A Risk Management Approach to Data Preservation

technology
from seed

Ricardo Vieira* (rjcv@ist.utl.pt)

Digital Preservation

- **Digital Preservation (DP)** aims at maintaining valuable digital objects accessible over long periods of time

» Problem initially triggered by memory institutions where DP is a main concern

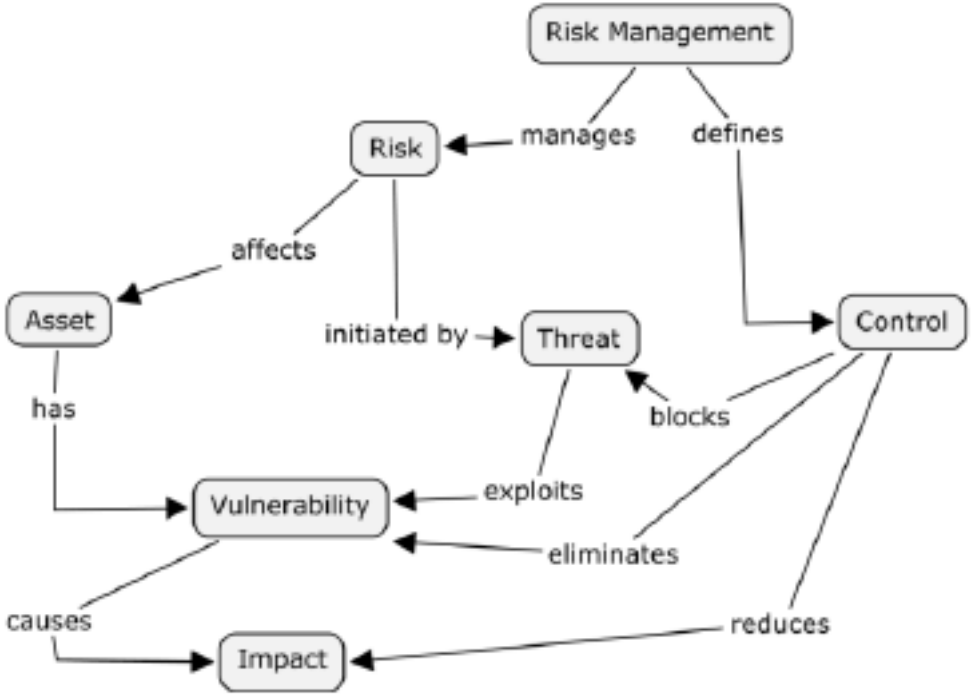
» Nowadays, widely present as a generalized concern of organizations

» **How to address it?**

Risk Management Overview

technology
from seed

- **Risk Management** defines prevention and control mechanisms to address the risk attached to specific activities and valuable assets



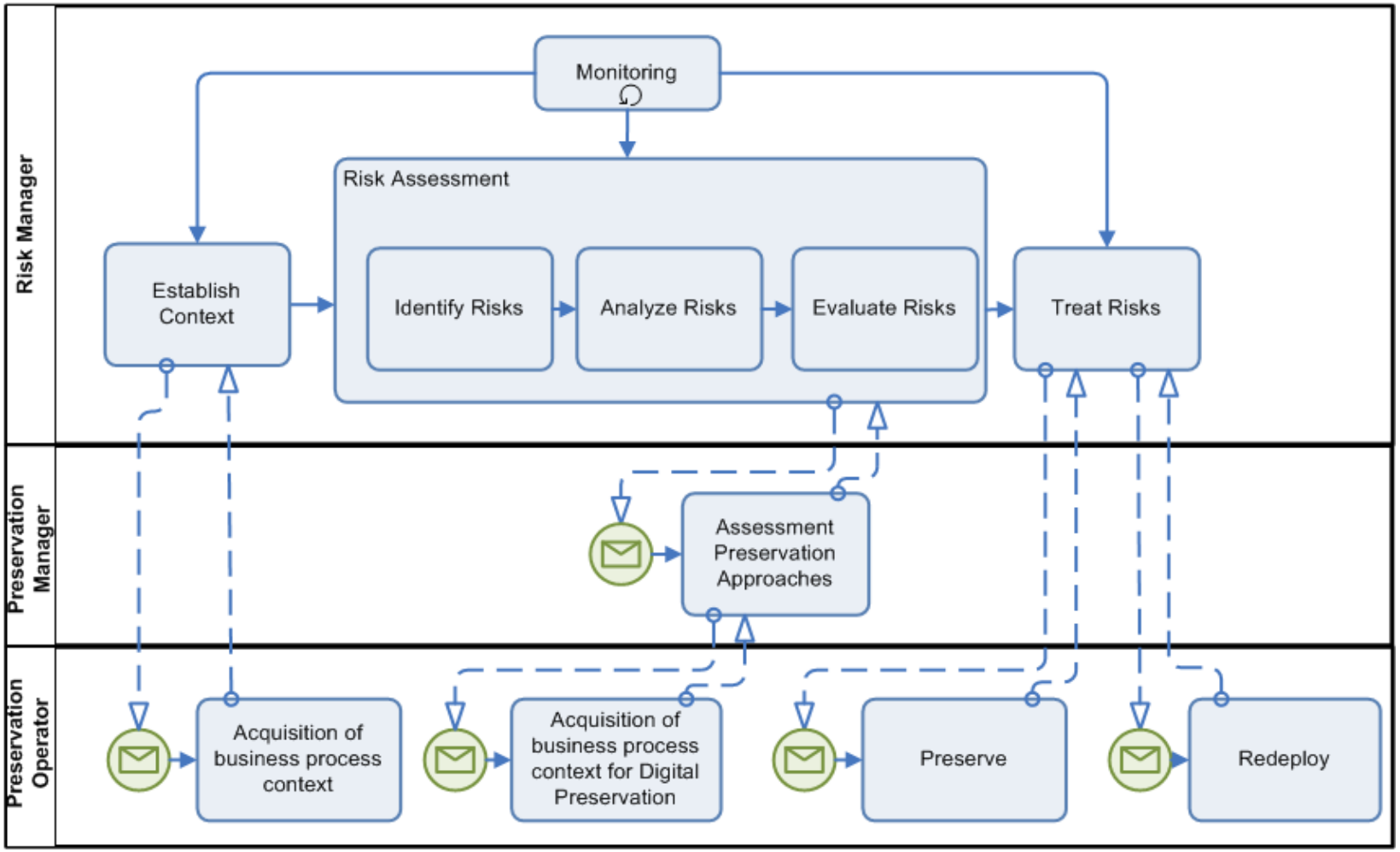
Digital Preservation as a Risk Management Activity

- The goal of DP is to protect digital information against several threats that can affect their proper use and interpretation

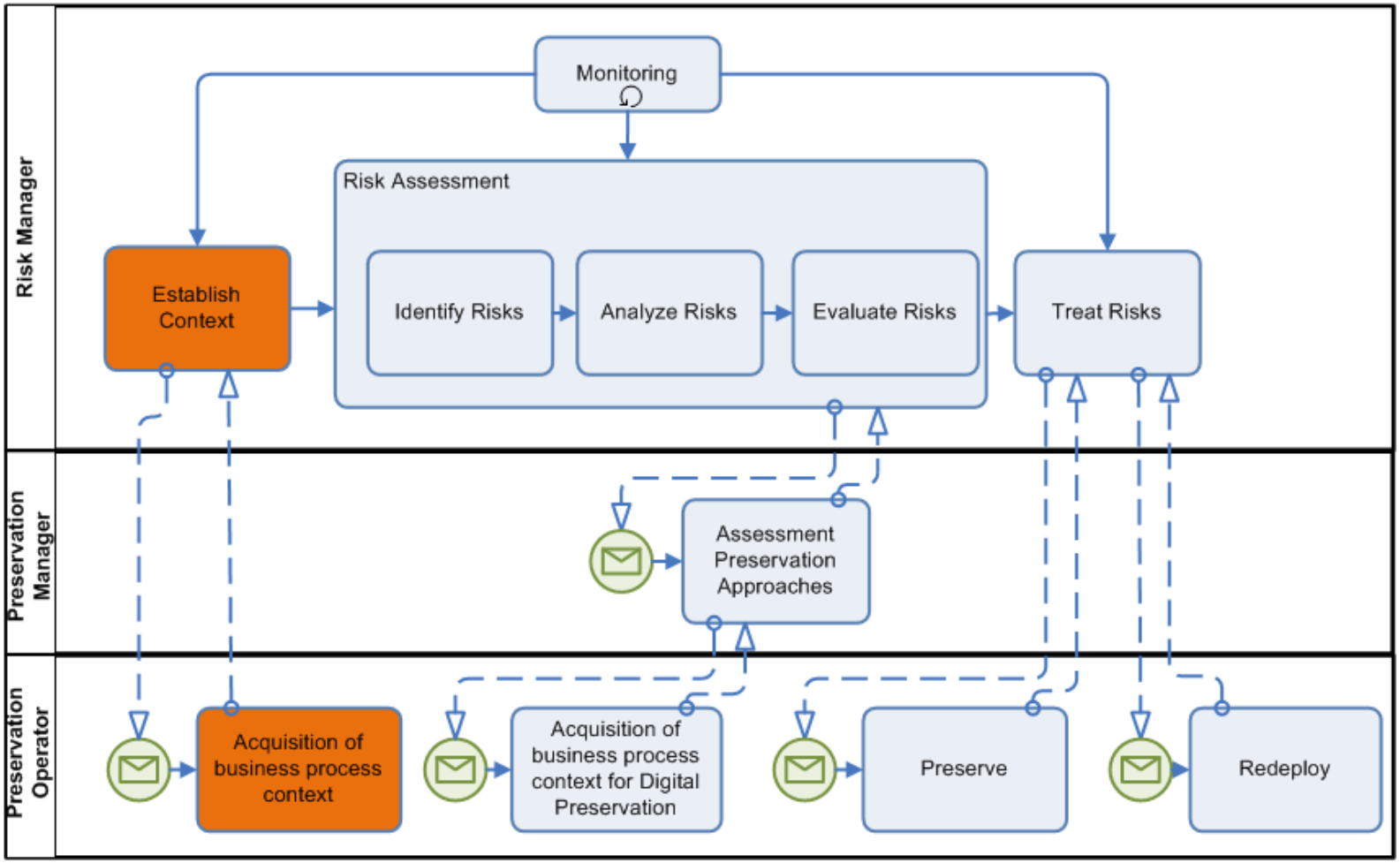
Threats and vulnerabilities			Techniques							
			Redundancy	Migration	Emulation	Refreshing	Diversity	Inertia	Metadata	Auditing
Vulnerabilities	Process	Software faults	-	-	-	r	r	-	-	R
		Software obsolescence	-	-	-	r	r	-	-	R
	Data	Media faults	R	-	-	r	-	-	R	R
		Media obsolescence	-	r	r	-	-	-	R	R
	Infrastructure	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 attack	R	-	-	-	r	r	R	R
	Management	Economic failures	-	-	-	-	r	-	-	R
		Organization failures	-	-	-	-	r	-	-	R
	Legislation	Legislation changes	-	-	-	-	r	-	r	-
		Legal requirements	-	-	-	-	r	-	r	-

r: Reduces the risk; R: Required for recovery; -: Does not fit

Risk Management Approach to Data Preservation



Establish the Context (1/2)

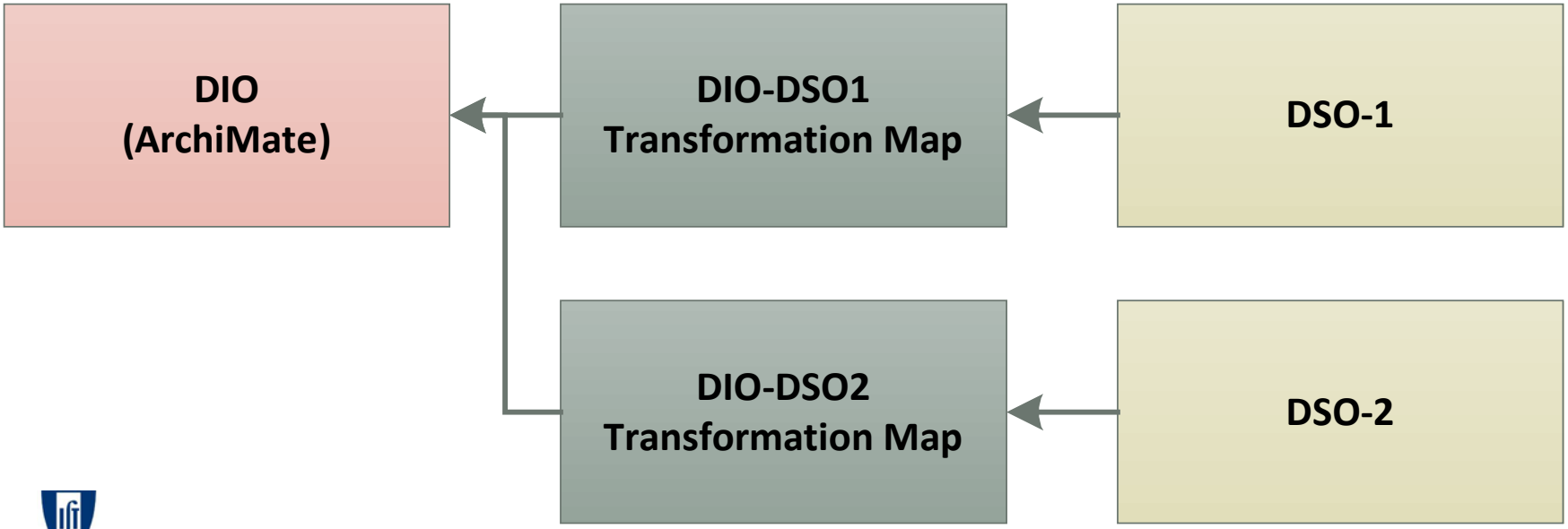


Establish the Context (2/2)

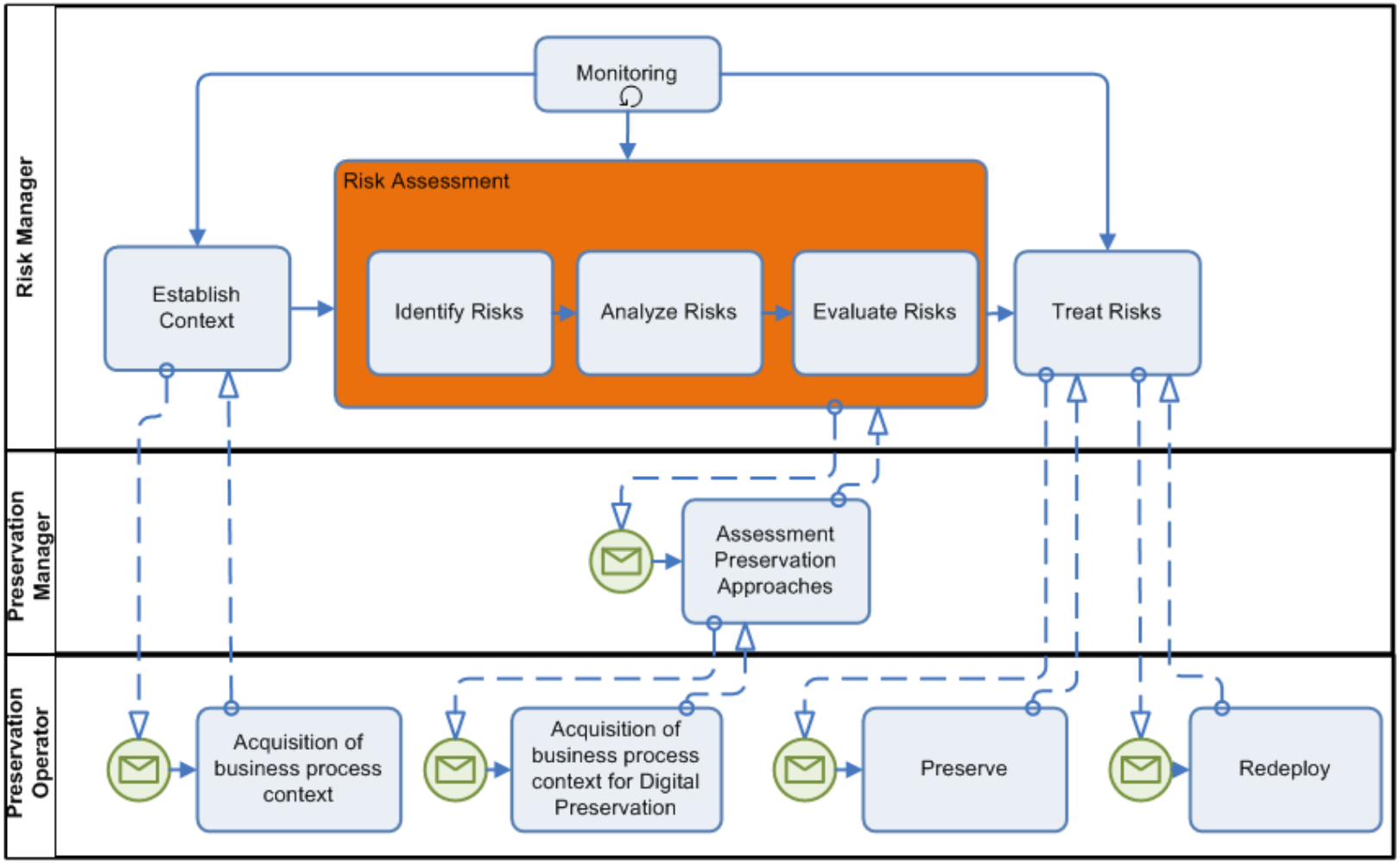
- Identify strategic objects and define criteria to determine which consequences are acceptable to the specific context.
  - Identify Stakeholders
  - Identify the context (Organisational, Technical and Legal)

Context Model

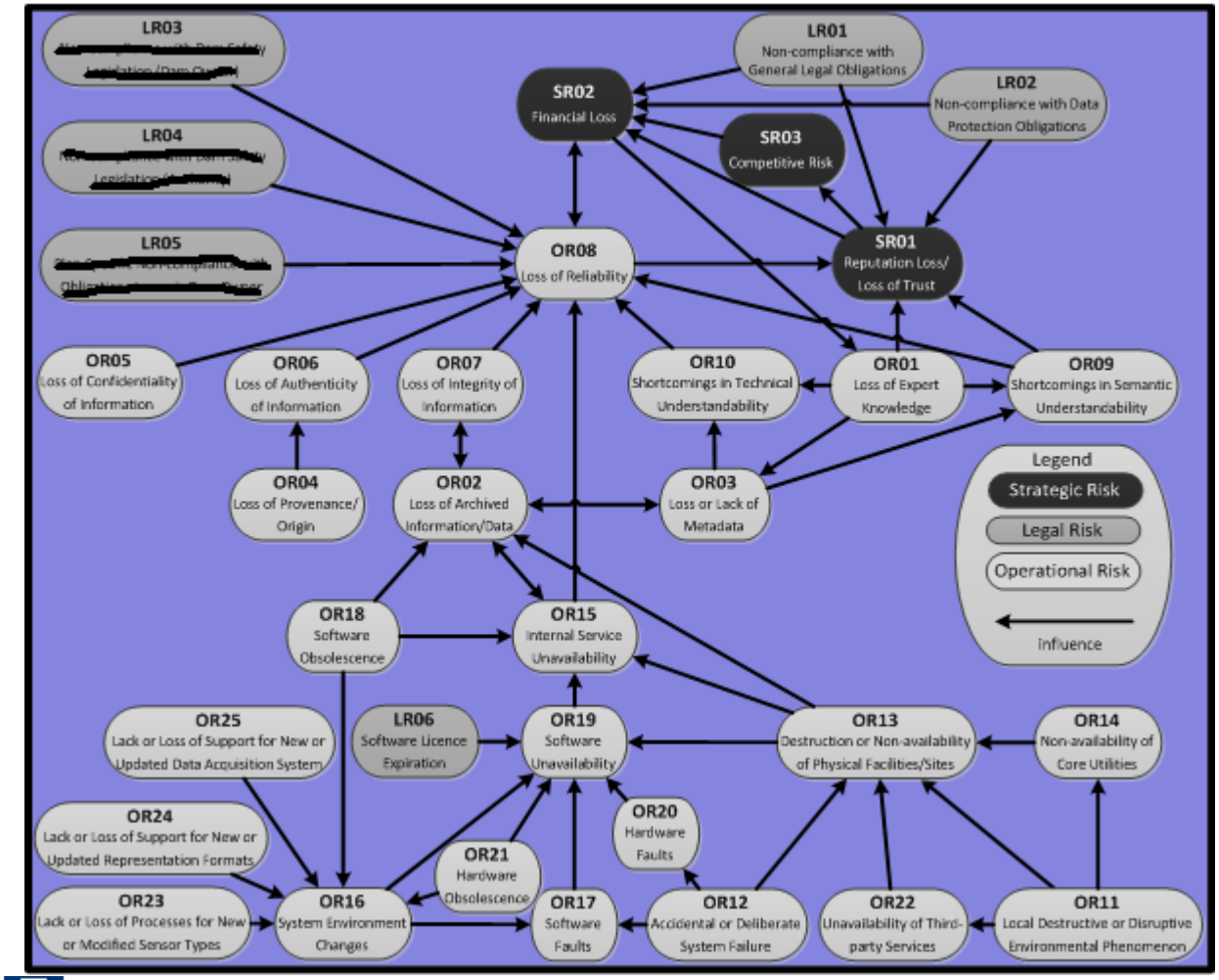
- The context model is represented as a set of ontologies
- Ontologies formalize knowledge representation
- Information can be extracted from ontologies through querying and processing (e.g. reasoning, logical inference)



Risk Assessment



Risk Identification



- Identify….
  - Assets
  - Vulnerabilities
  - Threats
  - Risks

Risk-DL Domain Language (1/2)

- Concepts mapped to ISO 31000
- Concepts formalized using the relational model
- Risk-DL to represente concepts
  - Interoperability
  - Sharing, discovery, reuse
  - Alignment between risks and organization artifacts
  - Reduce inconsistencies (formalization of risks)
  - Open specification -> support human-machine and machine-machine communication
  - XML properties:
    - Portability
    - Extensibility
    - Etc.

Risk-DL Domain Language (2/2)

technology
from seed

```
<xs:element name="asset">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="riskdl:name"/>
      <xs:element ref="riskdl:asset-Type"/>
      <xs:element ref="riskdl:description"/>
      <xs:element ref="riskdl:asset-value"/>
      <xs:element ref="riskdl:properties"/>
      <xs:element ref="riskdl:vulnerabilities"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="asset-Type" type="xs:NCName"/>
<xs:element name="asset-value" type="xs:NCName"/:
<xs:element name="properties">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="riskdl:property"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="property">
  <xs:complexType>
    <xs:attribute name="name" use="required"/>
    <xs:attribute name="value" use="required"/>
  </xs:complexType>
</xs:element>
```

Risk Analysis (1/2)

- Impact

Irrelevant	Minor	Noticeable	Major	Crucial	Catastrophic
<1000 €	1,000-6,000€	6,000-36,000€	36,000-216,000€	216,000-1,296,000€	>1,296,000€

- Likelihood

Very Low	Low	Moderate	High	Very High	Extreme
Never happened before	Once every 10 years	Once every 2 years	Once per half a year	Once every 2 months	More than once per month

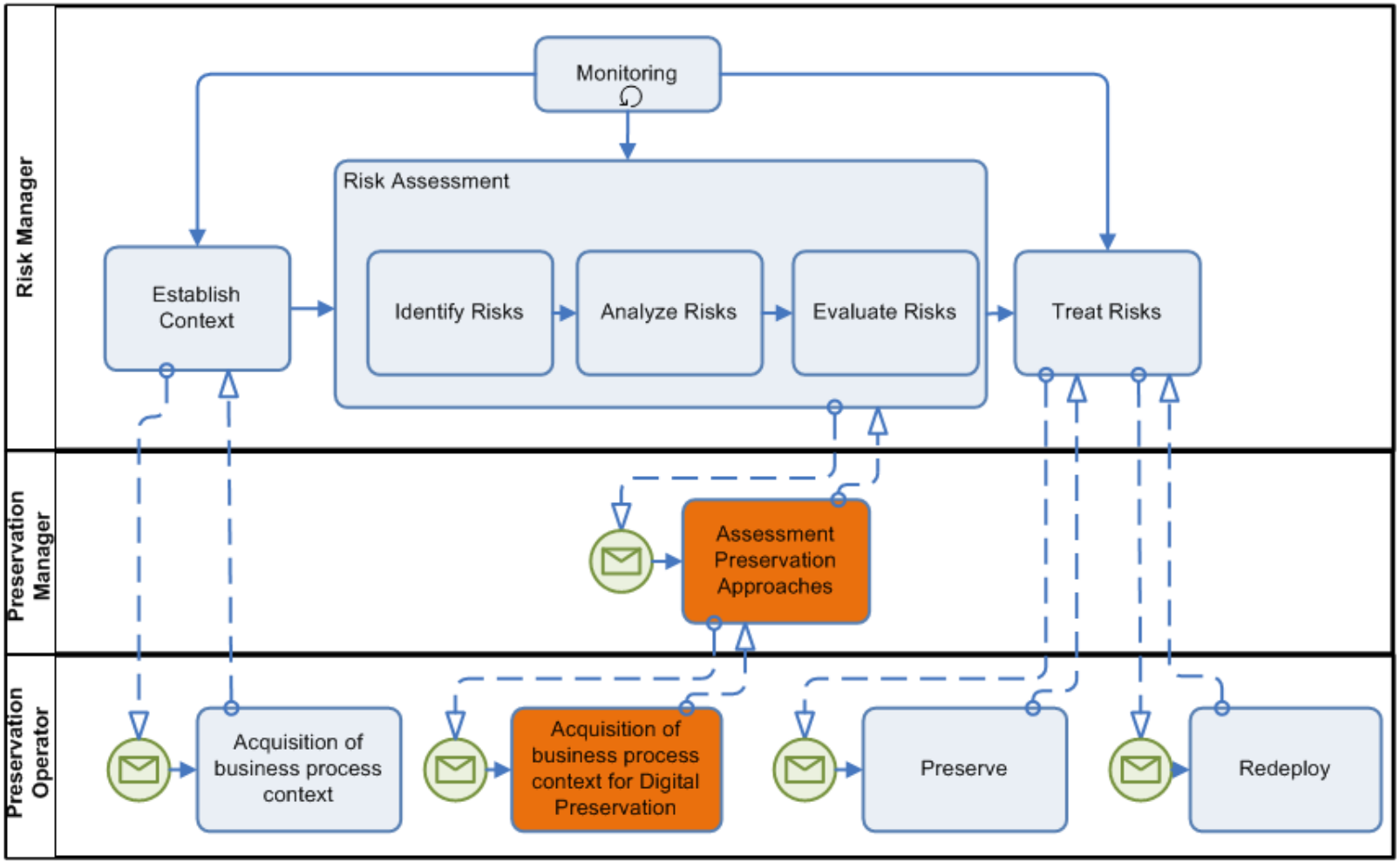
- Risk Level

Very Low	Low	Guarded	Moderate	High	Extreme
3-6	7-8	9-10	11-12	13-14	15-18

Risk Analysis (2/2)

Likelihood \ Impact	1	2	3	4	5	6
6						
5			OR03 OR15			
4		LR06 OR14	OR19 OR23	OR09	SR02	
3		OR04 OR06	LR02 LR05 OR20	OR01 OR10 OR17	LR01 OR07	OR08
2		OR05 OR22	SR03 OR24 OR25	LR03 OR21	LR04 OR02 OR12 OR16 OR18	SR01
1						OR11 OR13

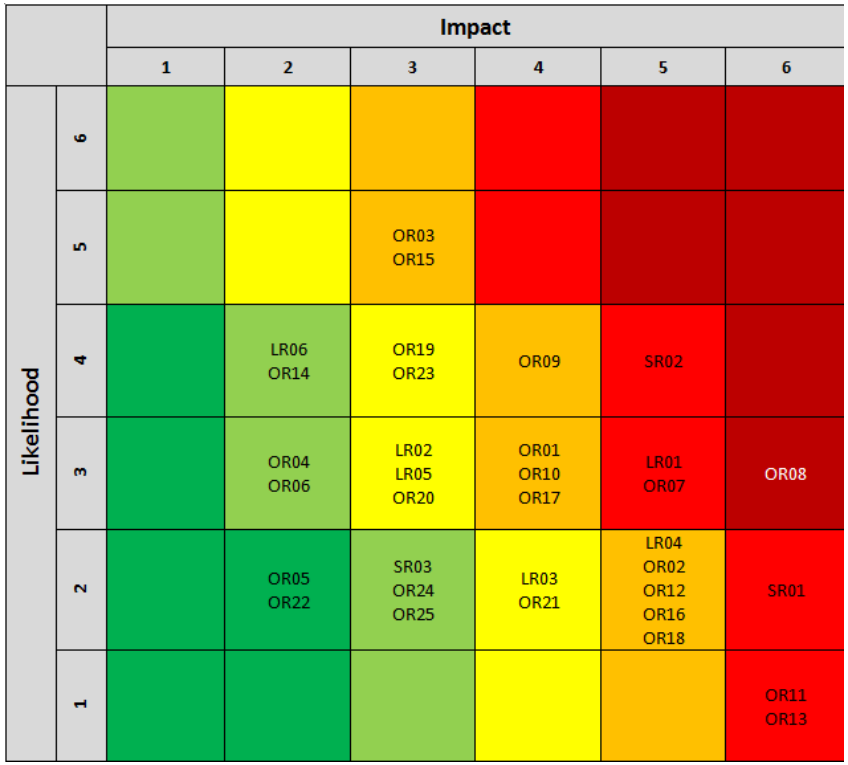
Scenario Evaluation



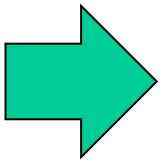
Risk Evaluation

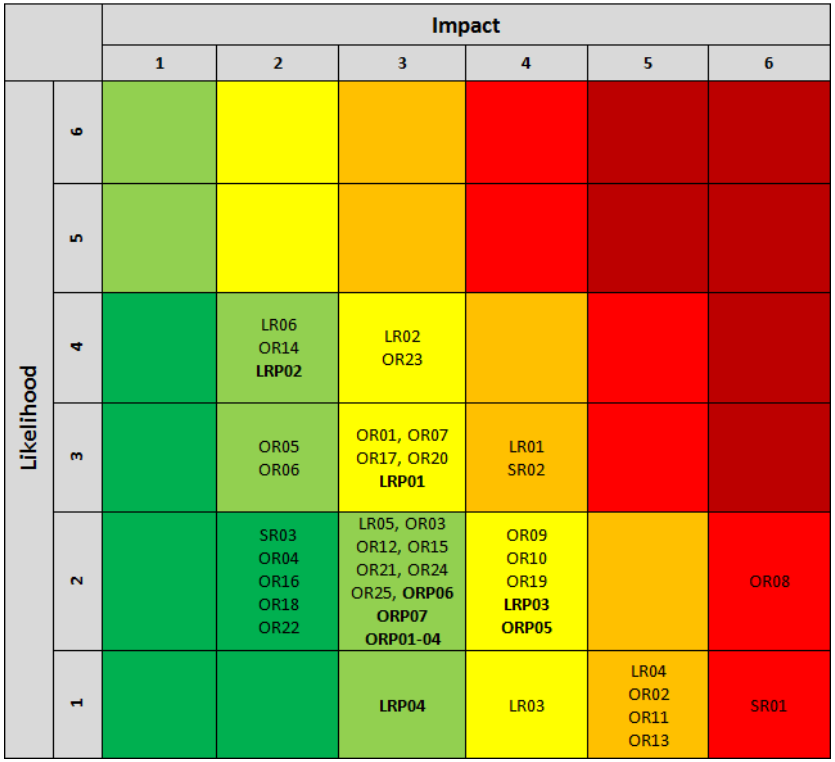
- Based on the outcome of risk analysis, decided which risks need treatment and the priority for treatment implementation
  - Compare level of risk with initial criteria

- Scenario evaluation: $Costs = \sum_i Likelihood_{Risk\ i} * Impact_{Risk\ i}$
  - DP mitigates the risk level for 19 risks
  - Costs without DP: 6,316,520 €/year
  - Costs with DP: 909,720€/year
  - Risk cost reduction of 5,406,800 €/year (=86%)



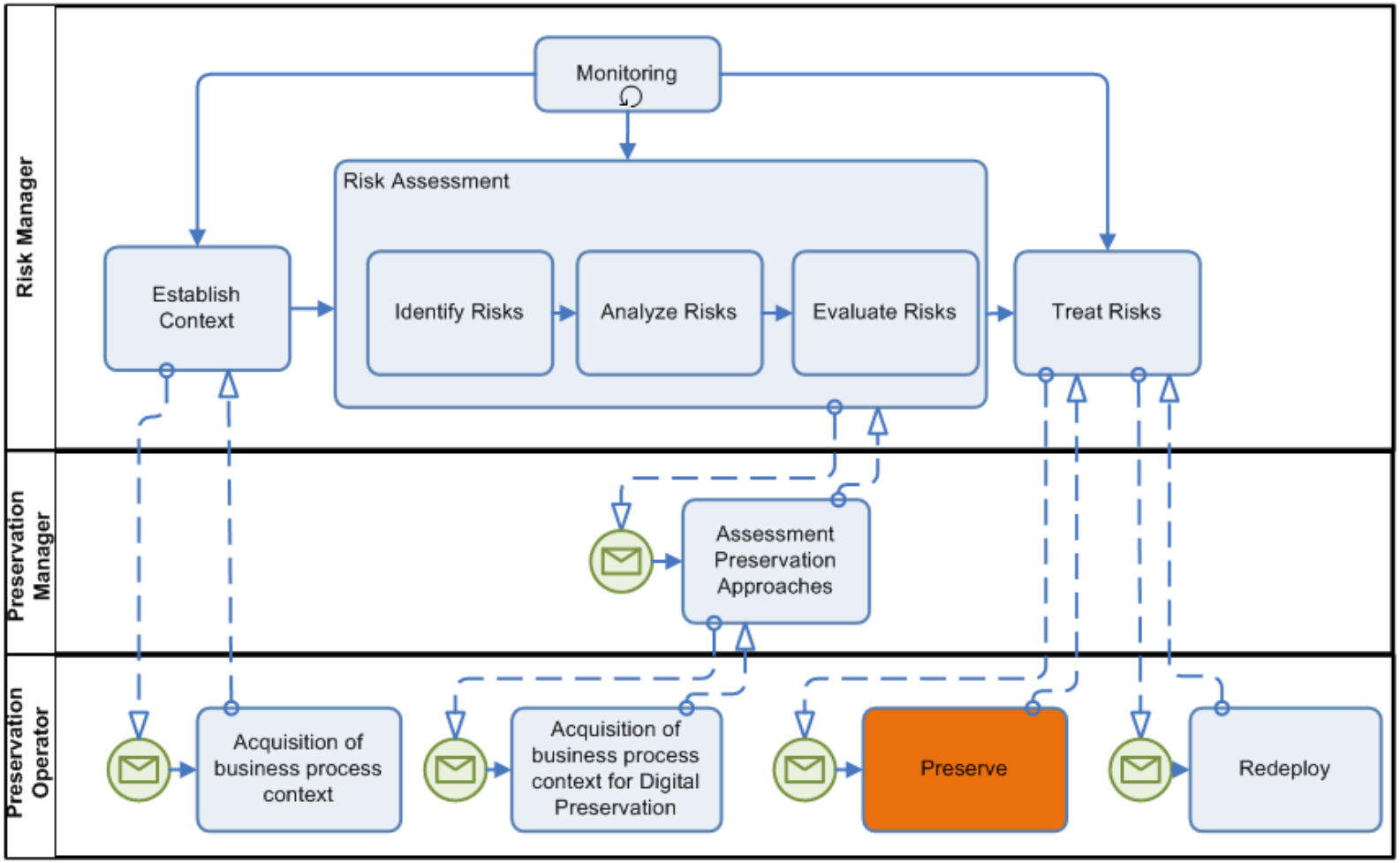
Likelihood	1	2	3	4	5	6
6						
5			OR03 OR15			
4		LR06 OR14	OR19 OR23	OR09	SR02	
3		OR04 OR06	LR02 LR05 OR20	OR01 OR10 OR17	LR01 OR07	OR08
2		OR05 OR22	SR03 OR24 OR25	LR03 OR21	LR04 OR02 OR12 OR16 OR18	SR01
1						OR11 OR13





Likelihood	1	2	3	4	5	6
6						
5						
4		LR06 OR14 **LRP02**	LR02 OR23			
3		OR05 OR06	OR01, OR07 OR17, OR20 **LRP01**	LR01 SR02		
2		SR03 OR04 OR16 OR18 OR22	LR05, OR03 OR12, OR15 OR21, OR24 OR25, **ORP06** **ORP07** **ORP01-04**	OR09 OR10 OR19 **LRP03** **ORP05**		OR08
1			**LRP04**	LR03	LR04 OR02 OR11 OR13	SR01

Risk Treatment



Conclusions

- Using risk management we are able to leverage the digital preservation problem
  - Better understanding of the problem
  - Better understanding of the solution

- **This is valid to other concerns!!!**

- Same approach is being applied to Data Management in e-Science project

  Ferreira, F., Coimbra, M., Vieira, R., Proença, D., Freitas, A., Russo, L., Borbinha, J., **Risk Aware Data Management in Metagenomics**, 5th INForum (INForum 2013), Évora, Portugal, September 5-6, 2013

technology from seed

Ricardo Vieira– rjcv@ist.utl.pt

inesc id lisboa