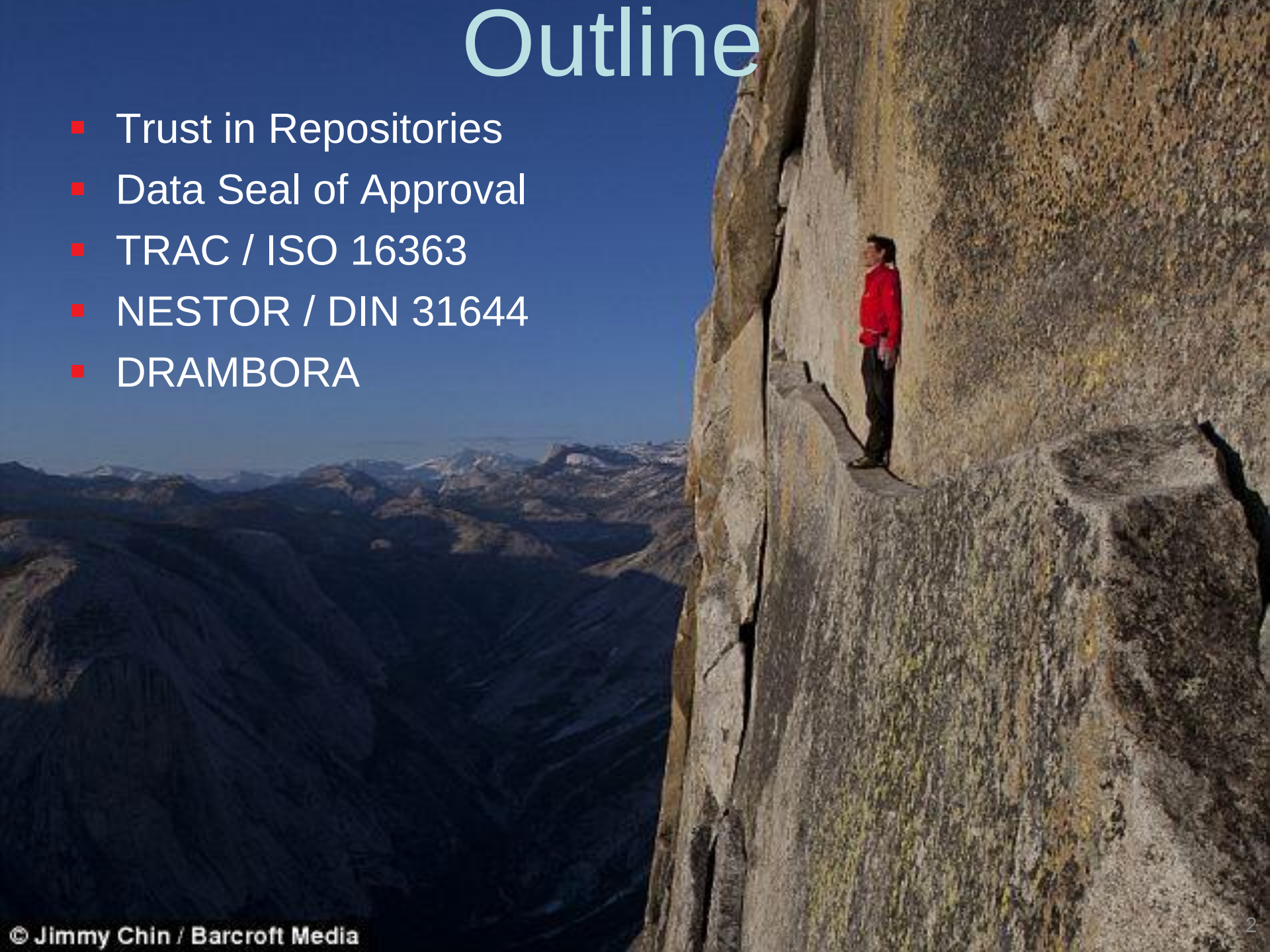
Trust, Audit and Certification

Stefan Pröll

Vienna University of Technology

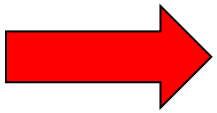
Outline

- Trust in Repositories
- Data Seal of Approval
- TRAC / ISO 16363
- NESTOR / DIN 31644
- DRAMBORA



Trust

- Repositories store sensitive information
- How can we trust an archive?
  - Minimum set of requirements
- Trustworthiness: How to
  - Establish?
  - Maintain?
  - Verify?



Audit and Certification

Audits

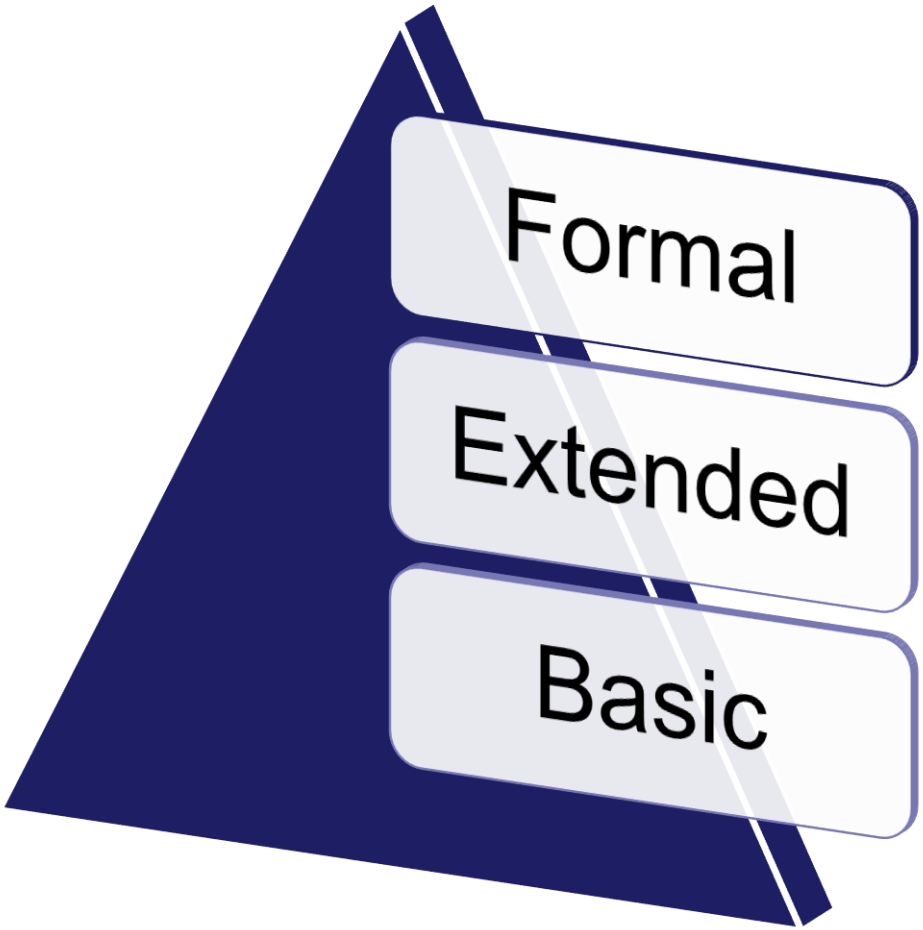
- An audit is a formal review or examination
  - Independent
  - External or internal
- Why should repositories undergo audits?
  - Review own processes and methods
  - Detect strengths, weaknesses and ga
  - Approve them by external professiona
  - Get feedback
  - Improve your repository
  - Demonstrate transparency
  - Show off with a certificate

http://kgi.org/blog/roger-doiron/school-garden-checklist

European Framework for Audit and Certification of Digital Repositories

- Three certification levels
  - Basic level:
    - Data Seal of Approval (DSA)
    - Self assessment based on DSA
  - Extended certification
    - Self assessment of DSA and either ISO 16363 or DIN 31644
  - Formal certification
    - Self assessment of DSA and full external audit of either ISO 16363 or DIN 31644

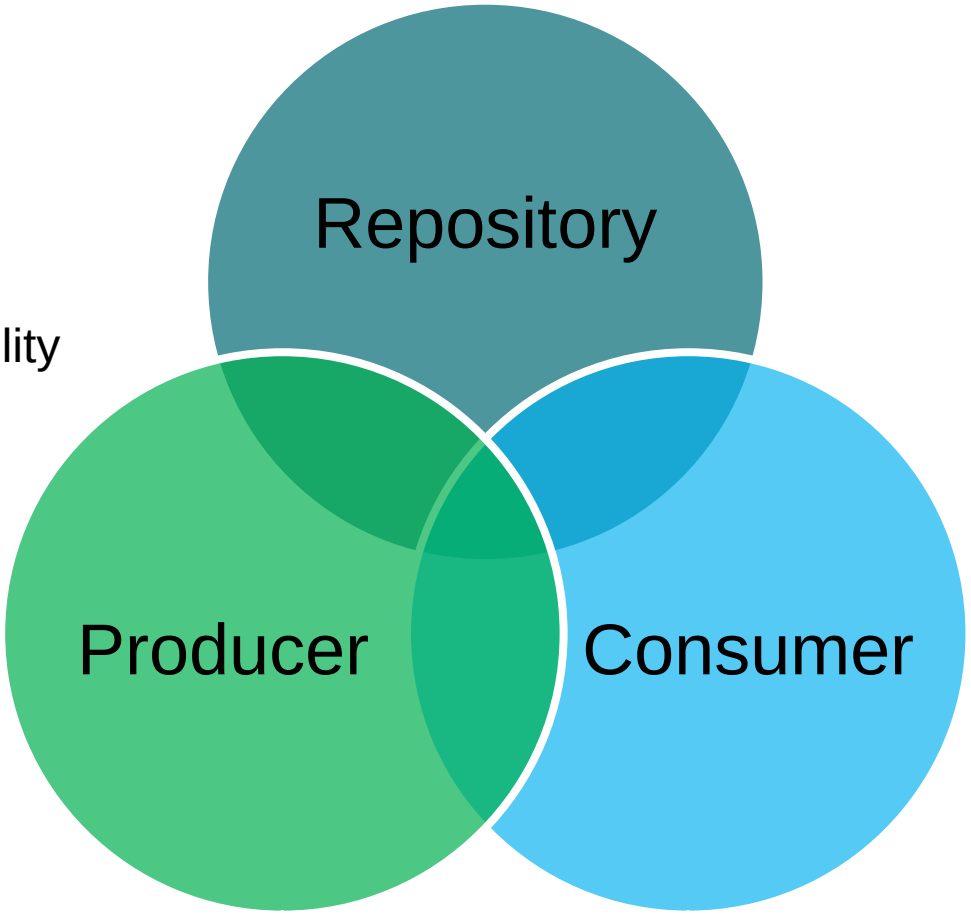
Types of Audits



Memorandum of Understanding: www.trusteddigitalrepository.eu

Stakeholders

- Producer
  - Responsible for the data quality
- Repository:
  - Data storage quality and availability
- Consumer
  - Quality of use



Data Seal of Approval

- 16 guidelines: http://datasealofapproval.org/en/information/guidelines/

- If all are followed → Data Seal of Approval
- www.datasealofapproval.org
- Complete self assessment
- Tool support available
- Peer review process
- Periodical renewal

DSA - Guidelines Relating to Data **Producers**

- 1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data and compliance with disciplinary and ethical norms.
- 2. The data producer provides the data in formats recommended by the data repository.
- 3. The data producer provides the data together with the metadata requested by the data repository.

Guidelines Related to **Repositories** 1

- 4. The data repository has an explicit mission in the area of digital archiving and promulgates it.
- 5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
- 6. The data repository applies documented processes and procedures for managing data storage.

DSA: Guidelines Related to **Repositories** 2

- 7. The data repository has a plan for long-term preservation of its digital assets.
- 8. Archiving takes place according to explicit work flows across the data life cycle.
- 9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.
- 10. The data repository enables the users to discover and use the data and refer to them in a persistent way.

DSA: Guidelines Related to Repositories 3

- 11. The data repository ensures the integrity of the digital objects and the metadata.
- 12. The data repository ensures the authenticity of the digital objects and the metadata.
- 13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.

DSA - Guidelines Related to Data **Consumers**

- 14. The data consumer complies with access regulations set by the data repository.
- 15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.
- 16. The data consumer respects the applicable licences of the data repository regarding the use of the data.

ISO 16363 / TRAC

- Trustworthy Repositories Audit & Certification
- ISO 16363 is based on TRAC
  - TRAC had focus on self audits of digital libraries
  - ISO 16363 supports full external audit of all kinds of repositories
- 109 criteria
- Was suggested as follow-on standard by OAIS (ISO 14721)
- The second standard ISO 16919 “Requirements for Bodies providing Audit and Certification” describes how to govern audit process.

ISO 16363 / TRAC

- ISO 16363 audit:
  - Includes preparation period
  - A visit on-site from the audit team
  - Audit team creates formal report
  - If the repository passed the audit it receives the certificate
  - The certificate needs to be periodically renewed
- Magenta book: http://public.ccsds.org/publications/archive/652x0m1.pdf
  - CCSDS 652.0-M-1 is equivalent to ISO 16363

ISO 16363 / TRAC

- Structure of the standard: Three pilars
  - Organizational infrastructure
  - Digital object management
  - Infrastructure and security management
- Each section provides metrics which can be measured and evaluated
- Each metric has an associated description of the evidence which is needed for compliance.

ISO 16363 / TRAC

- Governance and organizational viability

3 ORGANIZATIONAL INFRASTRUCTURE

3.1 GOVERNANCE AND ORGANIZATIONAL VIABILITY

Metric

3.1.1 The repository shall have a mission statement that reflects a commitment to the preservation of, long term retention of, management of, and access to digital information.

Description

Supporting Text

This is necessary in order to ensure commitment to preservation, retention, management and access at the repository's highest administrative level.

Evidence example

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Mission statement or charter of the repository or its parent organization that specifically addresses or implicitly calls for the preservation of information and/or other resources under its purview; a legal, statutory, or government regulatory mandate applicable to the repository that specifically addresses or implicitly requires the preservation, retention, management and access to information and/or other resources under its purview.

ISO 16363 / TRAC

- TRAC and Preservation Planning 1:

3.3.2.1 The repository shall have mechanisms for review, update, and ongoing development of its Preservation Policies as the repository grows and as technology and community practice evolve.

Supporting Text

This is necessary in order that the repository has up-to-date, complete policies and procedures in place that reflect the current requirements and practices of its community(ies) for preservation.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Current and past written documentation in the form of Preservation Policies, Preservation Strategic Plans and Preservation Implementation Plans, procedures, protocols, and workflows; specifications of review cycles for documentation; documentation detailing reviews, surveys and feedback. If documentation is embedded in system logic, functionality should demonstrate the implementation of policies and procedures.

- Watch Services, triggers
- Verification against changes in the environment
- Update of preservation plans

ISO 16363 / TRAC

- TRAC and Preservation Planning 2:

3.3.3 The repository shall have a documented history of the changes to its operations, procedures, software, and hardware.

Supporting Text

This is necessary in order to provide an 'audit trail' through which stakeholders can identify and trace decisions made by the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Capital equipment inventories; documentation of the acquisition, implementation, update, and retirement of critical repository software and hardware; file retention and disposal schedules and policies, copies of earlier versions of policies and procedures; minutes of meetings.

- History of preservation plans (created, reviewed and updated)
- Plato: Automated documentation of planning activities

ISO 16363 / TRAC

- TRAC and Preservation Planning 3:

3.3.4 The repository shall commit to transparency and accountability in all actions supporting the operation and management of the repository that affect the preservation of digital content over time.

Supporting Text

This is necessary because transparency, in the sense of being available to anyone who wishes to know, is the best assurance that the repository operates in accordance with accepted standards and practices.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Reports of financial and technical audits and certifications; disclosure of governance documents, independent program reviews, and contracts and agreements with providers of funding and critical services.

- Solid workflows in consistent manners enable informed and well-documented decisions
- Explicit definition of objectives and measurement units

ISO 16363 / TRAC

- TRAC and Preservation Planning 4:

4.1.1 The repository shall identify the Content Information and the Information Properties that the repository will preserve.

Supporting Text

This is necessary in order to make it clear to funders, depositors, and users what responsibilities the repository is taking on and what aspects are excluded. It is also a necessary step in defining the information which is needed from the information producers or depositors.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Mission statement; submission agreements/deposit agreements/deeds of gift; workflow and Preservation Policy documents, including written definition of properties as agreed in the deposit agreement/deed of gift; written processing procedures; documentation of properties to be preserved.

- Objective tree

ISO 16363 / TRAC

- TRAC and Preservation Planning 5:

4.3.1 The repository shall have documented preservation strategies relevant to its holdings.

Supporting Text

This is necessary in order that it is clear how the repository plans to ensure the information will remain available and usable for future generations and to provide a means to check and validate the preservation work of the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation identifying each preservation risk identified and the strategy for dealing with that risk.

• Preservation plan

ISO 16363 / TRAC

- TRAC and Preservation Planning 6:

4.3.3 The repository shall have mechanisms to change its preservation plans as a result of its monitoring activities.

Supporting Text

This is necessary in order for the repository to be prepared for changes in the external environment that may make its current preservation plans a bad choice as the time to implement draws near.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Preservation Plans tied to formal or informal technology watch(es); preservation planning or processes that are timed to shorter intervals (e.g., not more than five years); proof of frequent Preservation Policies and Preservation Plans updates; sections of Preservation Policies that address how plans may be updated and that address how often the plans are required to be reviewed and reaffirmed or updated.

- Watch Services, triggers
- Verification against changes in the environment
- Update of preservation plans

Example: HATHITrust Digial Library

- Review of the TRAC Checklist from 2010:
- http://www.hathitrust.org/trac
- HathiTrust is a large-scale digital repository
- Internet Archive
- Google Books

Nestor Kriterienkatalog (DIN 31644)

- 34 criteria for trustworthy repositories
- Version 2, 2008
- Three pilars
  - Organisational Framework
  - Object Management
  - Infrastructure and Security
- http://nbn-resolving.de/urn:nbn:de:0008-2010030806

Nestor – Organisational Framework 1

2.1	**The digital repository ensures its designated community/ communities can access the digital objects.**	Metric
	The DR should ensure that authorised users have access to the digital objects. This includes appropriate search possibilities. When determining its service portfolio, the DR takes the needs of its designated community/communities into account. The DR announces in advance its conditions of use and any costs which may arise, listing these in a transparent manner.	Description
	Access can be obtained by: ▪ Accessing the digital objects ▪ Creating or supplying an analogue copy (e.g., as print-out by the user or in the form of a print-on-demand service) ▪ Creating or supplying a digital copy (e.g. download to a storage medium by the user, email delivery) ▪ Creating interfaces to permit access via other systems to the digital objects.	Evidence example

Nestor – Organisational Framework 2

5.2 The digital repository documents all its elements based on a defined process.

The elements include: targets, plans, specifications, implementations, processes, software, objects and metadata etc.
The quality management system should include a suitable procedure for documentation, i.e. a system to manage all necessary documents. The DR should lay down rules regarding the completeness, correctness, validity, comprehensibility and accessibility of the documentation, implement these and monitor their observance.
This avoids knowledge being tied to certain individuals.

Standardised terminology which is adapted to the needs of the documentation users helps improve comprehensibility, for instance. Accordingly the documentation can be formal (e.g. for description of critical software processes), semi-formal (for conceptual description of processes and IT infrastructure) or natural (e.g. for external description of archive's objectives).

- Software documentation
- Process documentation
- Documentation of object formats

Nestor – Object Management 1

8 The digital repository has a strategic plan for its technical preservation measures (preservation planning).

In order to fulfil its responsibility for preserving information, the DR should have a strategic plan covering all outstanding or expected tasks, and the time of their realisation. This strategic planning (cf. 4.4) should be specified at the object level. Such measures should keep pace with ongoing technical developments (changes to data carriers, data formats, user demands etc.).

Measures for the physical preservation of the data (integrity, authenticity), its accessibility and the preservation of its interpretability should be used for the long-term preservation of the information represented by digital objects. Long-term preservation measures cover both content and metadata.

See 10.4 regarding implementation of the long-term preservation measures.

Output onto analogue media (e.g. microfilm) and redigitisation may be appropriate for certain digital objects.
The following are the main methods used to preserve interpretability:

Nestor – Object Management 2

9.2 The digital repository identifies which characteristics of the digital objects are significant for information preservation.

In determining the scope of the characteristics to be preserved, a balance should be struck between the goals concerning the technical possibilities and the costs of long-term preservation on the one hand and the needs of the designated community/communities on the other hand.

It may be effective to obtain different representations of an information object in order to preserve as many characteristics as possible.

Regarding information from databases, it may be sufficient to archive the data as so-called "flat files" (including a precise description of the data structure).

With regard to electronic files, the DOMEA system specifies saving the individual documents as image files. This precludes full-text searches and the executability of some documents (Excel tables or PowerPoint presentations).

Regarding web pages containing text-image information, one archive can store only text information, a second, only images, and a third, the entire interrelation. The different objectives lead to correspondingly different archiving strategies.

Screenshots from a standard browser are taken of web pages, but the text information is also stored for ease of research.

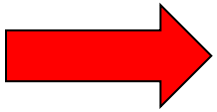
Digital Presevation and Risk

- Digital preservation is risk management
- We can‘t get rid of risks
  - But we can make them manageable
  - Try to predict them
  - Develop mitigation strategies
- Standard risk models exist in many disciplines

https://secure.flickr.com/photos/dsol/

Assessing Risk: A Practical Approach

- Criteria Catalogues are often abstract
  - How to measure if goals have been reached?
  - How to improve?
- So far:
  - Self assessment
  - Self audit
  - External audit
- How to evaluate risk?



DRAMBORA

DRAMBORA

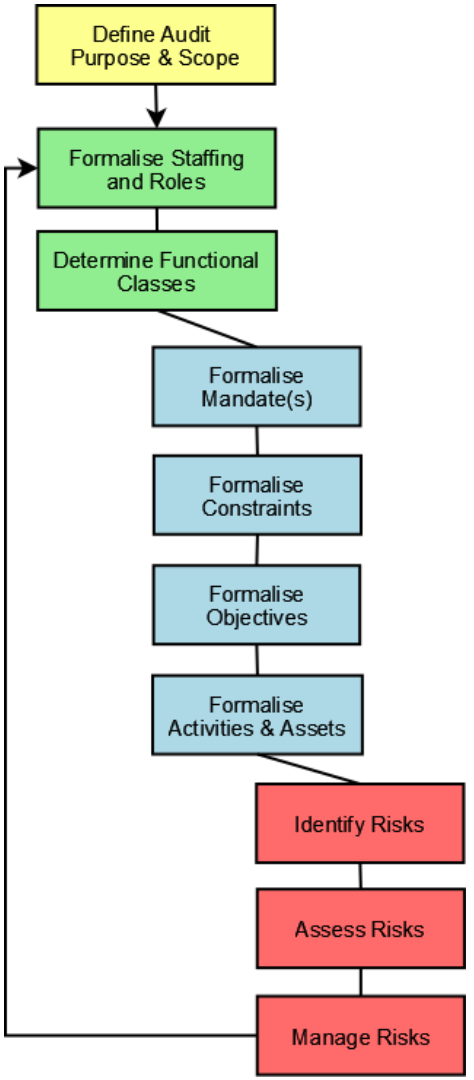
- Digital Repository Audit Method Based on Risk Assessment
- Aims to
  - Facilitate internal audit of digital preservation repositories
  - Assess capabilities of repository
  - Identify weaknesses
  - Recognise strengths
- Documented organisational self-awareness

DRAMBORA

- Covers the following aspects:
  - information assets (analogue/digital materials, databases, data files, contracts, agreements, documentation, policies and procedures);
  - software assets;
  - physical assets;
  - services and utilities;
  - business processes;
  - people (staffing and skills);
  - intangibles, such as reputation.

DRAMBORA Process Steps

1. Establish organisational profile
2. Develop contextual understanding
3. Identify and classify repository activities and assets
4. Derive risk register
5. Assess risks
6. Commit to management strategies



http://www.repositoryaudit.eu/

DRAMBORA Risk Register

- In DRAMBORA, risks have several attributes: *probability*, *impact*, *severity* (derived, p*i), *area of expression*, *owner(s)*, and *management strategies*. Risks may also link to other risks.

Risk Identifier:	*A text string provided by the repository to uniquely identify this risk and facilitate references to it within risk relationship expressions*
Risk Name:	*A short text string describing the risk*
Risk Description:	*A longer text string offering a fuller description of this risk*
Example Risk Manifestation(s):	*Example circumstances within which risk will or may execute*
Date of Risk Identification:	*Date that risk was first identified*
Nature of Risk:	*Physical environment*
	Personnel, management and administration procedures
	Operations and service delivery
	Hardware, software or communications equipment and facilities
Owner:	*Name of risk owner - usually the same as owner of corresponding activity*
Escalation Owner:	*The name of the individual who assumes ultimate responsibility for the risk in the event of the stated risk owner relinquishing control*

Table from http://www.repositoryaudit.eu/

DRAMBORA Risk Impact

- In terms of:
  - impact on repository staff or public well-being
  - impact of damage to or loss of assets
  - damage to reputation
  - damage to financial viability
  - deterioration of product or service quality
  - loss of digital object authenticity and understandability

DRAMBORA Risk Impact Scores

Risk Impact Score	Interpretation
0	*Zero* impact, results in **zero loss** of digital object authenticity and understandability
1	*Negligible* impact, results in **isolated but fully recoverable loss** of digital object authenticity and Understandability
2	*Superficial* impact, results in **widespread but fully recoverable** loss of digital object authenticity and Understandability
3	*Medium* impact, results in **total but fully recoverable loss** of digital object authenticity and understandability
4	*High* impact, results in **isolated loss**, including unrecoverable loss of digital object authenticity and Understandability
5	*Considerable* impact, results in **widespread loss**, including unrecoverable loss or loss that is recoverable only by third party of digital object authenticity and understandability
6	*Cataclysmic* impact, results in **total and unrecoverable loss** of digital object authenticity and Understandability

See http://www.rinascimento-digitale.it/eventi/conference2009/slides14/cirinna.pdf

DRAMBORA Risk Probability

Risk Probability Score	Interpretation
1	Minimal probability, occurs once every **100 years or more**
2	Very low probability, occurs once every **10 years**
3	Low probability, occurs once every **5 years**
4	Medium probability, occurs once **every year**
5	High probability, occurs once **every month**
6	Very high probability, occurs **more than once every month**

See http://www.rinascimento-digitale.it/eventi/conference2009/slides14/cirinna.pdf

DRAMBORA Risk Relationships

Risk Relationship	*Definition of Risk Relationship*
Explosive	where the simultaneous execution of *n* risks has an impact in excess of the sum of each risk occurring in isolation
Contagious	where a single risk's execution will increase the likelihood of another's
Complementry	where avoidance or treatment mechanisms associated with one risk also benefit the management of another
Domino	where avoidance or treatment associated with a single risk renders the avoidance or treatment of another less effective
Atomic	where risks exist in isolation, with no relationships with other risks

Table from http://www.repositoryaudit.eu/

DRAMBORA Risks

- Risk management
  - Combination of avoidance, tolerance and transfer
  - avoid circumstances in which risk arises
  - limit likelihood of risk
  - reduce potential impact of risk
  - share the risk
  - Transfer to others

DRAMBORA Risk Example

Risk Identifier:	R24
Risk Name:	Inability to evaluate staff effectiveness or suitability
Mitigation strategy(ies):	Avoidance strategies: • Establish internal means of assessment including risk management • Seek relevant external certification in order to demonstrate staff competence • Undertake regular staff development reviews
Risk Relationships:	→R01 [contagious] →R02 [contagious] →R19 [contagious]
Risk Probability:	4
Risk Potential Impact:	3
Risk Severity:	12
Owner:	Management
Escalation Owner:	Management
Stakeholders:	Management; financiers; staff; depositors; users; producers

See http://www.rinascimento-digitale.it/eventi/conference2009/slides14/cirinna.pdf

DRAMBORA Examples 2

- Loss of key member(s) of staff
  - Individuals with roles, responsibilities or aptitudes vital to the achievement of business objectives part company with the repository, rendering achievement of those objectives less straightforward
- Example manifestation
  - Repository head systems administrator, the sole individual with knowledge of the system's root password, leaves the organisation to work elsewhere

DRAMBORA Examples 3

- Budget cut or withdrawl from funding
- Legal liabilitiy
- Security breach
- Hardware damage
- Hardware obsolecence
- Physical intrusion
- Loss of trust
- Community feedback not received
- Inconsistencies between backups
- …..

DRAMBORA Results

- Risk score for each risk quantifies risks' severity
- Composite risk score for each category
- Illustrates vulnerabilities
- Facilitates resource investment

Summary

- Trust in digital repositories
- Criteria Catalogues and audit schemes
  - Data Seal of Approval
  - TRAC
  - Nestor
- Risk in digital preservation
  - DRAMBORA