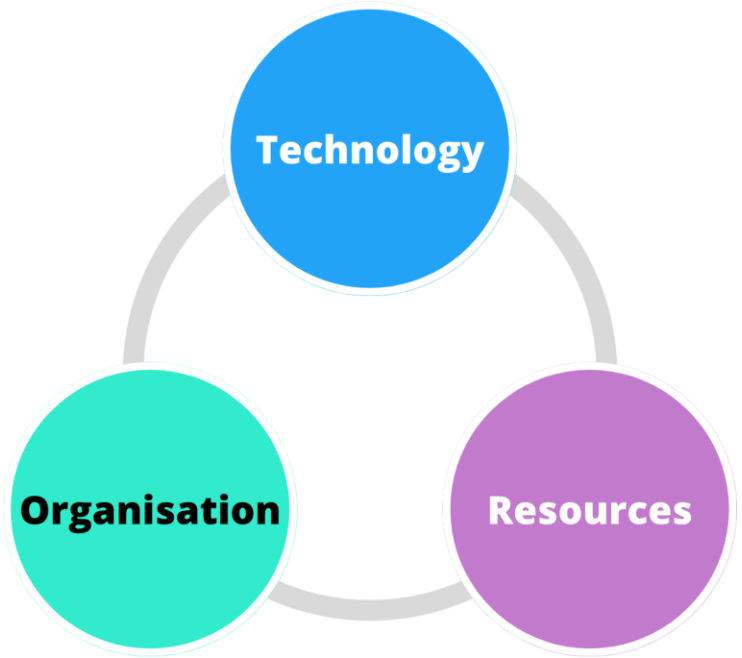
2023 EMA guidelines on computerized systems for clinical data

Digital Preservation

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

Digital Preservation Coalition

Long-term ingest, storage, preservation, management and access.



Preservation strategy, processes, procedures

Funding, staff, skills and expertise

http://www.dpworkshop.org/dpm-eng/eng_index.html

Archiving and Preservation Are Not The Same!

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 and application of digital preservation
- All the processes, techniques and systems for indefinite retention and use

cdisc

GxP Data Integrity: Risk Management

Consequences / Impact of a Data Integrity Failure

- Health and safety of study participants and patients
- Failed Inspections, CAPAs
- Rejection or delay to marketing application
- Removal of drug from market
- Financial penalties
- Quality issues with products
- Cost of doing repeat work
- Cost of doing additional work
- Reputational damage
- Delayed sale or MNA
- Ethical issues

Likelihood of a Data Integrity Failure

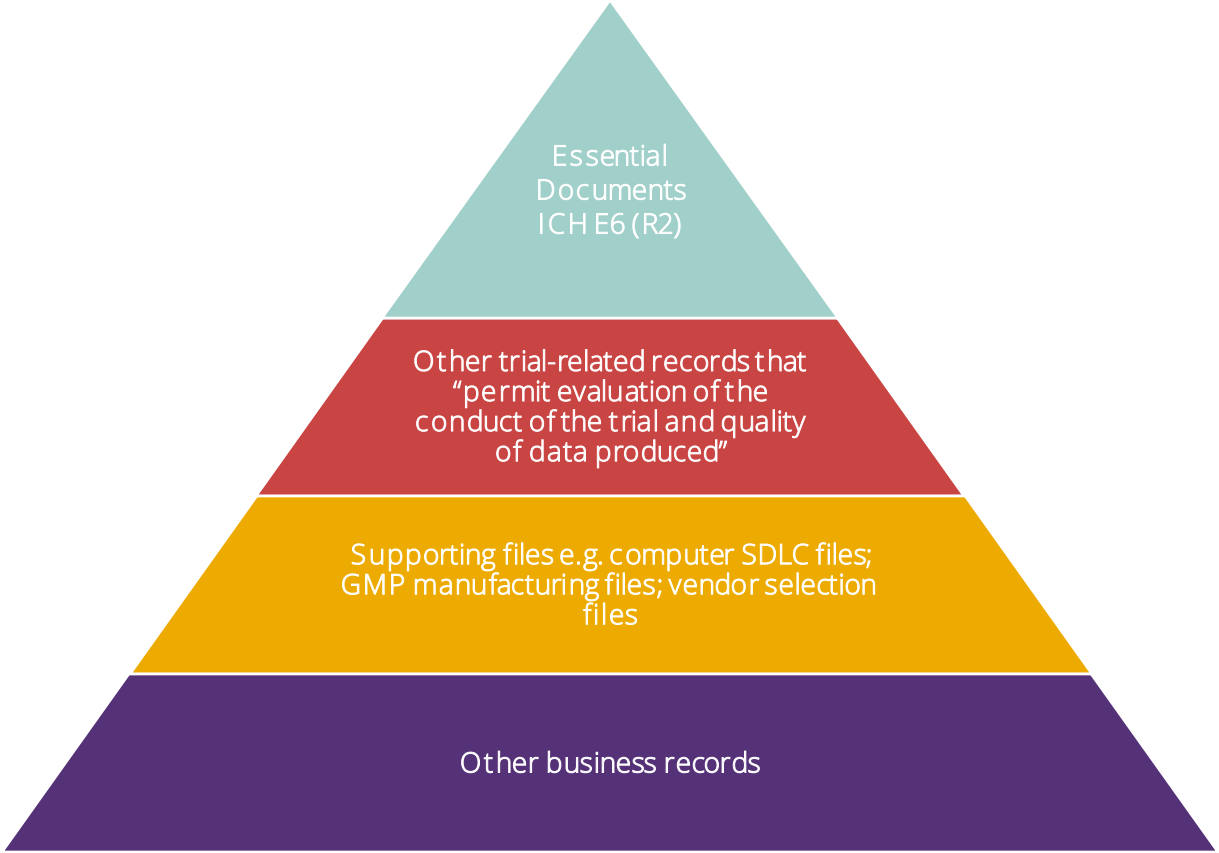
Probability	Harm severity			
	Minor	Marginal	Critical	Catastrophic
Certain	High	High	Very high	Very high
Likely	Medium	High	High	Very high
Possible	Low	Medium	High	Very high
Unlikely	Low	Medium	Medium	High
Rare	Low	Low	Medium	Medium
Eliminated	Eliminated			

Application of LTDP

cdisc

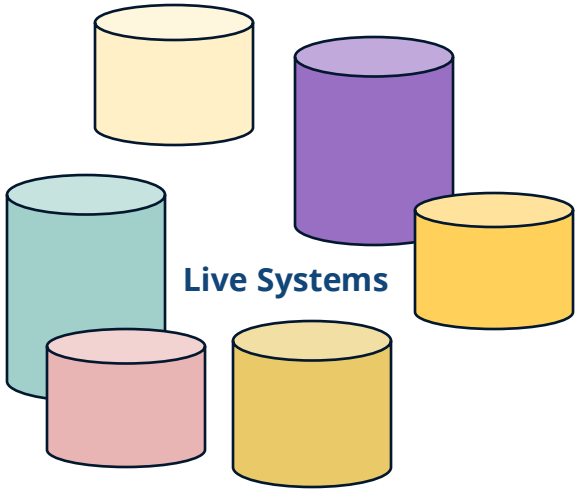
Assets: Data and Systems

- GCP critical
- GCP non-critical
- Other GxP
- Business records
- Retention periods
  - ATMP 30 years
  - MDR 10 years



https://www.diaglobal.org/en/resources/tools-and-downloads#Trial-Master-File

Live Systems, Migrations, Consolidated Archive



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

Migrations are a risk, but also an opportunity to check and document Data Integrity

Data Migration

Consolidated Archive and LTDP Environment

Manage risks in a consistent way
Specialist staff, systems and services
But can't just sit back and hope for the best!

cdisc

Challenges When Trying to Achieve Long Term ALCOA+

Legible and Traceable

- Will documents and data be readable after 30 years?
- Can the audit trail be used to recreate events from 30 years ago?
- Is there documentation so someone can still understand the data?
- Will you be locked into proprietary / legacy formats and software?

Enduring

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

Attributable, Accurate, Contemporaneous

- Can timestamps be altered?
- Is the audit trail permanent?
- Will signatures always validate?
- Can any deletions or changes go unnoticed or unapproved?

Complete, Correct

- Can data integrity issues go undetected during transfers or archiving?
- Can you prove the entire TMF (structure, files, metadata, audit logs) is complete?
- Can you prove migrations (formats, systems, people) were successful?

Available (25 years)

- Can data be discovered easily (metadata)?
- Can data be retrieved quickly (ready access)?
- Is everything documented?
- Will data become spread across lots of 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?

Digital Preservation: Good Practice and Maturity Models

- Good practice: internationally recognised and tested
- Practical and specific things to do in the real world
- Created by organisations with decades of experience
- Covers all the bases
  - Organisation (e.g. people and skills)
  - Resources (e.g. business cases and sustainability)
  - Technology (e.g. systems and processes)

NDSA
Levels of Digital Preservation

https://www.dpconline.org/digipres/dpc-ram

https://www.coretrustseal.org/

https://ndsa.org/publications/levels-of-digital-preservation/

cdisc

Service capabilities		
G	Acquisition, transfer and ingest	Processes to acquire or transfer content and ingest it into a digital archive.
H	Bitstream preservation	Processes to ensure the storage and integrity of digital content to be preserved.
I	Content preservation	Processes to preserve the meaning or functionality of the digital content and ensure its continued accessibility and usability over time.
J	Metadata management	Processes to create and maintain sufficient metadata to support preservation, discovery and use of preserved digital content.
K	Discovery and access	Processes to enable discovery of digital content and provide access for users.

Example 1: Bit Preservation

H - Bitstream preservation Processes to ensure the storage and integrity of digital content to be preserved.	
0 - Minimal awareness	The organization has minimal awareness of either the need for bitstream preservation or basic principles for applying it.
1 – Awareness	The organization is aware of the need for bitstream preservation, and has an understanding of basic principles.
2 – Basic	The organization has implemented a basic process for bitstream preservation, for example: • Dedicated storage is available to meet current preservation needs. • Staff know where content is stored. • Replication is based on simple backup regimes. • Checksums are generated for all content. • There is an understanding of which staff members should be authorized to access the content.
3 – Managed	The organization stores content in a managed way consistent with preservation good practice for replication and integrity checking. For example: • Content is managed with a combination of integrity checking and content replication to one or more locations. • Decisions on the frequency of integrity checking and the number of copies held take into consideration risks, value of the content and costs (both financial and environmental). • Content failing integrity checks is repaired. • Authorizations to access the content by staff are enforced and documented. • Tests are routinely carried out to verify the effectiveness of backups, replication and integrity checking.
4 – Optimized	The organization applies a highly managed storage regime with proactive risk management, for example: • Geographically separated copies are held to minimise the risk of loss due to disaster. • Different storage technologies or services are in use. • Future storage needs are regularly predicted and updated and storage capacity is monitored and revised accordingly. • Content integrity and processes to ascertain integrity are independently reviewed • All access to content is logged and reviewed for unauthorized use and/or changes made: which content when and by whom.

ALCOA++: Enduring and Available

- Store multiple copies of data (files, metadata, audit trails, documentation)
- Different geographically separated locations
- Different types/tiers/classes of storage technology or services
- Immutable data with controlled access and audit trails of any changes
- Initial and periodic data checks (using checksums)
- Storage risk assessment and migration plans
- Automated data replication and monitoring
- BCDR and exit strategy, e.g. if a vendor or system fails

Example 2: Content Preservation

I - Content preservation Processes to preserve the meaning or functionality of the digital content and ensure its continued accessibility and usability over time.	
0 - Minimal awareness	The organization has minimal awareness of either the need for content preservation or basic principles for applying it.
1 – Awareness	The organization is aware of the need for content preservation, and has an understanding of basic principles.
2 – Basic	The organization has implemented a basic process to understand the content that they hold, for example: • File formats are identified. • Content is characterized and assessed for preservation and quality issues such as encrypted, broken or incomplete content and invalid files. • There is a basic understanding of current and future users and use cases for the content.
3 – Managed	The organization has implemented a managed process to monitor and plan for accessibility of content over time, for example: • Technology watch activities are carried out and 'at risk' content is identified. • Technical dependencies are detected and documented. • Actions are occasionally carried out to ensure preservation and quality of content such as migration, emulation or modification of creation or capture workflows. • Preservation actions occur with an understanding of the properties of the digital object that should be retained to support current and future use cases. • All changes to digital content are recorded, including details of when, what, how, why and who.
4 – Optimized	The organization takes a proactive approach to prioritize and mitigate preservation risks to ensure content is accessible over time, for example: • Risks to specific file formats or types of content held are well understood. • A rigorous preservation planning process identifies appropriate preservation actions for risk mitigation. • Decisions on whether to enact preservation actions take into account risks, value of content, costs (both financial and environmental) and use cases. • Format migrations, normalizations, emulation and other preservation actions are implemented in accordance with preservation plans. • Quality control is in place to assess (and record) the meaning and/or functionality of the content has been retained as required. • Digital content and metadata are version co where appropriate.

ALCOA++: Legible and Complete

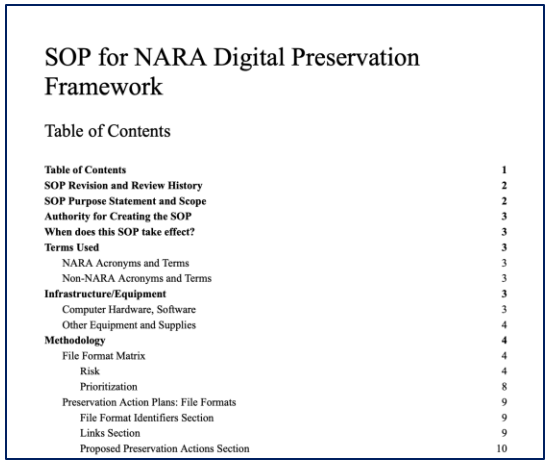
- Document the formats of your data
- Perform Risk Assessment
- File format conversion: migration / normalisation
- Software preservation: emulation, virtualization
- Use CDISC Standards, e.g. eTMF RM and EMS
- Validate format migrations, e.g. dynamic data
- Include LTDP in DMP / eTMF plan

https://github.com/usnationalarchives/digital-preservation/blob/master/Digital_Preservation_Risk_Matrix

https://github.com/usnationalarchives/digital-preservation/blob/master/Digital_Preservation_Risk_Matrix

File Format Risk Assessment

- Perform Risk Assessment
  - Proprietary format?
  - Open specification or standard?
  - Widely adopted? Currently supported?
  - Available tools/applications?
  - Patents/licensing issues?



SOP for NARA Digital Preservation Framework

Table of Contents

Table of Contents 1
SOP Revision and Review History 2
SOP Purpose Statement and Scope 2
Authority for Creating the SOP 3
When does this SOP take effect? 3
Terms Used 3
NARA Acronyms and Terms 3
Non-NARA Acronyms and Terms 3
Infrastructure/Equipment 3
Computer Hardware, Software 3
Other Equipment and Supplies 4
Methodology 4
File Format Matrix 4
Risk 4
Prioritization 8
Preservation Action Plans: File Formats 9
File Format Identifiers Section 9
Links Section 9
Proposed Preservation Actions Section 10

Numeric Risk Rating	Risk Level	NARA Format ID	Format Name	File Extension(s)	Category/Plan(s)
30.00	Low Risk	NF00336	MPEG-1 Program Stream	|m2a|mpa|mpv	Digital Video
-8.00	Low Risk	NF00778	MPEG-2 Program Stream	m2p|mpg|mpeg	Digital Video
26.00	Low Risk	NF00337	MPEG-2 Video	mp2p|x-mpg|x-pn-mpg|x-mpeg|x-	Digital Video
34.00	Low Risk	NF00595	MPEG-4 Advanced Video Coding (H.264)	mp4|mpa	Digital Video
29.00	Low Risk	NF00339	MPEG-4 Media File	mp4|mpa	Digital Video
27.00	Low Risk	NF00393	QuickTime File Format (MOV)	mov	Digital Video
19.00	Moderate Risk	NF00709	Sonic Scenarist Closed Caption Format	scc	Digital Video
31.00	Low Risk	NF00723	SubRip Video Subtitle	srt	Digital Video
36.00	Low Risk	NF00689	Synchronized Accessible Media Interchange	smi|sami	Digital Video
32.00	Low Risk	NF00101	Third Generation Partnership Project (3GPP)	3gp|3gpp	Digital Video
25.00	Low				
7.00	Mo				
7.00	Mo				
7.00	Mo				

Numeric Risk Rating	Risk Level	NARA Format ID	Format Name	File Extension(s)	Category/Plan(s)
-6.00	Moderate Risk	NF00230	Lotus 1-2-3 Worksheet 3.0	wk3	Spreadsheets
-6.00	Moderate Risk	NF00231	Lotus 1-2-3 Worksheet 4.0/5.0	wk4	Spreadsheets
3.00	Moderate Risk	NF00259	Microsoft Excel 2.x	xls	Spreadsheets
18.00	Moderate Risk	NF00260	Microsoft Excel 2000-2003	xls	Spreadsheets
5.00	Moderate Risk	NF00261	Microsoft Excel 3.0	xls	Spreadsheets
10.00	Moderate Risk	NF00262	Microsoft Excel 4.0	xls	Spreadsheets
17.00	Moderate Risk	NF00263	Microsoft Excel 97	xls	Spreadsheets
0.00	Moderate Risk	NF00264	Microsoft Excel Backup	xlk	Spreadsheets
5.00	Moderate Risk	NF00265	Microsoft Excel for Macintosh 2001	xls	Spreadsheets
5.00	Moderate Risk	NF00266	Microsoft Excel for Macintosh 2002	xls	Spreadsheets
5.00	Moderate Risk	NF00267	Microsoft Excel for Macintosh 2004	xls	Spreadsheets
0.00	Moderate Risk	NF00268	Microsoft Excel for Macintosh 3.0	xls	Spreadsheets
5.00	Moderate Risk	NF00269	Microsoft Excel for Macintosh 4.0	xls	Spreadsheets
5.00	Moderate Risk	NF00270	Microsoft Excel for Macintosh 98	xls	Spreadsheets
5.00	Moderate Risk	NF00271	Microsoft Excel for Macintosh v.X	xls	Spreadsheets
21.00	Moderate Risk	NF00665	Microsoft Excel Macro-enabled	xlsm	Spreadsheets
30.00	Low Risk	NF00272	Microsoft Excel Office Open XML	xlsx	Spreadsheets
18.00	Moderate Risk	NF00770	Microsoft Excel Template	xlt	Spreadsheets
4.00	Moderate Risk	NF00656	Microsoft Excel unspecified version	xls	Spreadsheets
2.00	Moderate Risk	NF00273	Microsoft Excel Workspace	xlw	Spreadsheets
-1.00	Moderate Risk	NF00278	Microsoft Multiplan 4.0	mod	Spreadsheets
39.00	Low Risk	NF00349	OpenDocument Spreadsheet 1.0	ods|fods|ots	Spreadsheets
39.00	Low Risk	NF00513	OpenDocument Spreadsheet 1.1	ods|fods|ots	Spreadsheets
41.00	Low Risk	NF00514	OpenDocument Spreadsheet 1.2	ods|fods|ots	Spreadsheets
45.00	Low Risk	NF00800	OpenDocument Spreadsheet 1.3	ods|fods|ots	Spreadsheets

Numeric Risk Rating	Risk Level	NARA Format ID	For		
-22.00	High Risk	NF00476	Sig		
-26.00	High Risk	NF00149	Cor		
-20.00	High Risk	NF00435	Vis		
-19.00	High Risk	NF00820	Bey		
-18.00	High Risk	NF00815	Apple Soundtrack Pro Project	stmp	Digital Audio
-19.00	High Risk	NF00477	Sony RAW	srf	Digital Still Image
-34.00	High Risk	NF00407	Sound Designer II Audio File	sd2	Digital Audio
-13.00	High Risk	NF00405	Scitex CT Image	sct|ct|ch	Digital Still Image
-32.00	High Risk	NF00726	Rocket Book eBook format	rb	Presentation and Publis
-21.00	High Risk	NF00470	FujiFilm RAW	raf	Digital Still Image
-17.00	High Risk	NF00812	Pixel Image File	px	Digital Still Image
-18.00	High Risk	NF00817	Avid Pro Tools Session 5.1-6.9	pts	Digital Audio
-14.00	High Risk	NF00354	PaintShop Pro Image version 2	psp	Digital Still Image
-14.00	High Risk	NF00516	PaintShop Pro Image version 3	psp	Digital Still Image
-14.00	High Risk	NF00517	PaintShop Pro Image version 4	psp	Digital Still Image

Summary

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			

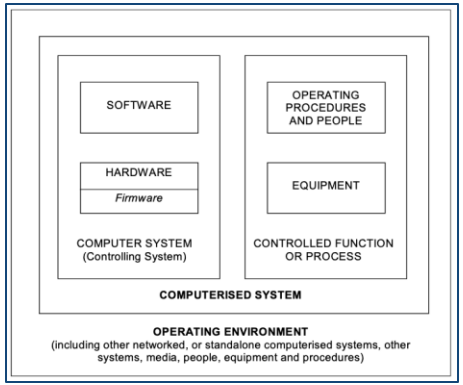
ALCOA+ Principles

- A – Attributable
- L – Legible
- C – Contemporary
- O – Original
- A – Accurate
- C – Complete
- C – Correct
- E – Enduring
- A – Available

NDSA Levels of Digital Preservation

CORE TRUST SEAL

1. Categorise the criticality of your data and the systems it resides within
2. Assess your risks based on long-term GxP Data Integrity requirements
3. Implement proportionate and appropriate LTDP good practice
4. Validated archiving and preservation solutions from qualified suppliers



https://picscheme.org/docview/3444

cdisc

Thank You!