A Risk-based Approach to long Term Retention and Access of Electronic Documents and Records

Matthew Addis

Arkivum Ltd

Matthew.addis@arkivum.com

www.diahome.org

Arkivum

Every bit archived

Disclaimer

The views and opinions expressed in the following PowerPoint slides are those of the individual presenter and should not be attributed to Drug Information Association, Inc. (“DIA”), its directors, officers, employees, volunteers, members, chapters, councils, Special Interest Area Communities or affiliates, or any organization with which the presenter is employed or affiliated.

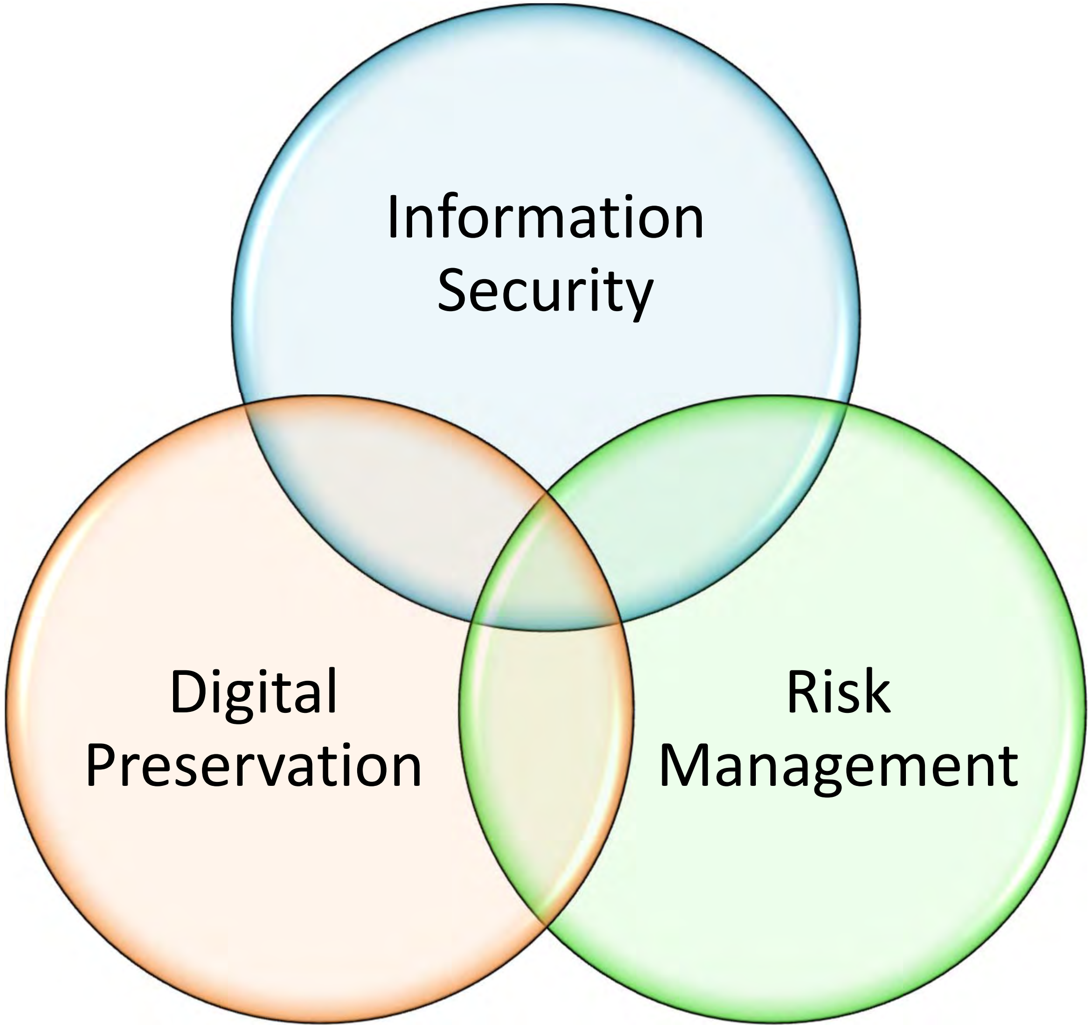
These PowerPoint slides are the intellectual property of the individual presenter and are protected under the copyright laws of the United States of America and other countries. Used by permission. All rights reserved. Drug Information Association, DIA and DIA logo are registered trademarks or trademarks of Drug Information Association Inc. All other trademarks are the property of their respective owners.

http://www.youtube.com/watch?v=pbBa6Oam7-w

Objectives

- Integrity and authenticity
- Confidentiality
- Usability / ready access / readability
- Responsibility
- Risk management

Three approaches



Information Security

Digital Preservation

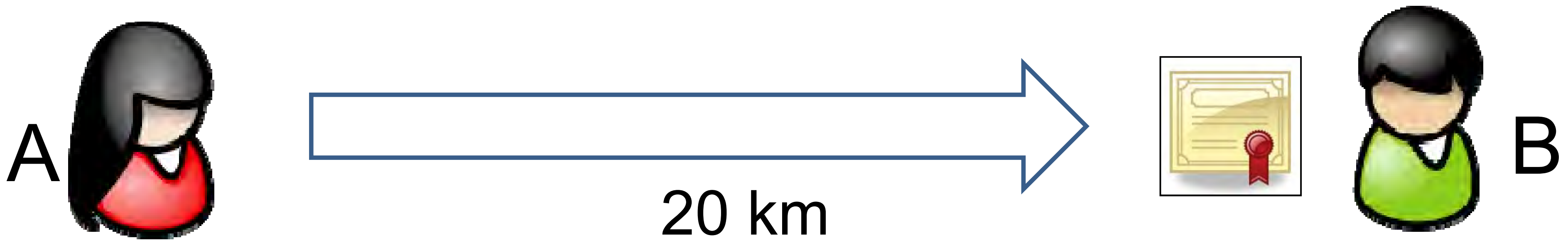
Risk Management

Getting from A to B

- Sending information from A → B
  - Authentication
  - Confidentiality
  - Integrity
  - Availability

→ Risk Management

Information Security

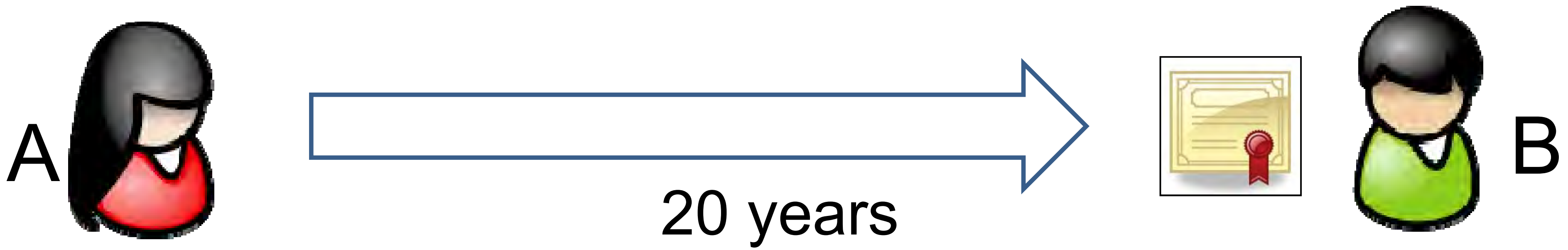


Communicating with the future

- Sending information through time
  - Readability
  - Integrity and authenticity
  - Accessibility
  - Safety and security

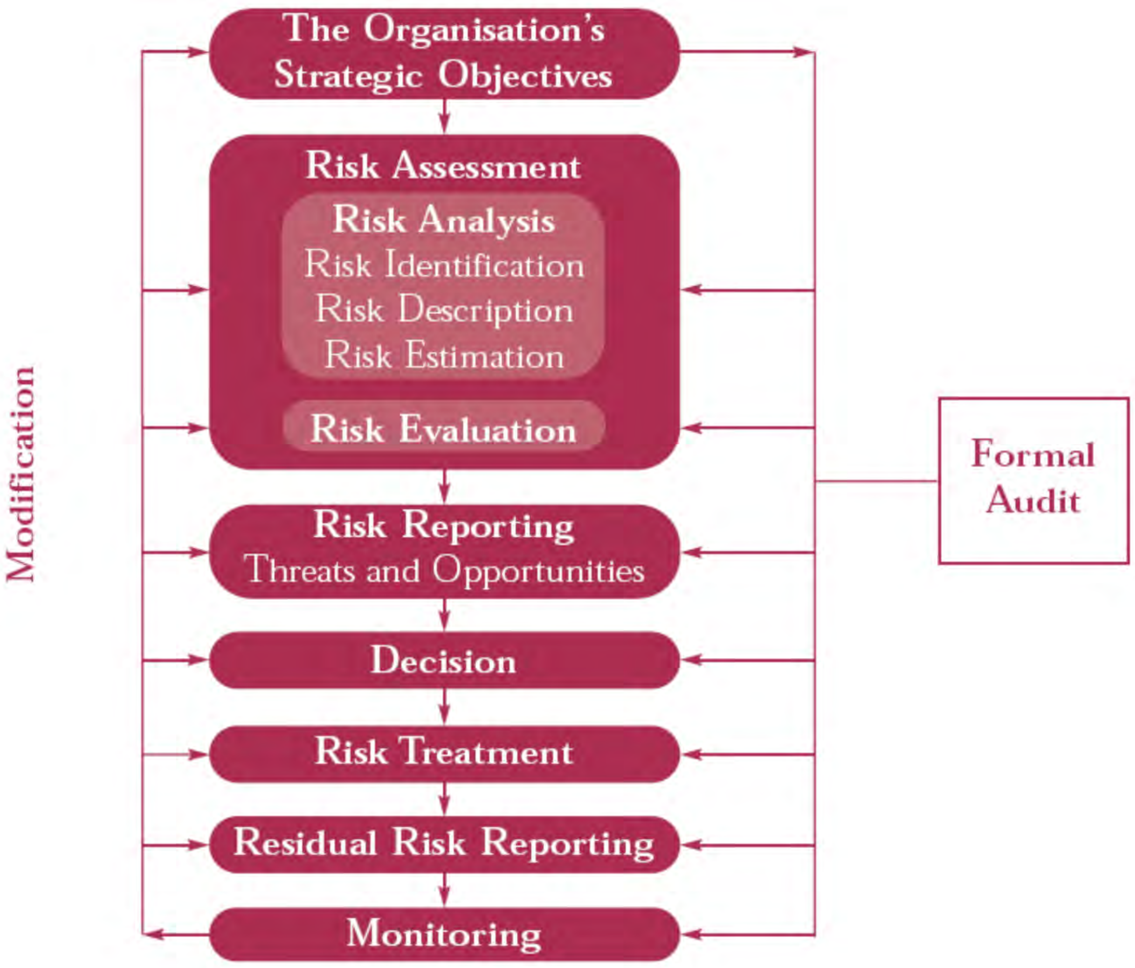
→ Risk Management

Digital Preservation



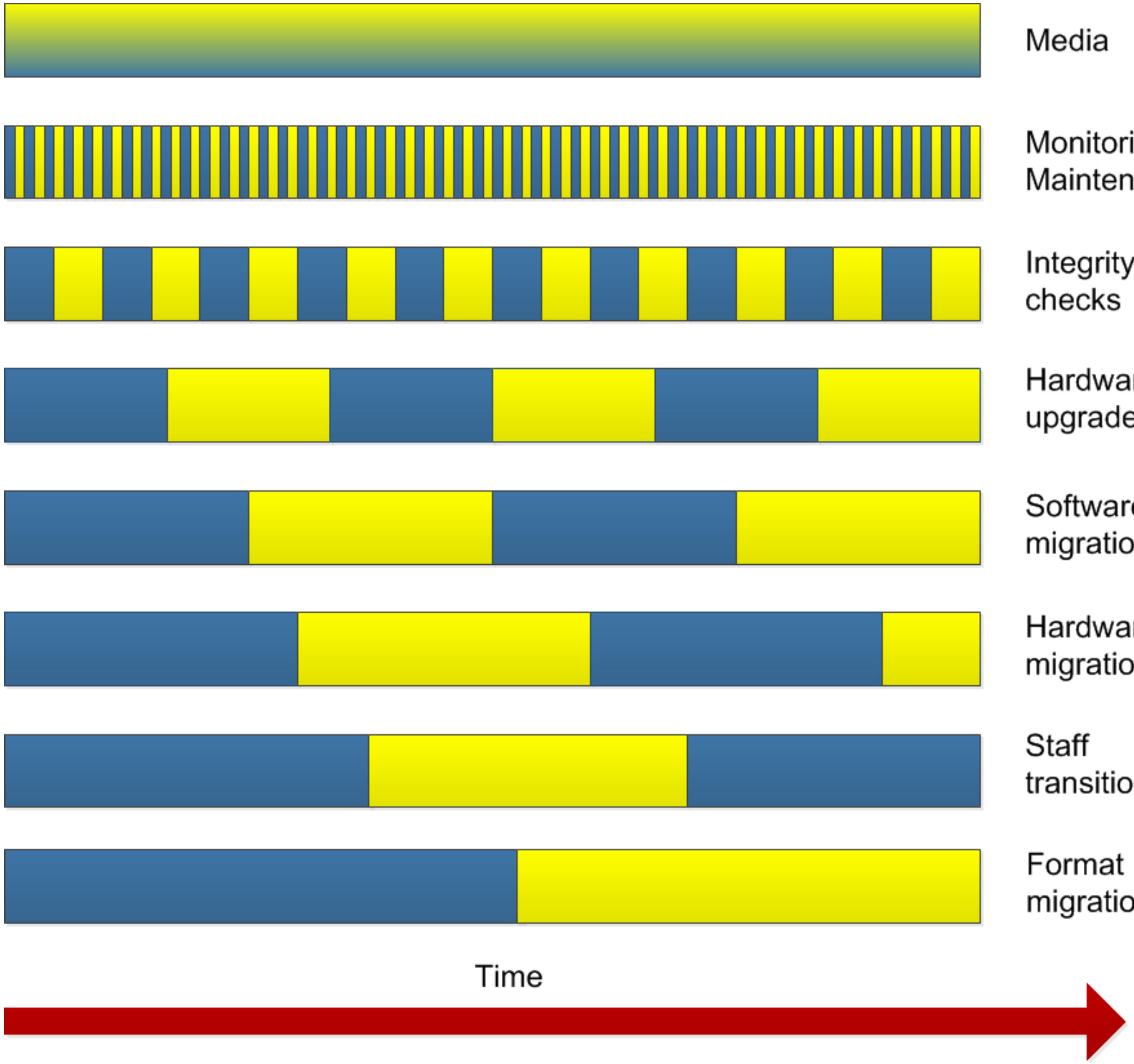
All roads lead to risk management

- Assets
- Value
- Threats
- Impact
- Treatment
- Management



CHALLENGES

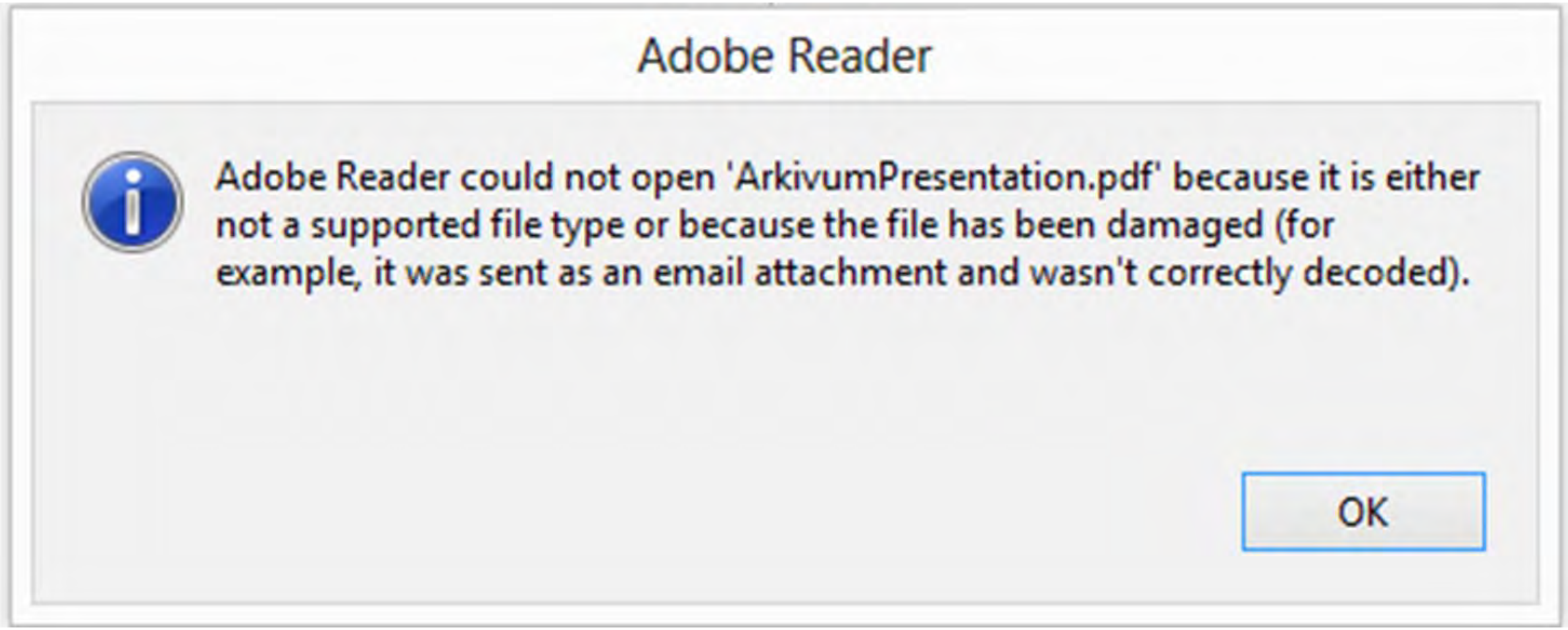
20 years of keeping content alive



Digital preservation

“Digital information lasts forever -
or five years, whichever comes first."
Jeff Rothenberg

Preservation =
No Continuity Failures



Adobe Reader

Adobe Reader could not open 'ArkivumPresentation.pdf' because it is either not a supported file type or because the file has been damaged (for example, it was sent as an email attachment and wasn't correctly decoded).

YOU
ARE
FIRED

Continuity

- Continuity costs money
- Continuity takes time
- Continuity introduces risk
- Continuity requires validation
- Continuity needs planning
- Continuity needs management

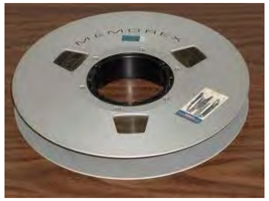
Obsolescence

30,000 years

2,200 years

1,200 years

200 years



20 years

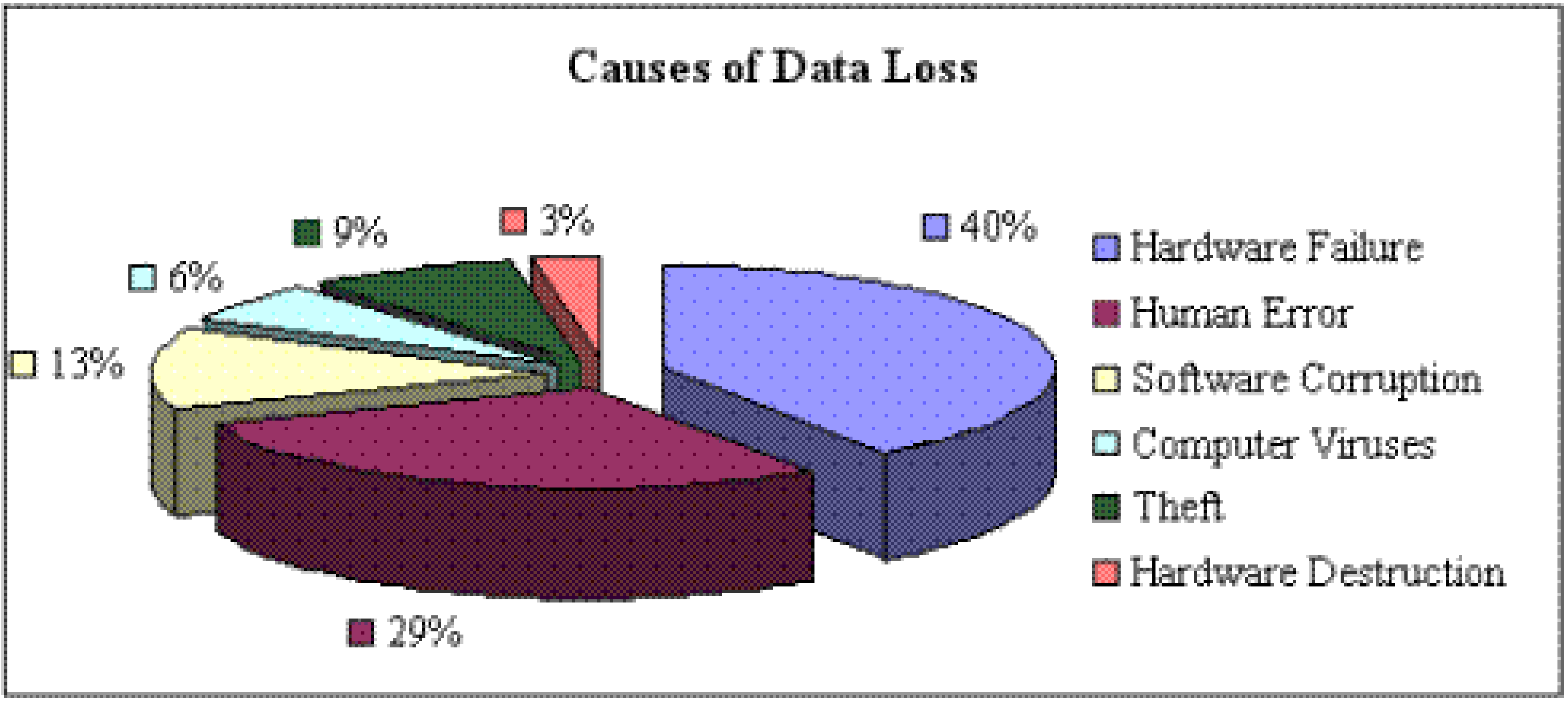
Each new generation of technology:

1000x times denser
lasts 1/10th as long

IT storage is not 100% safe

Examples of damaged discs.

People can cause data loss too

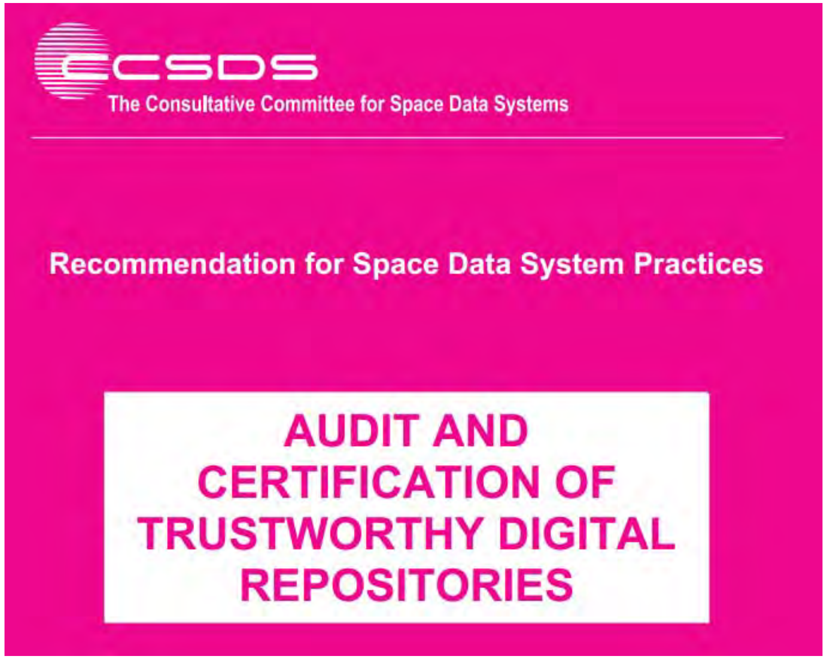


There's plenty more ways to lose data!

- Technical obsolescence, e.g. formats and players
- Hardware failures, e.g. digital storage systems
- Loss of staff, e.g. skilled transfer operators
- Insufficient budget, e.g. digitisation too expensive
- Accidental loss, e.g. human error during QC
- Stakeholders, e.g. preservation no longer a priority
- Underestimation of resources or effort
- Fire, flood, meteors, aliens…

SOLUTIONS

Digital preservation standards

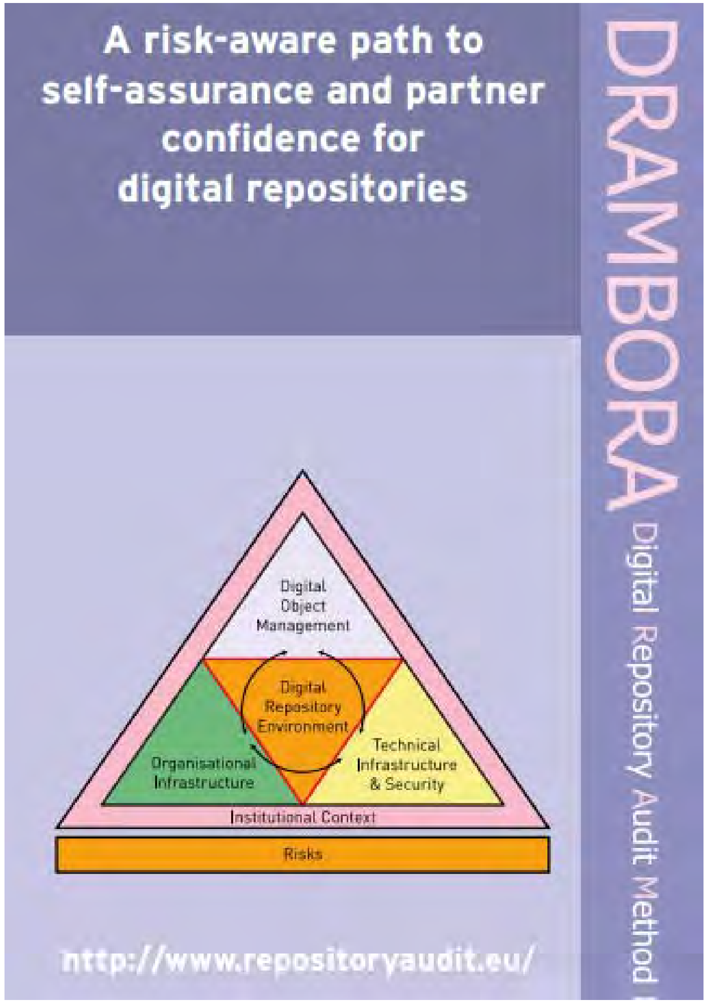


CCSDS

The Consultative Committee for Space Data Systems

Recommendation for Space Data System Practices

AUDIT AND CERTIFICATION OF TRUSTWORTHY DIGITAL REPOSITORIES



A risk-aware path to self-assurance and partner confidence for digital repositories

Digital Object Management
Digital Repository Environment
Organisational Infrastructure
Technical Infrastructure & Security
Institutional Context
Risks

http://www.repositoryaudit.eu/

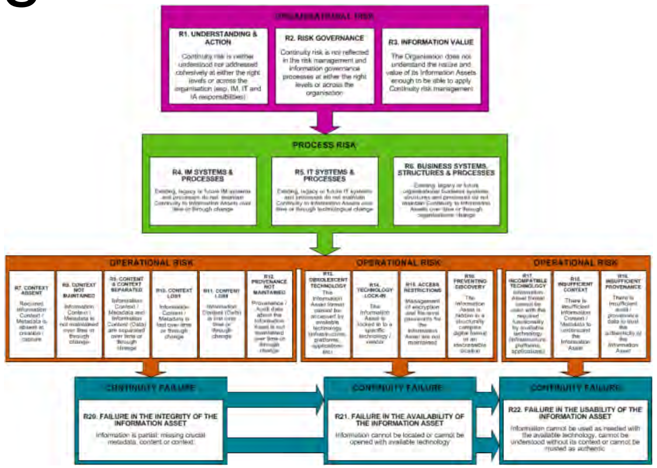
DRAMBORA Digital Repository Audit Method

ISO/IEC 27001:2005

Others have trodden the road already

- Organisational risks
  - Understanding, governance, value
- Process risks
  - Business, IT and IM processes
- Operational risks
  - Technology failure, lock-in, obsolescence
- Continuity risks
  - Failure in integrity, availability, usability

The National Archives



ORGANISATIONAL RISK

R1. UNDERSTANDING & ACTION
Continuity risk is neither understood nor addressed cohesively at either the right levels or across the organisation (esp. IM, IT and IA responsibilities)

R2. RISK GOVERNANCE
Continuity risk is not reflected in the risk management and information governance processes at either the right levels or across the organisation

R3. INFORMATION VALUE
The Organisation does not understand the nature and value of its Information Assets enough to be able to apply Continuity risk management

PROCESS RISK

R4. IM SYSTEMS & PROCESSES
Existing, legacy or future IM systems and processes do not maintain Continuity to Information Assets over time or through change

R5. IT SYSTEMS & PROCESSES
Existing, legacy or future IT systems and processes do not maintain Continuity to Information Assets over time or through technological change

R6. BUSINESS SYSTEMS, STRUCTURES & PROCESSES
Existing, legacy or future organisational business systems, structures and processes do not maintain Continuity to Information Assets over time or through organisational change

OPERATIONAL RISK

R7. CONTEXT ABSENT
Required Information Context / Metadata is absent at creation / capture

R8. CONTEXT NOT MAINTAINED
Information Context / Metadata is not maintained over time or through change

R9. CONTENT & CONTEXT SEPARATED
Information Context / Metadata and Information Content (Data) are separated over time or through change

R10. CONTEXT LOST
Information Context / Metadata is lost over time or through change

R11. CONTENT LOST
Information Content (Data) is lost over time or through change

R12. PROVENANCE NOT MAINTAINED
Provenance / Audit data about the Information Asset is not maintained over time or through change

OPERATIONAL RISK

R13. OBSOLESCENT TECHNOLOGY
The Information Asset format cannot be accessed by available technology (infrastructure, platforms, applications etc.)

R14. TECHNOLOGY LOCK-IN
The Information Asset is locked in to a specific technology / vendor

R15. ACCESS RESTRICTIONS
Management of encryption and file-level passwords for the Information Asset are not maintained

R16. PREVENTING DISCOVERY
The Information Asset is hidden in a structurally complex digital format or an inaccessible location

OPERATIONAL RISK

R17. INCOMPATIBLE TECHNOLOGY
Information Asset format cannot be used with the required functionality by available technology (infrastructure, platforms, applications)

R18. INSUFFICIENT CONTEXT
There is insufficient Information Context / Metadata to understand the Information Asset

R19. INSUFFICIENT PROVENANCE
There is insufficient audit / provenance data to trust the authenticity of the Information Asset

CONTINUITY FAILURE

R20. FAILURE IN THE INTEGRITY OF THE INFORMATION ASSET
Information is partial: missing crucial metadata, content or context.

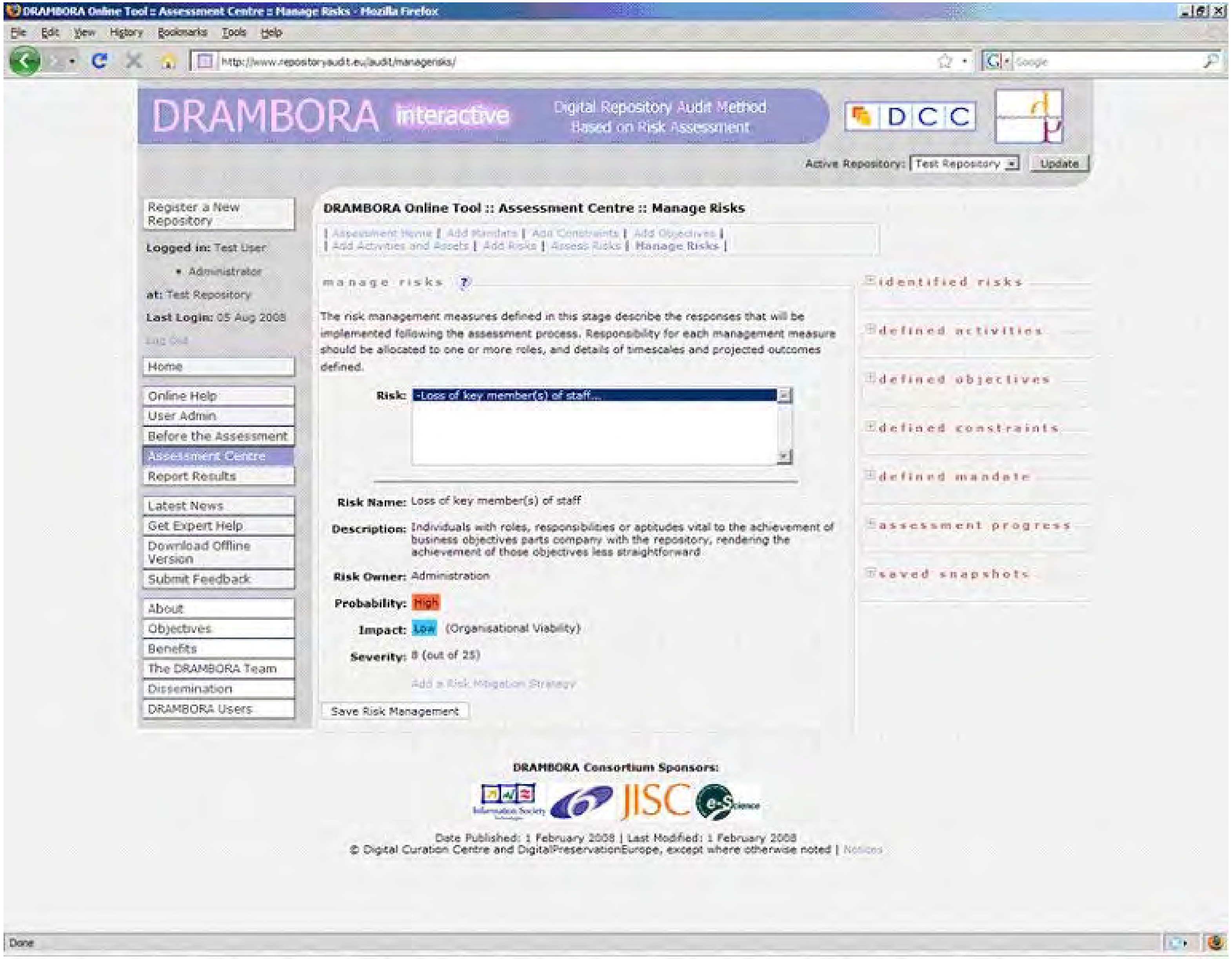
CONTINUITY FAILURE

R21. FAILURE IN THE AVAILABILITY OF THE INFORMATION ASSET
Information cannot be located or cannot be opened with available technology

CONTINUITY FAILURE

R22. FAILURE IN THE USABILITY OF THE INFORMATION ASSET
Information cannot be used as needed with the available technology, cannot be understood without its context or cannot be trusted as authentic

DRAMBORA



DRAMBORA Online Tool :: Assessment Centre :: Manage Risks

manage risks ?

The risk management measures defined in this stage describe the responses that will be implemented following the assessment process. Responsibility for each management measure should be allocated to one or more roles, and details of timescales and projected outcomes defined.

Risk: -Loss of key member(s) of staff...

Risk Name: Loss of key member(s) of staff

Description: Individuals with roles, responsibilities or aptitudes vital to the achievement of business objectives parts company with the repository, rendering the achievement of those objectives less straightforward

Risk Owner: Administration

Probability: High

Impact: Low (Organisational Viability)

Severity: 8 (out of 25)

DRAMBORA Consortium Sponsors:

Date Published: 1 February 2008 | Last Modified: 1 February 2008
© Digital Curation Centre and DigitalPreservationEurope, except where otherwise noted | Notices

Example risks

Risk ID	Title	Example
R30	Hardware Failure	A storage system corrupts files (bit rot) or loses data due to component failures (e.g. hard drives).
R31	Software Failure	A software upgrade to the system looses or corrupts the index used to locate files.
R32	Systems fail to meet archive needs	The system can't cope with the data volumes and the backups fail.
R33	Obsolescence of hardware or software	A manufacturer stops support for a tape drive and there is insufficient head life left in existing drives owned by the archive to allow migration
R34	Media degradation or obsolescence	The BluRay optical discs used to store XDCAM files develop data loss.
R35-R38	Security	Insufficient security measures allow unauthorised access that results undetected modification of files.

Prioritising risks

	Impact		
Likelihood	Major	Moderate	Minor
Likely	🟥	🟥	🟧
Possible	🟥	🟧	🟩
Unlikely	🟧	🟩	🟩

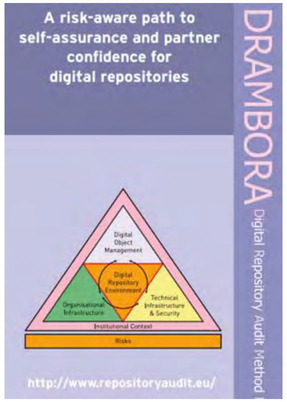
ISO 16363 (Trusted Digital Repositories) 2012

- Cause
  - e.g. failure to maintain storage systems
- What is affected
  - e.g. eTMF, audit trail, contracts
- Consequence
  - e.g. fines, loss of reputation, recreate content
- Priority
  - e.g. high
- Best practice
  - e.g. technology watch, migration planning
- Mitigate, avoid, accept, transfer
  - e.g. regular tests and migrations, multiple copies of data

CCSDS
The Consultative Committee for Space Data Systems

Recommendation for Space Data System Practices

AUDIT AND CERTIFICATION OF TRUSTWORTHY DIGITAL REPOSITORIES



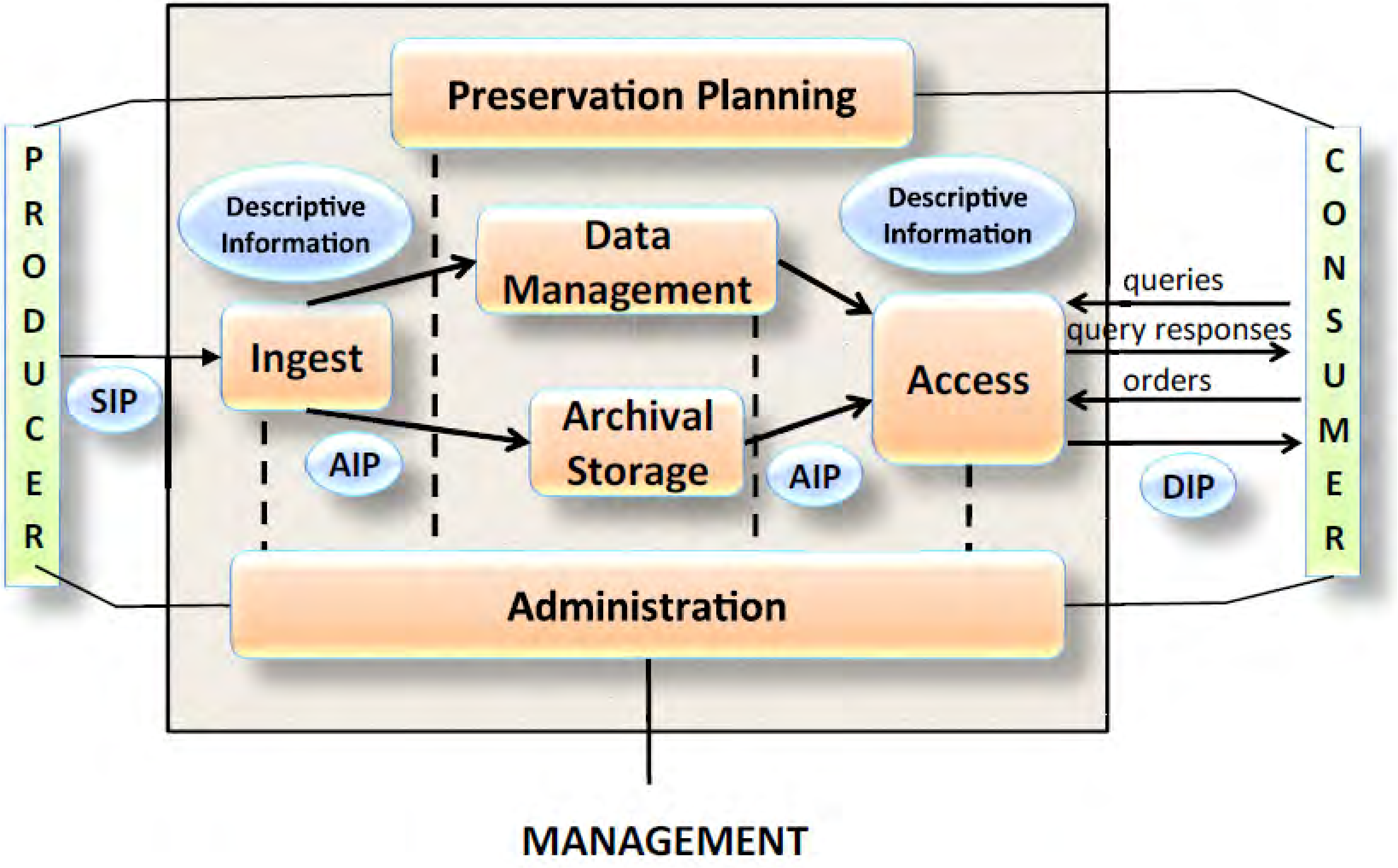
A risk-aware path to self-assurance and partner confidence for digital repositories

Digital Object Management
Digital Repository Environment
Organisational Infrastructure
Technical Infrastructure & Security
Institutional Context
Risks

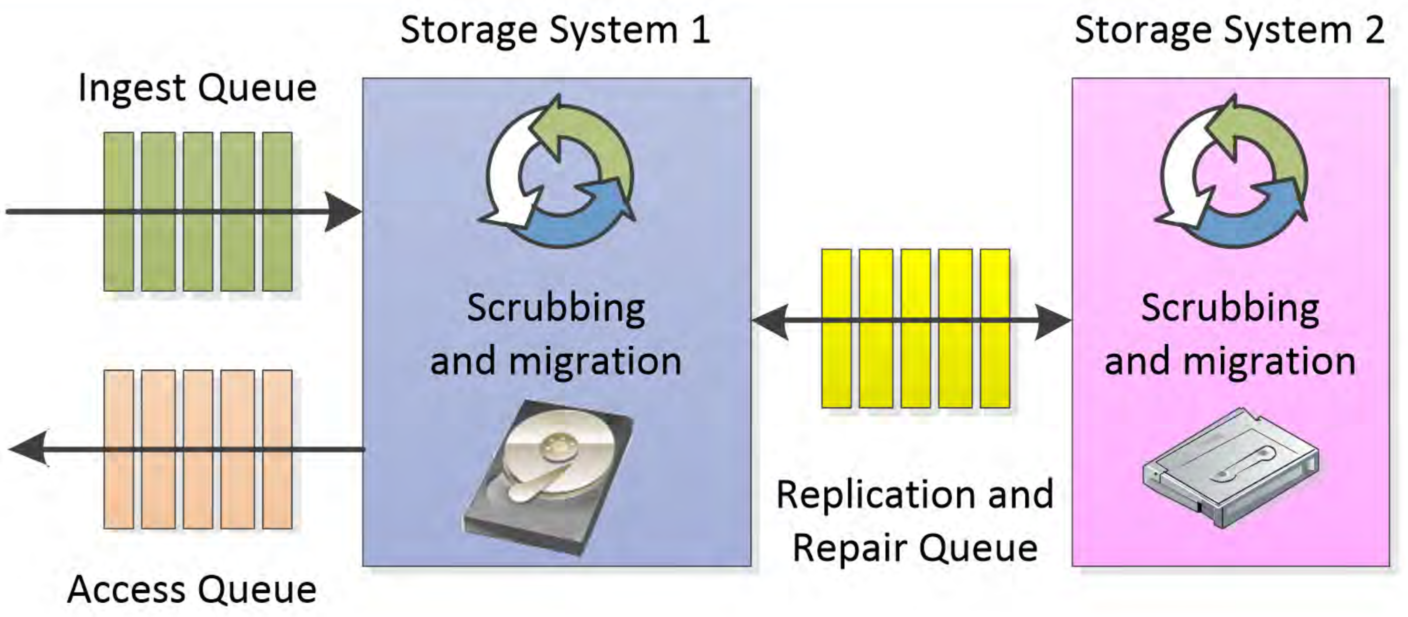
http://www.repositoryaudit.eu/

DRAMBORA Digital Repository Audit Method

ISO 14721 (Open Archival Information System) 2012

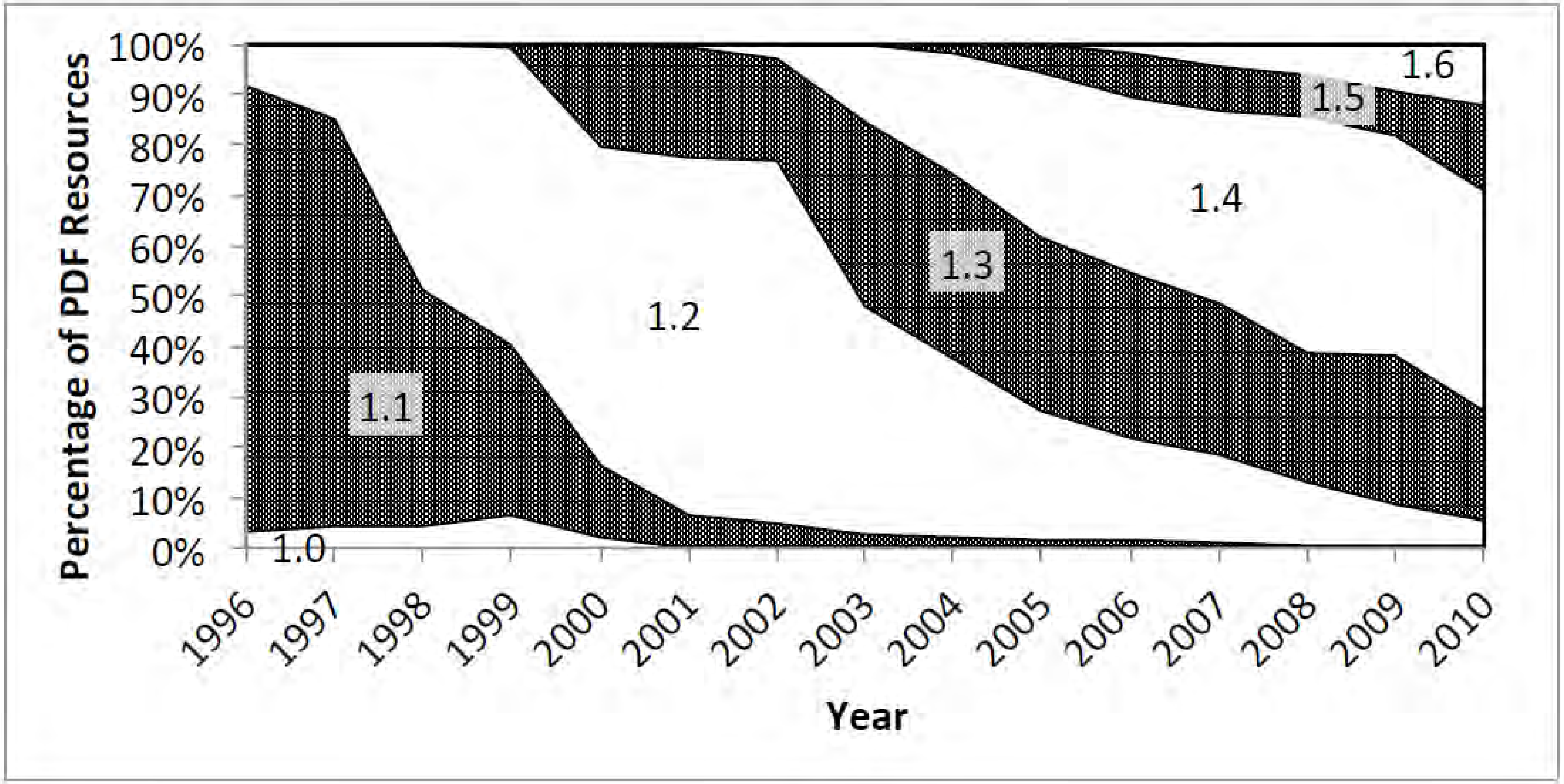


Active archiving



- Preservation best practice (diversity, intervention)
  - – Multiple copies in different locations
  - – Different technologies and different people
  - – Active management: migration, integrity

Format obsolescence and proliferation



File formats and migration

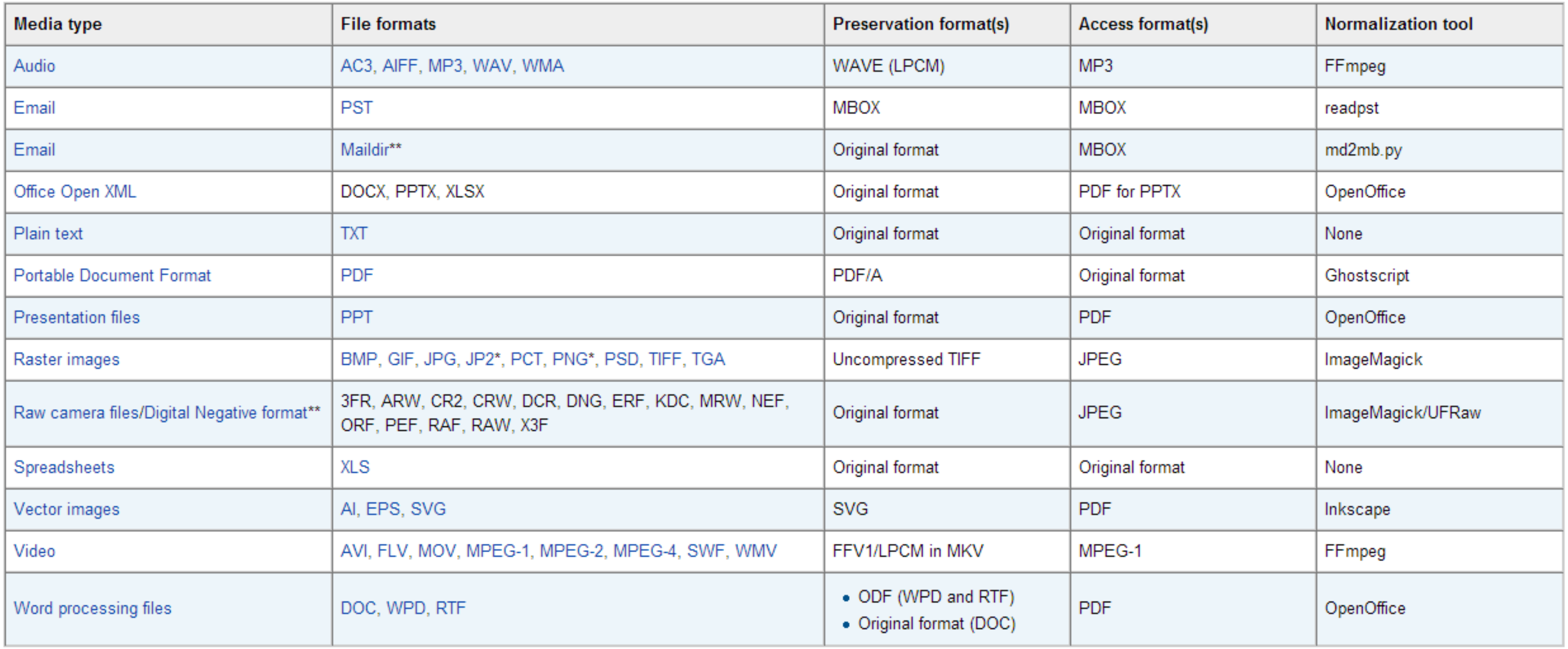
www.diahome.org

Droid

DROID (Digital Record and Object Identification)

The technical registry

PRONOM



Media type	File formats	Preservation format(s)	Access format(s)	Normalization tool
Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Email	PST	MBOX	MBOX	readpst
Email	Maildir**	Original format	MBOX	md2mb.py
Office Open XML	DOCX, PPTX, XLSX	Original format	PDF for PPTX	OpenOffice
Plain text	TXT	Original format	Original format	None
Portable Document Format	PDF	PDF/A	Original format	Ghostscript
Presentation files	PPT	Original format	PDF	OpenOffice
Raster images	BMP, GIF, JPG, JP2*, PCT, PNG*, PSD, TIFF, TGA	Uncompressed TIFF	JPEG	ImageMagick
Raw camera files/Digital Negative format**	3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F	Original format	JPEG	ImageMagick/UFRaw
Spreadsheets	XLS	Original format	Original format	None
Vector images	AI, EPS, SVG	SVG	PDF	Inkscape
Video	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV	FFV1/LPCM in MKV	MPEG-1	FFmpeg
Word processing files	DOC, WPD, RTF	• ODF (WPD and RTF) • Original format (DOC)	PDF	OpenOffice

Metadata standards

ISO 23081-1:2006

Information and documentation -- Records management processes -- Metadata for records -- Part 1: Principles

MoReq2010®

Modular Requirements for Records Systems

PREMIS

PRESERVATION METADATA
MAINTENANCE ACTIVITY

METS

Metadata Encoding & Transmission Standard

Official Web Site

SIMPLE STRATEGY

1. Assess and manage the risks

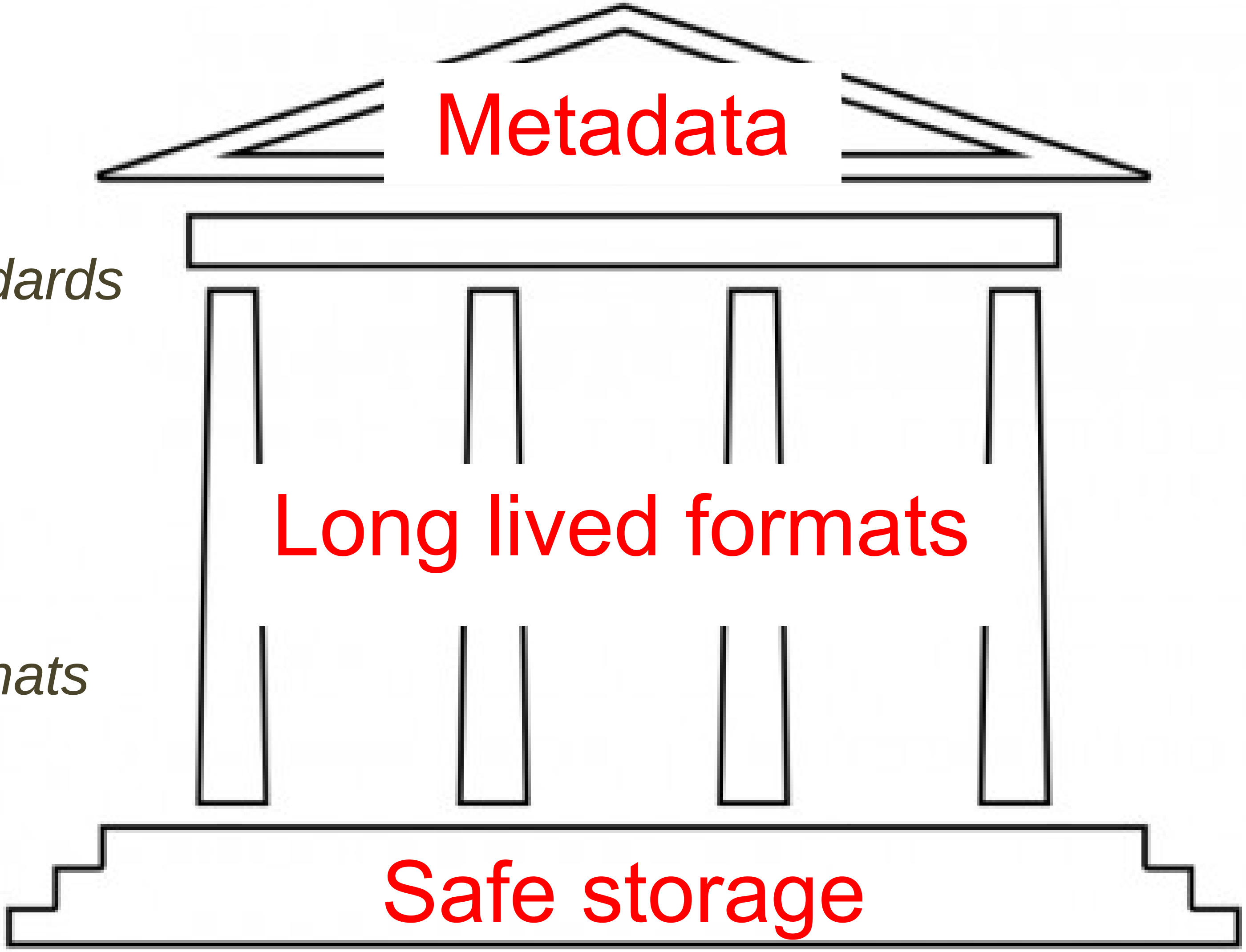
• ISO27001	Information Security Management
• ISO16363	Trusted Digital Repositories
• DSA	Data Seal of Approval

2. Start at the bottom and build up

Descriptive
Technical
Preservation
Rights
Metadata Standards

Creation
Preservation
Access
Canonical Formats

Optimise
for archive

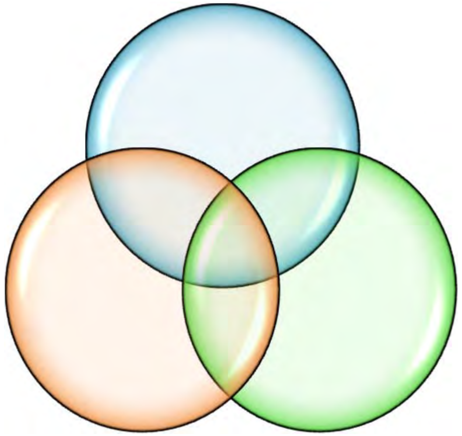


3. Establish a burn-line

- Metadata and data on a file system
- Open standards and formats
- Drive down costs and risks

Summary

- Information Security
- Digital Preservation
- Risk Management



- OAIS, TDR, PREMIS, MoReq, BagIt…
- People, Processes, Infrastructure
- Active management, validation, audit