technology
from seed

José Barateiro*   Gonçalo Antunes   Filipe Freitas   José Borbinha

Designing Digital Preservation Solutions: A Risk Management Based Approach

5th international Digital Curation Conference
“Moving to Multi-Scale Science: Managing Complexity and Diversity”
2009, London

Outline

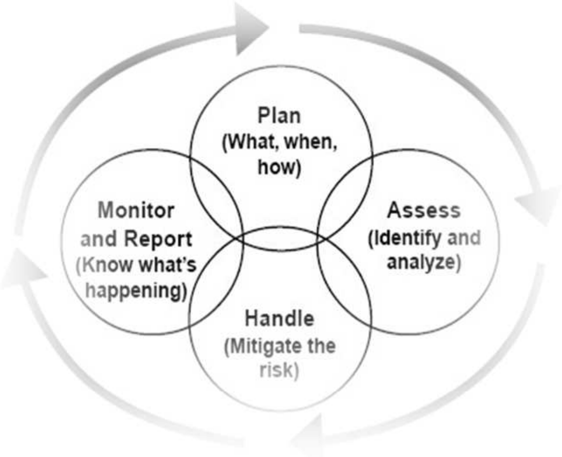
- Digital Preservation
- The Risk Management Approach
- Applying RM to Digital Preservation
  - Requirements
  - Threats and Vulnerabilities
  - Techniques
  - Addressing DP Threats and Vulnerabilities
- Related work
- Context
- Conclusions

Digital Preservation

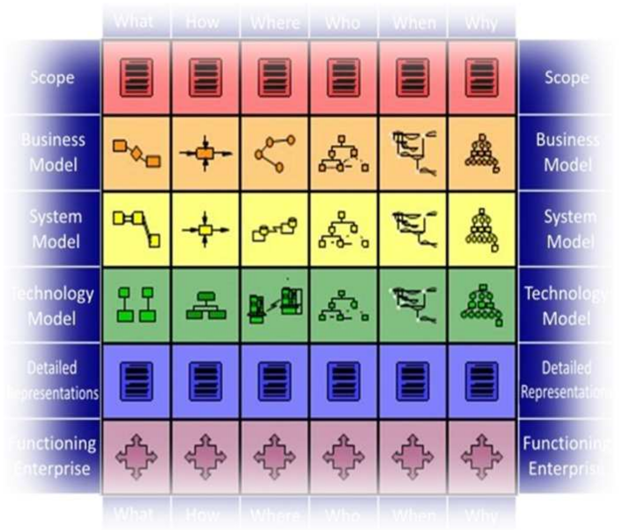
- “***Digital preservation** aims at maintaining digital objects accessible over long periods of time, ensuring the authenticity and integrity of these digital objects*”.
- IEEE defines **interoperability** as “*…the ability of two or more systems or components to exchange and use information...*”.
- Digital preservation stresses the **time dimension of interoperability**.

Approaching Digital Preservation

Systems Engineering



Risk Management



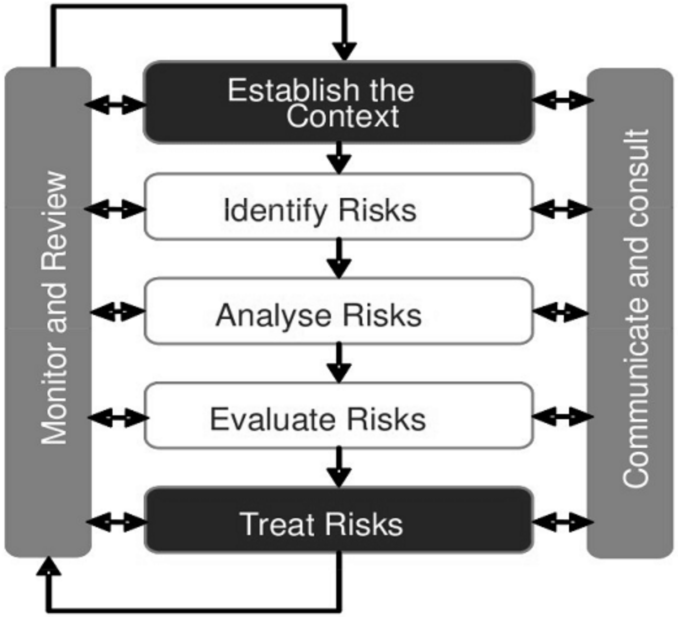
Enterprise Architecture

Plan (What, when, how)
Monitor and Report (Know what's happening)
Assess (Identify and analyze)
Handle (Mitigate the risk)

The Risk Management Perspective

- Risk Management: To define prevention and control mechanisms address the risk attached to specific activities and valuable assets, where risk is defined as the combination of the probability of an event and its consequences.

- A standard: ISO/FDIS 31000.


Establish the Context

Identify Risks

Analyse Risks

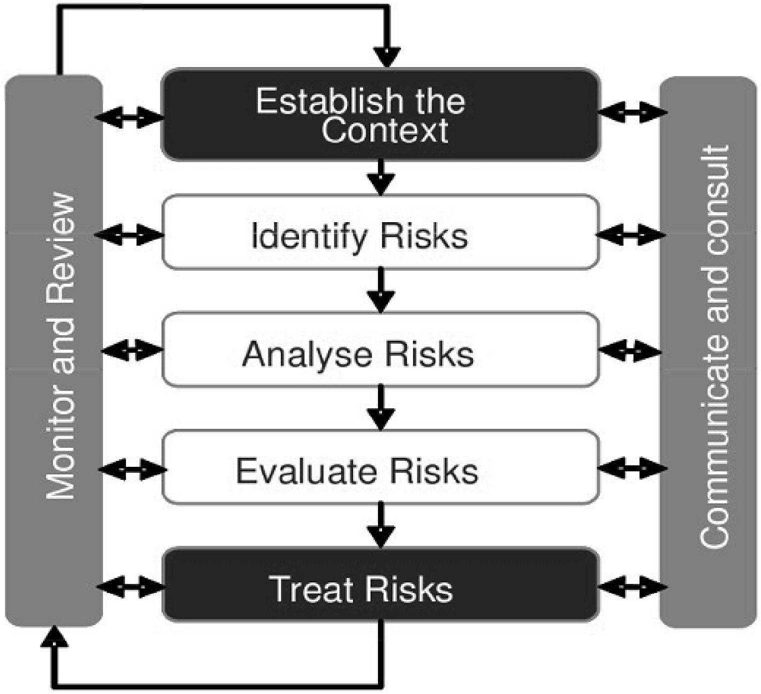
Evaluate Risks

Treat Risks

Monitor and Review

Communicate and consult

- **Establish the Context**: Digital Preservation Requirements!



Plan
(What, when, how)
Monitor
and Report
(Know what's
happening)
Assess
(Identify and
analyze)
Handle
(Mitigate the
risk)

Digital Preservation Requirements

technology
from seed

- **Reliability**: A copy (or representation) of any preserved object must survive over its system's lifetime.
- **Authenticity Assurance**: A future consumer may require the accessed information to be trustworthy.
- **Provenance**: A future consumer may require information concerning the origins of the object.
- **Integrity**: Effective preservation requires that the informational content of objects remains unchanged through its lifetime.

Digital Preservation Requirements

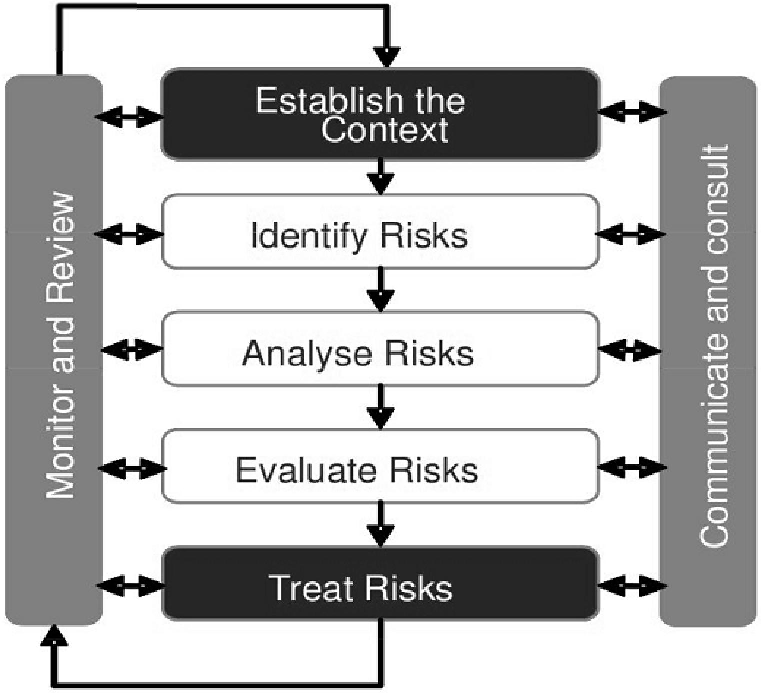
- **Dealing with Obsolescence**: Digital objects should be able to be exploited independently of any technological context (ideally…).

- **Scalability**: Digital preservation systems might be required to face technological evolution through the addition of new components.

- **Heterogeneity**: Digital preservation system's components should be heterogeneous due to technology disruption.

The Risk Management Perspective

- **Risk Assessment**: Identification, analysis and evaluation of threats and vulnerabilities!



Plan
(What, when, how)
Monitor and Report (Know what's happening)
Assess (Identify and analyze)
Handle (Mitigate the risk)

Digital Preservation
Threats and Vulnerabilities

Vulnerabilities	Process	Software faults / Software obsolescence
	Data	Media faults / Media obsolescence
	Infrastructure	Hardware faults / Hardware obsolescence / Communication faults / Network service failures
Threats	Disasters	Natural disasters / Human operational errors
	Attacks	Internal attack / External attacks
	Management	Economic failures / Organization failures
	Legislation	Legislation changes / Legal requirements

Vulnerability: ”…weakness, design, or implementation error that can lead to an unexpected, undesirable event…”.

Threat: ”…event with the potential to adversely impact an asset through unauthorized access, destruction, disclosure, modification of data…”.

(ISO/IEC Guide 73, 2002).

Plan (What, when, how)
Assess (Identify and analyze)
Handle (Mitigate the risk)
Monitor and Report (Know what's happening)

Digital Preservation Threats and Vulnerabilities

Vulnerabilities	Process	Software faults / Software obsolescence
	Data	Media faults / Media obsolescence
	Infrastructure	Hardware faults / Hardware obsolescence / Communication faults / Network service failures
Threats	Disasters	Natural disasters / Human operational errors
	Attacks	Internal attack / External attacks
	Management	Economic failures / Organization failures
	Legislation	Legislation changes / Legal requirements

Vulnerability.Data: affecting the information entities.

Digital Preservation
Threats and Vulnerabilities

Vulnerabilities	Process	Software faults / Software obsolescence
	Data	Media faults / Media obsolescence
	Infrastructure	Hardware faults / Hardware obsolescence / Communication faults / Network service failures
Threats	Disasters	Natural disasters / Human operational errors
	Attacks	Internal attack / External attacks
	Management	Economic failures / Organization failures
	Legislation	Legislation changes / Legal requirements

Vulnerability.Process: affecting the execution of processes (manual or supported by computational services) that control information entities.

Digital Preservation
Threats and Vulnerabilities

Vulnerabilities	Process	Software faults / Software obsolescence
	Data	Media faults / Media obsolescence
	Infrastructure	Hardware faults / Hardware obsolescence / Communication faults / Network service failures
Threats	Disasters	Natural disasters / Human operational errors
	Attacks	Internal attack / External attacks
	Management	Economic failures / Organization failures
	Legislation	Legislation changes / Legal requirements

Vulnerability.Infrastructure: technical problems in the infrastructure’s components.

Digital Preservation
Threats and Vulnerabilities

Vulnerabilities	Process	Software faults / Software obsolescence
	Data	Media faults / Media obsolescence
	Infrastructure	Hardware faults / Hardware obsolescence / Communication faults / Network service failures
Threats	Disasters	Natural disasters / Human operational errors
	Attacks	Internal attack / External attacks
	Management	Economic failures / Organization failures
	Legislation	Legislation changes / Legal requirements

Threat.Disasters: non-deliberate actions affecting the system's behaviour.

Digital Preservation
Threats and Vulnerabilities

Vulnerabilities	Process	Software faults / Software obsolescence
	Data	Media faults / Media obsolescence
	Infrastructure	Hardware faults / Hardware obsolescence / Communication faults / Network service failures
Threats	Disasters	Natural disasters / Human operational errors
	Attacks	Internal attack / External attacks
	Management	Economic failures / Organization failures
	Legislation	Legislation changes / Legal requirements

Threat.Attacks: deliberate actions affecting the system’s behaviour.

Digital Preservation
Threats and Vulnerabilities

Vulnerabilities	Process	Software faults / Software obsolescence
	Data	Media faults / Media obsolescence
	Infrastructure	Hardware faults / Hardware obsolescence / Communication faults / Network service failures
Threats	Disasters	Natural disasters / Human operational errors
	Attacks	Internal attack / External attacks
	Management	Economic failures / Organization failures
	Legislation	Legislation changes / Legal requirements

Threat.Management: consequences of wrong management and planning decisions.

Digital Preservation
Threats and Vulnerabilities

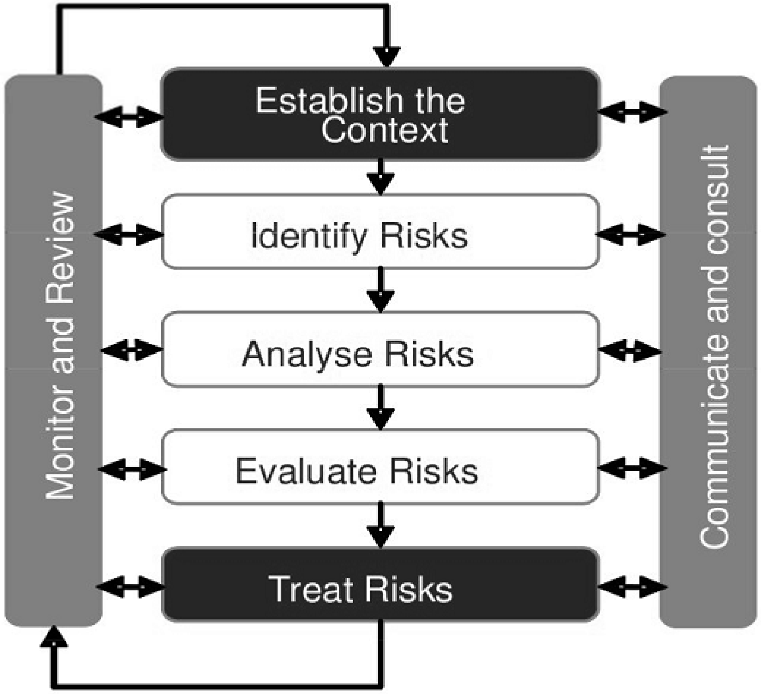
technology
from seed

Vulnerabilities	Process	Software faults / Software obsolescence
	Data	Media faults / Media obsolescence
	Infrastructure	Hardware faults / Hardware obsolescence / Communication faults / Network service failures
Threats	Disasters	Natural disasters / Human operational errors
	Attacks	Internal attack / External attacks
	Management	Economic failures / Organization failures
	Legislation	Legislation changes / Legal requirements

Threat.Legislation: digital preservation processes or preserved data violate new or updated legislation.

The Risk Management Perspective

- **Treat Risks**: Digital preservation techniques!



Plan
(What, when, how)
Monitor and Report (Know what's happening)
Assess (Identify and analyze)
Handle (Mitigate the risk)

Digital Preservation Techniques

- **Redundancy:** several copies of data can be stored across different components.
- **Migration:** keep digital object in recent formats. Lossless vs. loss migrations.
  - (*i*) Analog media; (*ii*) Version update; (*iii*) Conversion; (*iv*) Normalization.
- **Emulation:** simulation of the original environment.
- **Refreshing:** replacement of infrastructure’s components by most recent ones.

Digital Preservation Techniques

- **Diversity:** diversifying the properties of the system to avoid correlated failures.
  - (*i*) physical location; (*ii*) software; (*iii*) hardware; (*iv*) administration; (*v*) storage; (*vi*) funding.

- **Inertia**: “A system that works quickly also fails quickly!”. Thus, limiting the speed of the system can reduce the risk of abrupt failures.

- **Metadata:** “Data about data”.
  - (*i*) descriptive; (*ii*) technical; (*iii*) structural; (*iv*) preservation; (*v*) rights.

- **Auditing:** supports the detection of latent faults, allowing the system to recover faster and reducing the chance of losses.

Plan (What, when, how)
Monitor and Report (Know what's happening)
Assess (Identify and analyze)
Handle (Mitigate the risk)

Addressing digital preservation threats and vulnerabilities

Threats and vulnerabilities			Techniques							
			Redundancy	Migration	Emulation	Refreshing	Diversity	Inertia	Metadata	Auditing
Vulnerabilities	Process	Software faults	-	-	-	r	r	-	-	R
		Software obsolescence	-	-	-	r	r	-	-	R
	Data	Media faults	R	-	-	r	-	-	R	R
		Media obsolescence	-	r	r	-	-	-	R	R
	Infrastructure	Hardware faults	-	-	-	r	r	-	-	R
		Hardware obsolescence	-	-	-	r	r	-	-	R
		Communication faults	-	-	-	r	r	-	-	R
		Network service failures	-	-	-	r	r	-	-	R
Threats	Disasters	Natural disasters	R	-	-	-	r	-	-	-
		Human operational errors	R	-	-	-	r	r	R	R
	Attacks	Internal attack	R	-	-	-	r	r	R	R
		External attacks	R	-	-	-	r	r	R	R
	Management	Economic failures	-	-	-	-	r	-	-	R
		Organization failures	-	-	-	-	r	-	-	R
	Legislation	Legislation changes	-	-	-	-	r	-	r	-

- r: reduces the risk of threat/vulnerability; R: required for recovery; -: does not fit

Related work

- ***TRAC Criteria and Checklist*** is meant to identify potential risks to digital content held in repositories.

- ***DRAMBORA*** focuses on risks, and their classification and evaluation according to the activities, assets and contextual constraints of individual repositories.

- **The *Managing Information Risk guide for Accounting Officers, Board Members and Senior Information Risk Owners*** propose the following risk categories: Governance and culture; Information management and information integrity; The human dimension; Information availability and use.

- ***The UK Archives*** classifies risks into: Organizational; Process; Operational.

Context

technology
from seed

(http://grito.intraneia.pt)

- National project
- **Exclusive storage** clusters (dedicated to digital preservation)
- **Extended storage** clusters (using surplus resources of computing clusters)

SHAMAN - Sustaining Heritage Access through Multivalent ArchiviNg

(http://shaman-ip.eu/shaman)

- European project
- Three domains of focus: memory institutions, **engineering** and **e-Science**
- Strong focus on authenticity and integrity
- Definition of frameworks and architectures for digital preservation

Common ground: use of data grids (massive data sets, file management, user management, networking etc.)

Conclusions

- Risk Management based approach in three phases:
  - Establishing digital preservation requirements.
  - Identifying digital preservation threats and vulnerabilities.
  - Treating the risks associated with the identified threats and vulnerabilities.

- Provide generic requirements, threats, vulnerabilities and techniques.

- Future/ongoing work: simulator that can be used to evaluate the risk of threats (natural disasters) and infrastructure failures, on a preservation environment using redundancy and diversity techniques.

technology
from seed

José Barateiro – jbarateiro@lnec.pt

Gonçalo Antunes – goncalo.antunes@ist.utl.pt

Filipe Freitas – ffreitas@cc.isel.ipl.pt

José Borbinha – jlb@ist.utl.pt

LNEC

inesc id lisboa