



Metadata for Preserving Computing Environments

Angela Dappert

Digital Preservation Coalition

The TIMBUS Project

PREMIS Editorial Committee

Overview: Current Research Activities

TIMELESS BUSINESS   



□ A vision:

❖ Processes and their execution environments:

- 🖨️ The TIMBUS Project

□ The building blocks:

❖ OAIS repository objects:

- 🖨️ PREMIS data dictionary;
- 🖨️ PREMIS Environment working group;
- 🗄️ file format work (Jhove, Pronom, UDFR, LoC, etc.);

❖ Software preservation:

- 🖨️ The Significant Properties of Software: A Study;
- 🖨️ Preserving Virtual Worlds;
- 🖨️ SWOP; 🖨️ DOAP;
- 🗄️ NSRL National Software Reference Library; 🗄️ AMINET

❖ Virtualized infrastructure:

- 🖨️ VRDF; 🖨️ TIMBUS Dependency Analysis; 🖨️ CDMI; 🖨️ WSDL

❖ Computing environments: 🗄️ KEEP Totem;

The TIMBUS Consortium

TIMELESS BUSINESS   



Digital
Preservation
Coalition

+ Members

- SAP - Lead partner (NI, CH)
- Intel (Ireland)
- Software Quality Systems (Germany)

Industry

- Digital Preservation Coalition (UK)
- INESC - ID (Portugal)
- Karlsruhe Institute for Technology (Germany)
- Laboratorio de Instrumentacao e Fisica Experimental de Particulas (Portugal)
- Laboratorio Nacional de Engenharia Civil (Portugal)
- Muenster University (Germany)

Research

- Caixa Magica Software (Portugal)
- iPharro Media (Germany)
- Secure Business Austria (Austria)

SMEs

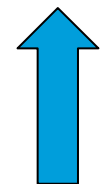
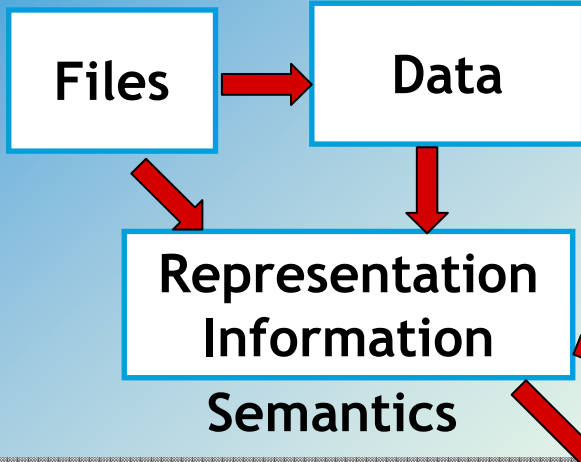
A Preservation Continuum

TIMELESS BUSINESS 

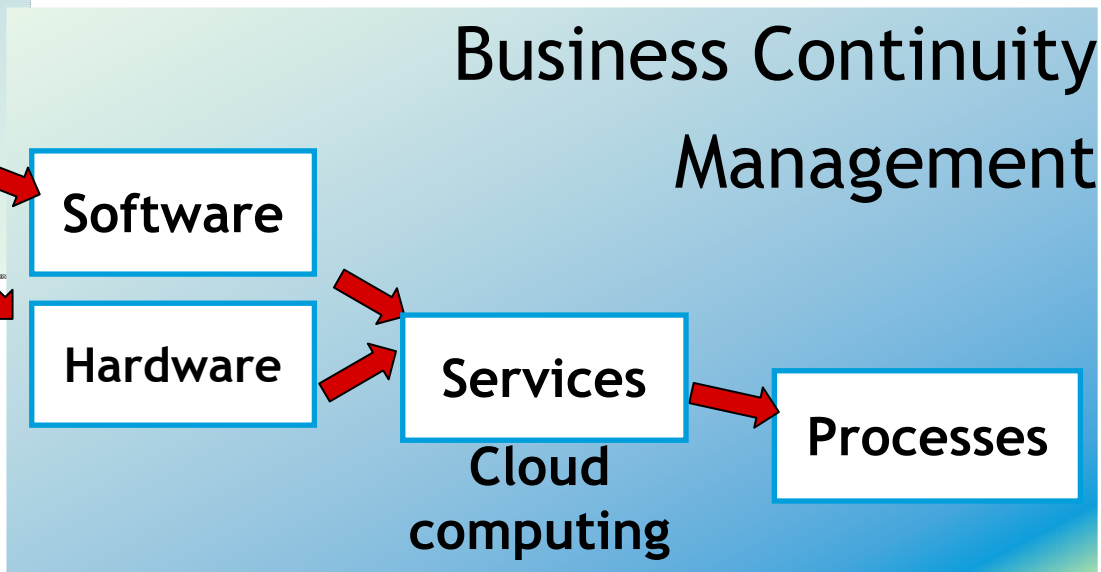


Longevity

Digital Preservation



Business Continuity Management



Complexity



Distribution of Work

TIMELESS BUSINESS   



R&D

- Processes
- Architecture
- Tools

Use Cases

- Engineering Services & Systems for Digital Preservation
- Civil Engineering Infrastructures
- eScience & Mathematical Simulations

- Exploitation
- Dissemination and Training



that ensure long-term continued access to business processes and the underlying software and hardware infrastructure

Components

TIMELESS BUSINESS   



Processes and Standards for Digital Preservation of Business Processes

Future Simulated Test Bed

- Intelligent Enterprise Risk Management
- Service Dependency Analysis
- Business Process Context Capture
- Legalities Lifecycle Management

- Business Process Virtualization and Storage (of distributed interdependent services)
- Validation of preserved business processes

- Business Process Exhumation
- Integration Support with Testbed
- Verification of Preserved Business Processes

Security
Authorization

Expediency

Execution

Exhumation

Case Study: Digital Asset Migration

TIMELESS BUSINESS   

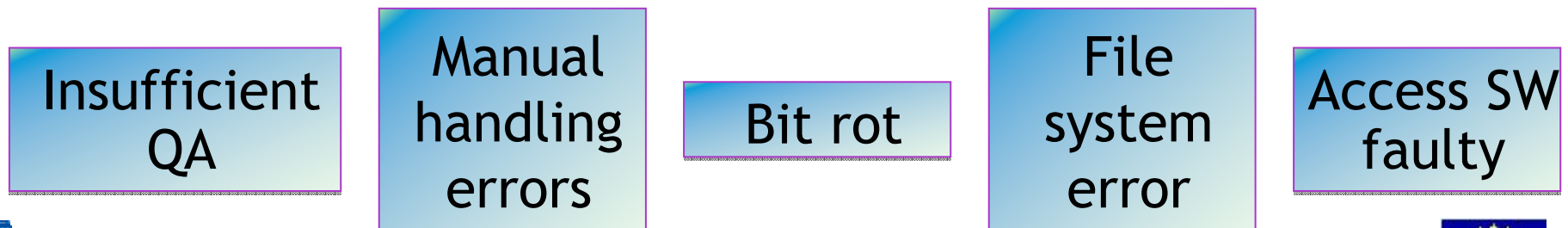


Case Study: Digital Asset Migration

TIMELESS BUSINESS   



Enterprise Risk Management

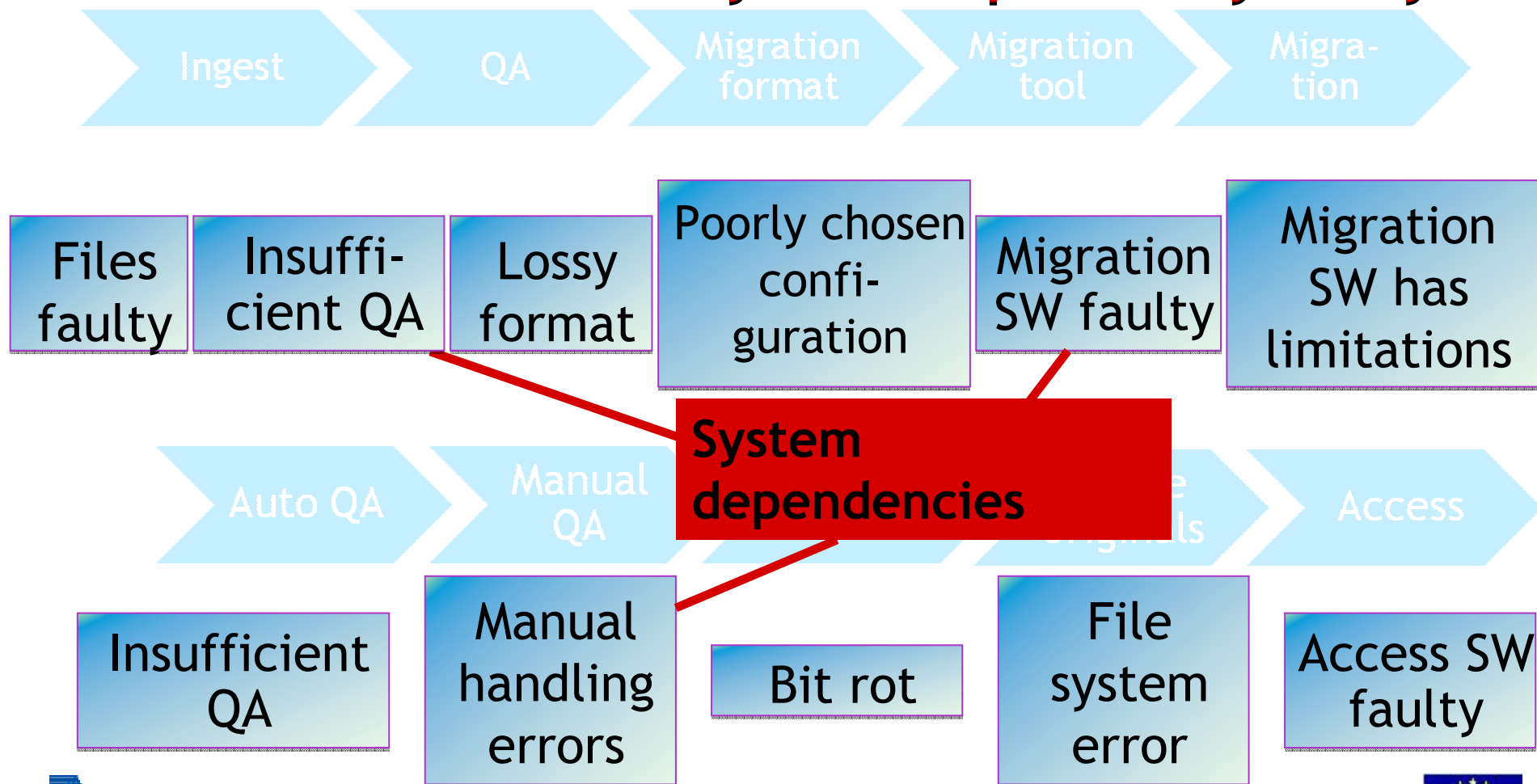


Case Study: Digital Asset Migration

TIMELESS BUSINESS



System Dependency Analysis

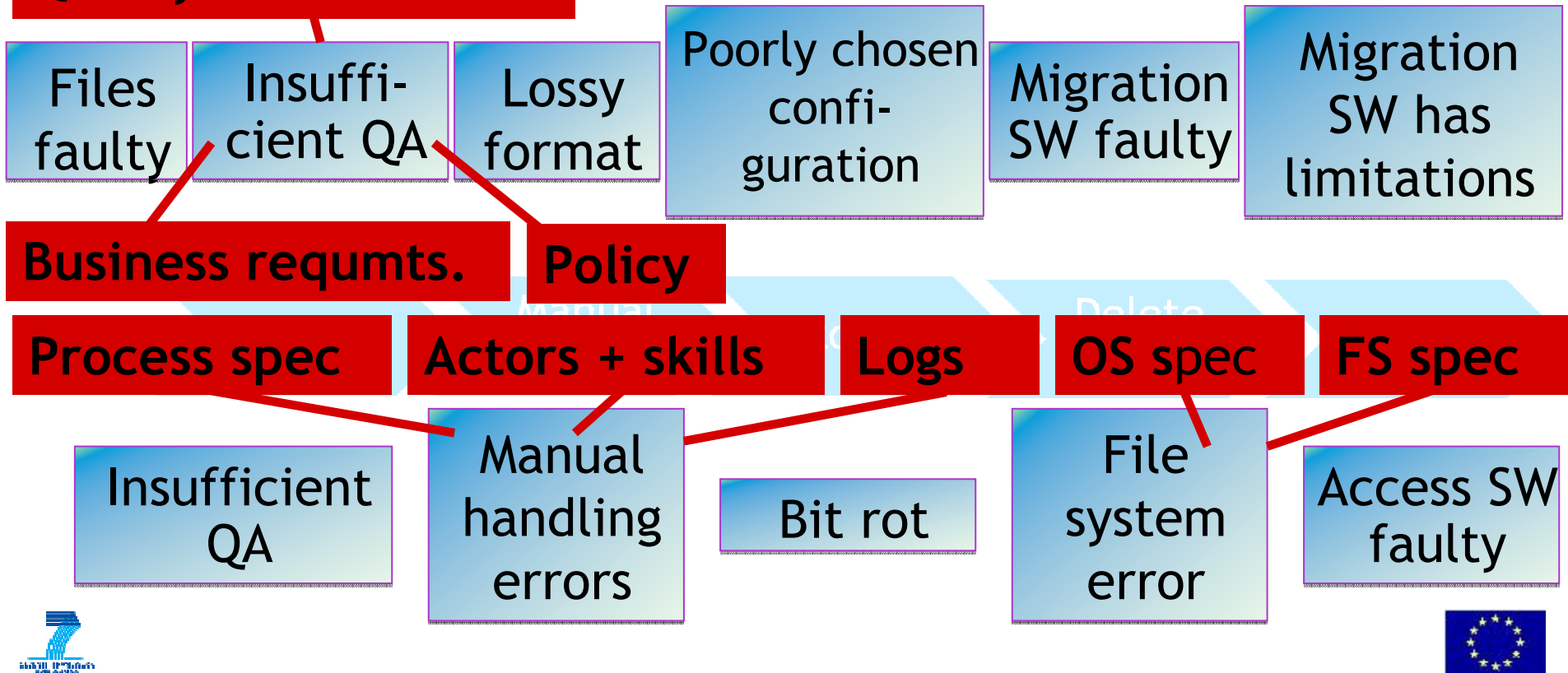


Case Study: Digital Asset Migration



Sampling methods
Test sets used
Test methods
Exception treatment
Processes and logs
Quality criteria

Business Process Context Capture

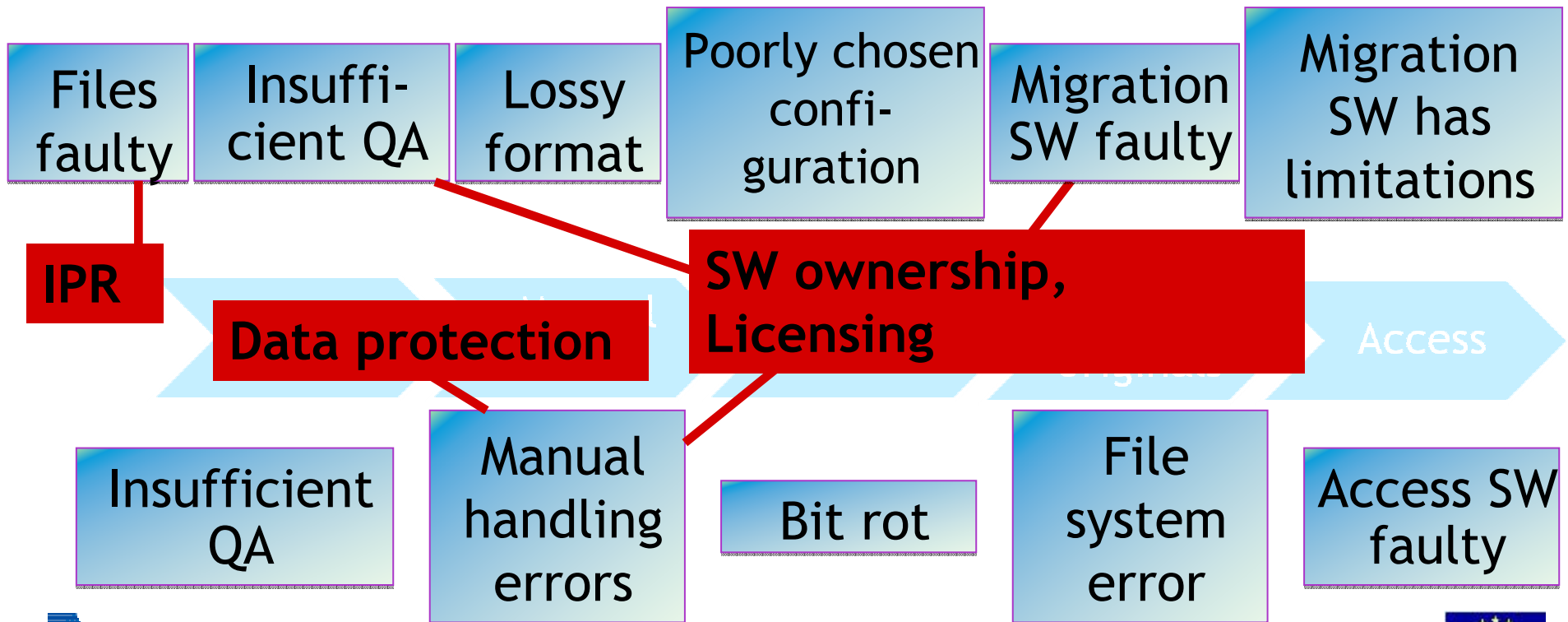


Case Study: Digital Asset Migration

TIMELESS BUSINESS   



Legalities Lifecycle Management



Components

TIMELESS BUSINESS   



Processes and Standards for Digital Preservation of Business Processes

Future Simulated Test Bed

- Intelligent Enterprise Risk Management
- Service Dependency Analysis
- Business Process Context Capture
- Legalities Lifecycle Management

- Business Process Virtualization and Storage (of distributed interdependent services)
- Validation of preserved business processes

- Business Process Exhumation
- Integration Support with Testbed
- Verification of Preserved Business Processes

Security
Authorization

Expediency

Execution

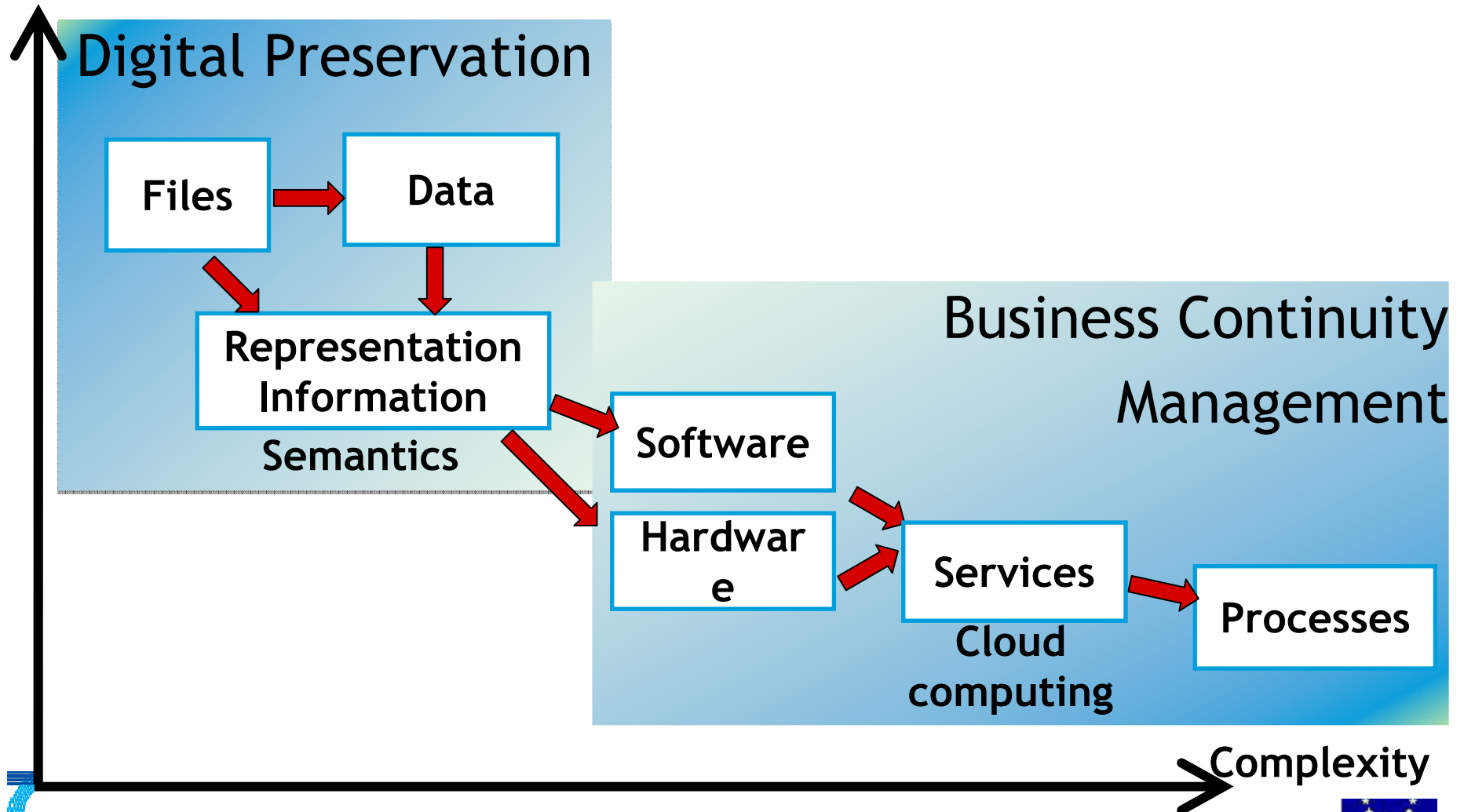
Exhumation

A Preservation Continuum

TIMELESS BUSINESS   



Longevity



Overview: Current Research Activities

TIMELESS BUSINESS   



□ A vision:

❖ Processes and their execution environments:

- 🖨️ The TIMBUS Project

□ The building blocks:

❖ OAIS repository objects:

- 🖨️ PREMIS data dictionary;
- 🖨️ PREMIS Environment working group;
- 🗄️ file format work (Jhove, Pronom, UDFR, LoC, etc.);

❖ Software preservation:

- 🖨️ The Significant Properties of Software: A Study;
- 🖨️ Preserving Virtual Worlds;
- 🖨️ SWOP; 🖨️ DOAP;
- 🗄️ NSRL National Software Reference Library; 🗄️ AMINET

❖ Virtualized infrastructure:

- 🖨️ VRDF; 🖨️ TIMBUS Dependency Analysis; 🖨️ CDMI; 🖨️ WSDL

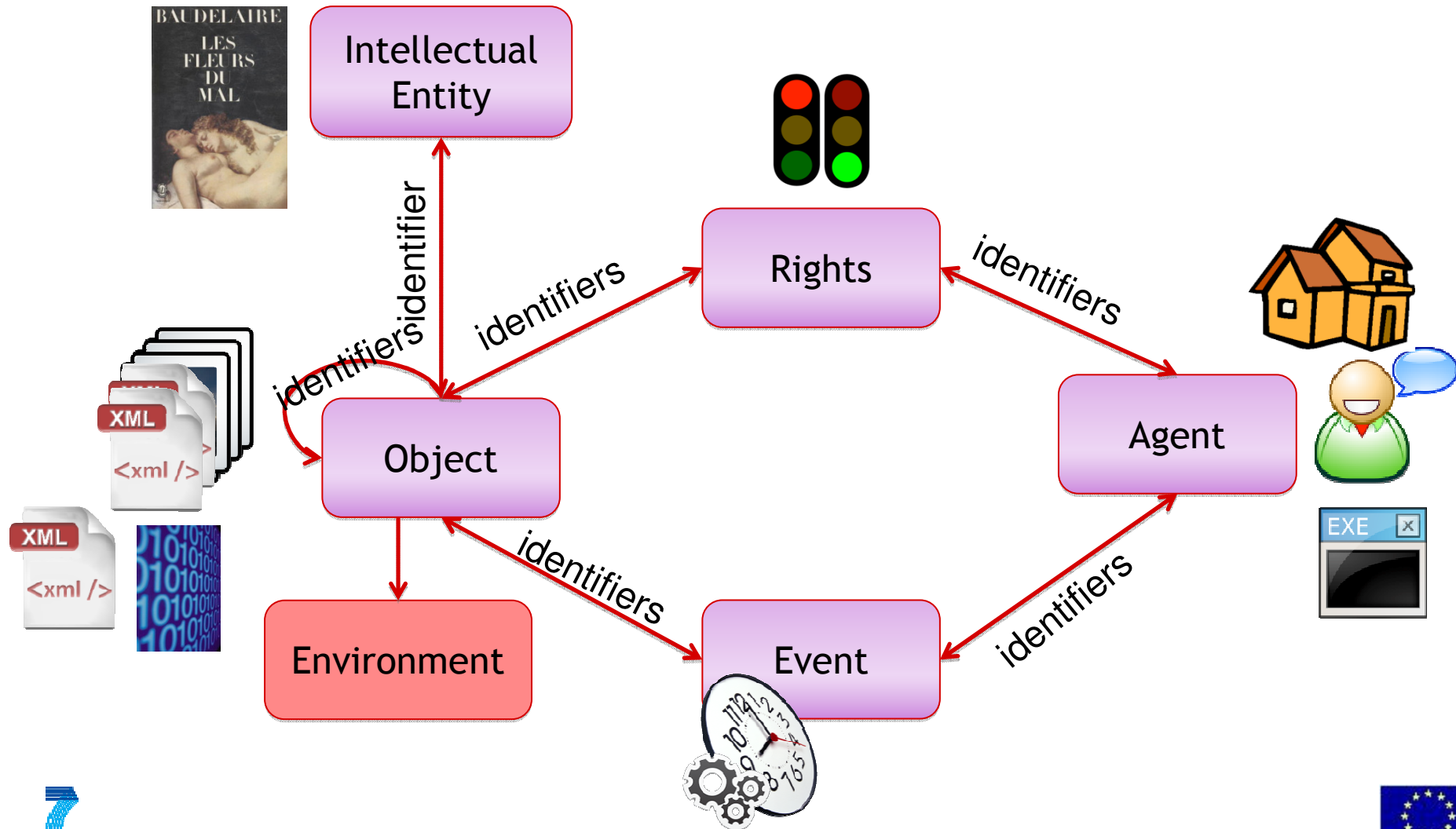
❖ Computing environments: 🗄️ KEEP Totem;

The PREMIS Data Model

TIMELESS BUSINESS ◀ © ▶



Slide by S. Peyrard



Example: Object Entity

TIMELESS BUSINESS ◀ © ▶



□ Main types of information

- ❖ identifier
- ❖ technical object characteristics
- ❖ creation information
- ❖ software and hardware environment
- ❖ digital signatures
- ❖ relationships to other objects
- ❖ links to other types of entity



1.5.5 creatingApplication

1.5.5.1 creatingApplicationName

1.5.5.2 creatingApplicationVersion

1.5.5.3 dateCreatedByApplication

1.5.5.4 creatingApplicationExtension

Semantic Unit: Environment

TIMELESS BUSINESS ◀ © ▶



- What is needed to render or use an object
 - ❖ Operating system
 - ❖ Application software
 - ❖ Computing resources

PREMIS - Environment Metadata

TIMELESS BUSINESS ◀ © ▶



1.8 environment

- 1.8.1 environmentCharacteristic
- 1.8.2 environmentPurpose
- 1.8.3 environmentNote

1.8.4 dependency

- 1.8.4.1 dependencyName
- 1.8.4.2 dependencyIdentifier
 - 1.8.4.2.1 dependencyIdentifierType
 - 1.8.4.2.2 dependencyIdentifierValue

1.8.5 software

- 1.8.5.1 swName
- 1.8.5.2 swVersion
- 1.8.5.3 swType
- 1.8.5.4 swOtherInformation
- 1.8.5.5 swDependency

1.8.6 hardware

- 1.8.6.1 hwName
- 1.8.6.2 hwType
- 1.8.6.3 hwOtherInformation

1.8.7 environmentExtension

Environment Example: PDF File

TIMELESS BUSINESS ◀ © ▶



environmentCharacteristic =
known to work
environmentPurpose = render

hardware/hwName =
Intel Pentium II
hardware/hwType =
processor

dependency/dependencyName
= Mathematica 5.2
True Type math fonts

software/swName =
Adobe Acrobat Reader
software/swVersion = 6.1
software/swType = renderer
software/swDependency =
Windows NT

software/swName =
Windows NT
software/swVersion = 5.0
software/swType =
operatingSystem



❑ Too specific:

- ❖ Environments are too complex to handle in an object repository
- ❖ Need to link to external registry
- ❖ Need to be first class entities

❑ OAIS focus on Object:

- ❖ May want to only describe Environments, not Objects - Registry
- ❖ Creating Applications are Environments



- ❑ **Too redundant:** Environments are rarely specific to a single object. Repeating each environment for a dedicated object can be unnecessarily verbose.
- ❑ **Too cumbersome to manage:** Environments descriptions evolve. Managing the updates can become cumbersome when the information is redundantly spread across different objects.



❑ **Not generic enough: Environments**

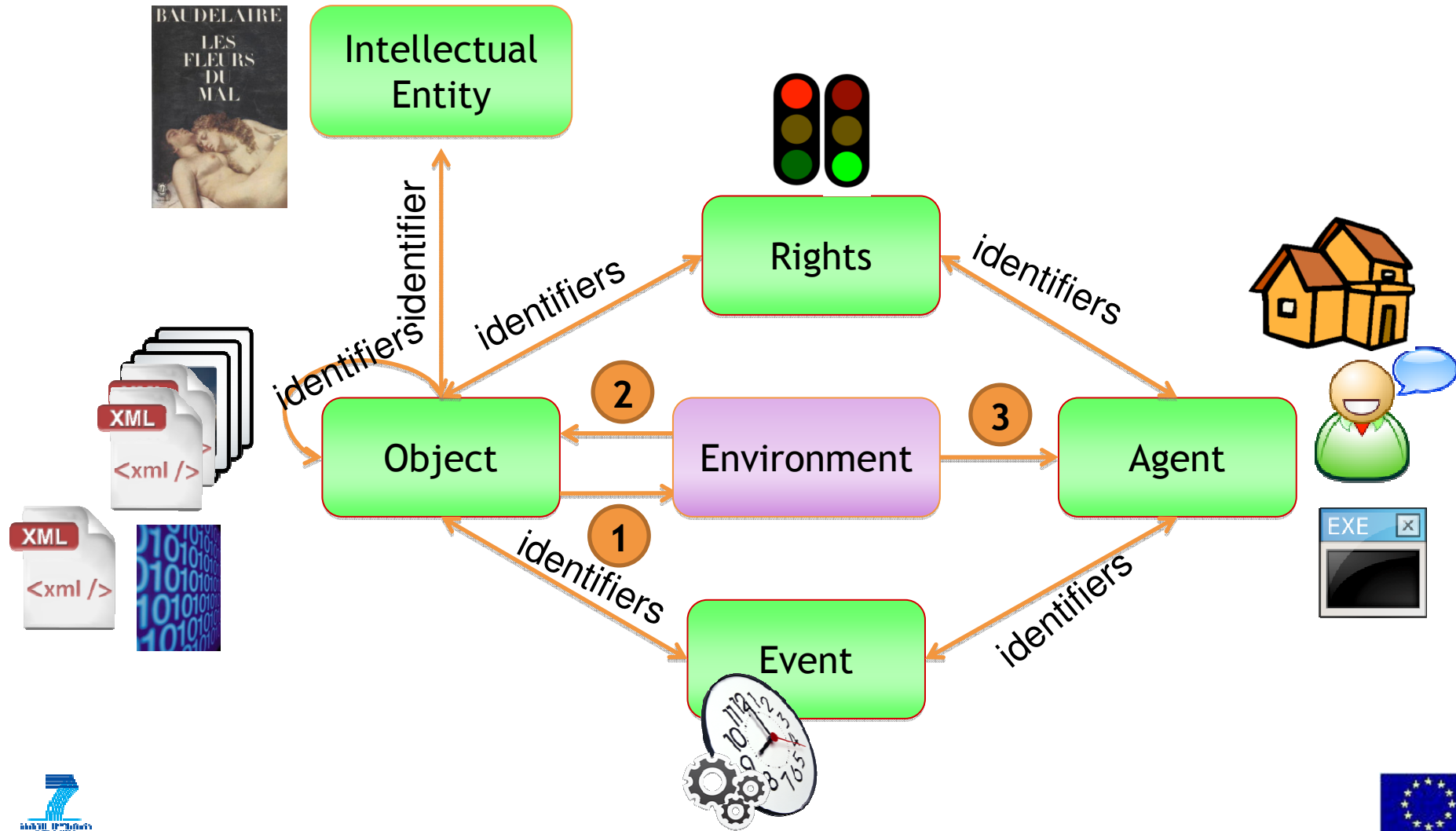
- ❖ Can be related to Objects
- ❖ Can be Objects
- ❖ Can be software Agents

The PREMIS Data Model

TIMELESS BUSINESS ◀ ◉ ▶



Slide adapted from S. Peyrard





❑ **Not generic enough: Environments**

- ❖ Can be related to Objects
- ❖ Can be Objects
- ❖ Can be software Agents

❑ **Environments may link to other Environments**

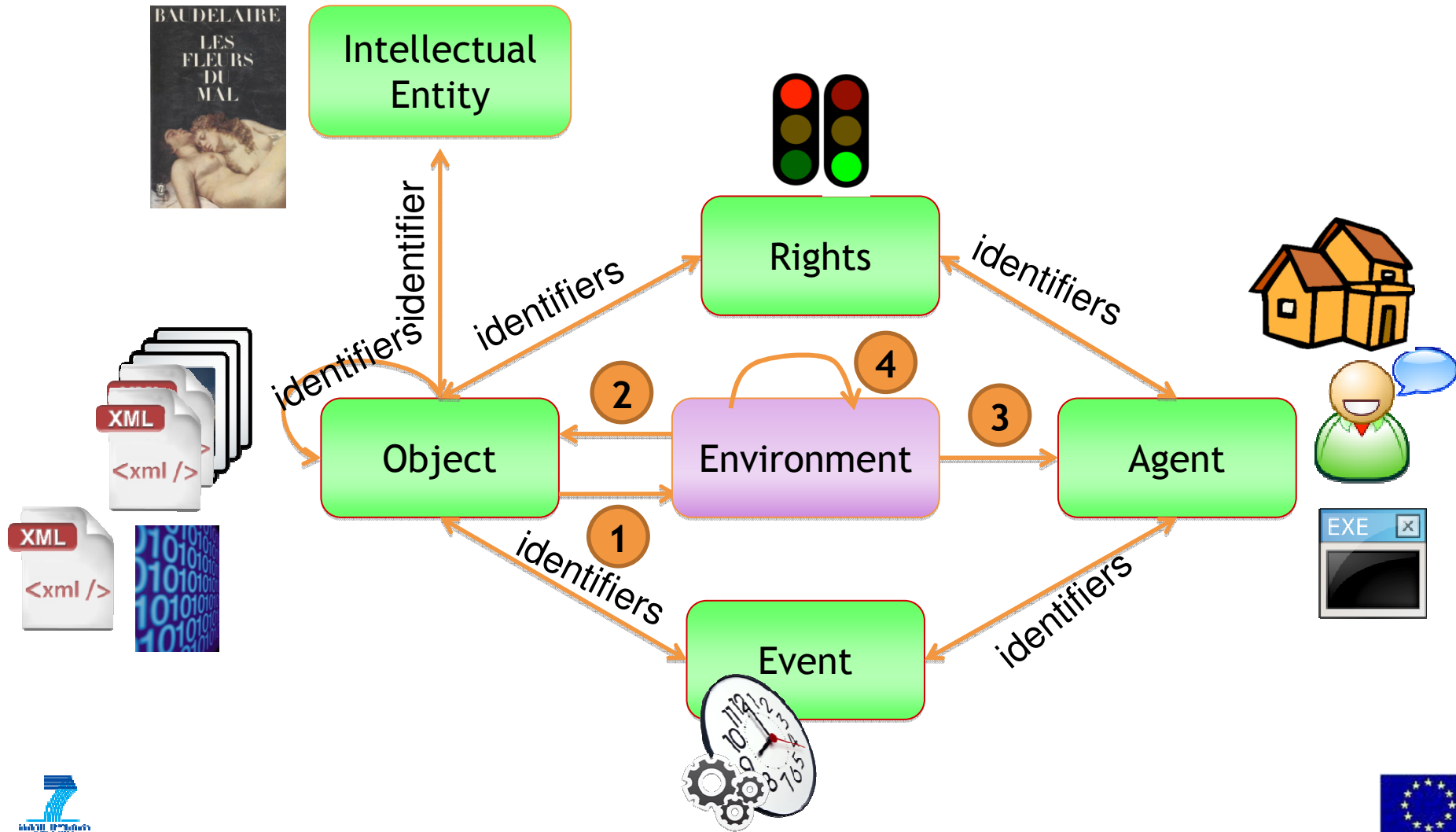
- ❖ E.g. software application linking to its hardware platform

The PREMIS Data Model

TIMELESS BUSINESS ◀ ◉ ▶



Slide adapted from S. Peyrard





❑ **Not generic enough: Environments**

- ❖ Can be related to Objects
- ❖ Can be Objects
- ❖ Can be software Agents

❑ **Environments may link to other Environments**

- ❖ E.g. software app linking to hardware platform

❑ **Environments may link to Events:**

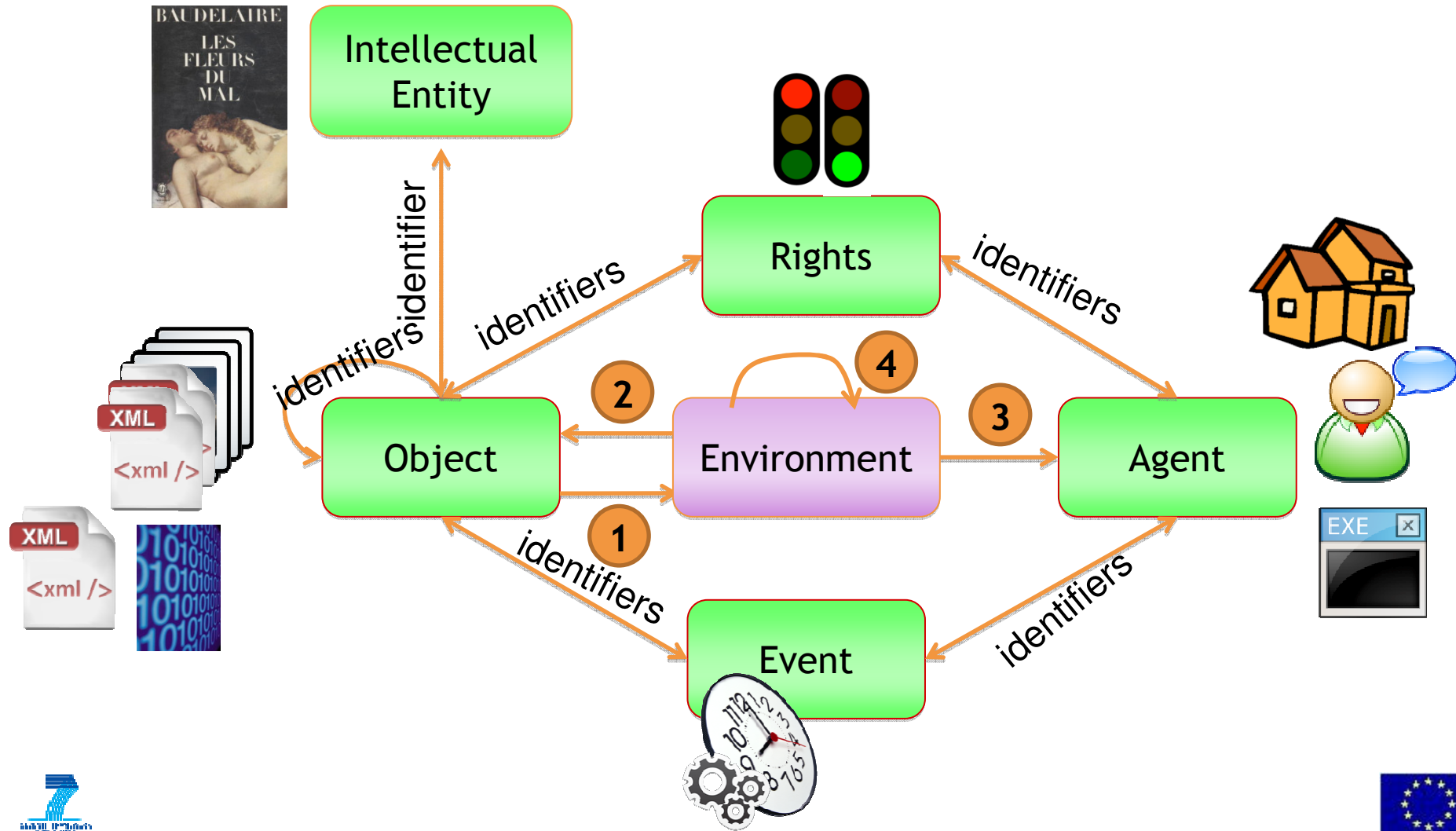
- ❖ Creation, adding memory, ...
- ❖ Environments may need to be versioned
- ❖ Role as object

The PREMIS Data Model

TIMELESS BUSINESS ◀ ● ▶



Slide adapted from S. Peyrard



PREMIS Environment Working Group

TIMELESS BUSINESS ◀ ◉ ▶



- ❑ Environments may be generic or instances
- ❑ Environments may be tools or services
- ❑ Virtual machines blur the distinction between software and hardware

Overview: Current Research Activities

TIMELESS BUSINESS   



□ A vision:

❖ Processes and their execution environments:

- 🖨️ The TIMBUS Project

□ The building blocks:

❖ OAIS repository objects:

- 🖨️ PREMIS data dictionary;
- 🖨️ PREMIS Environment working group;
- 🗄️ file format work (Jhove, Pronom, UDFR, LoC, etc.);

❖ Software preservation:

- 🖨️ The Significant Properties of Software: A Study;
- 🖨️ Preserving Virtual Worlds;
- 🖨️ SWOP; 🖨️ DOAP;
- 🗄️ NSRL National Software Reference Library; 🗄️ AMINET

❖ Virtualized infrastructure:

- 🖨️ VRDF; 🖨️ TIMBUS Dependency Analysis; 🖨️ CDMI; 🖨️ WSDL

❖ Computing environments: 🗄️ KEEP Totem;

The Significant Properties of Software: A Study

TIMELESS BUSINESS ◀ ○ ▶



http://www.jisc.ac.uk/media/documents/programmes/preservation/spssoftware_report_redacted.pdf

- Functionality**
- Software Composition**
- Provenance and Ownership**
- User Interaction**
- Software Environment**
- Software Architecture**
- Operating Performance**

The Significant Properties of Software: A Study

TIMELESS BUSINESS ◀ ○ ▶



	Package	Version	Variant	Download
Functionality				
Provenance and Ownership				
Software Environment				
Software Architecture				
Operating Performance				
Software Composition	software overview tutorials	source manual Installation	binary source	file
User Interaction	requirements	test cases specification	configuration	

Package Properties

TIMELESS BUSINESS ◀ ○ ▶



Property Category	Software Property	
Functionality	purpose	Description of overall functionality of software system
	keyword	Classification of software under a specified controlled vocabulary
Provenance and Ownership	package_name	Name of the package
	owner	Owner of the package, with contact details
	licence	Overall licensing agreement
	location	URL of website of software
Software Environment	-	
Software Architecture	overview	Overview of software architecture
Operating Performance	-	
Software Composition	software overview	Documentation on the overview of the software
	tutorials	Teaching material on the system.
	requirements	requirements of package

Version Properties

TIMELESS BUSINESS ◀ ○ ▶



Property Category	Software Property	
Functionality	functional_description	Description of relationship of between inputs and outputs of the version.
	release_notes	Description of changes of this version from other versions.
	algorithm	Description of the algorithm used.
	input_parameter	Details of names and formats of inputs
	output_parameter	Details of names and formats of outputs
	interface	API description
	error_handling	Description of how errors are handled.
Provenance and Ownership	version_identifier	Identifier for this particular version
	licence	Licence specific to this version.
Software Environment	programming_language	Programming language used for this version.
	hardware_device	Category of hardware device which the software version depends upon.
Software Architecture	detailed_architecture	Detailed description of architectural dependencies of the version.



VariantProperties

TIMELESS BUSINESS ◀ ○ ▶



Property Category	Software Property	
Functionality	variant_notes	Description of the variations in behaviour specific to this variant.
Provenance and Ownership	licence	Licence specific to this variant.
Software Environment	platform	Target hardware machine architecture of version.
	operating_system	Version of operating system
	compiler	Version of compiler used to construct this variant.
	dependent_library	Version of dependent software libraries used.
	hardware_device	Specific auxiliary hardware devices supported by the variant.
Software Architecture	dependent_package	Dependency on another software package being installed.



Download Properties

TIMELESS BUSINESS ◀ ○ ▶

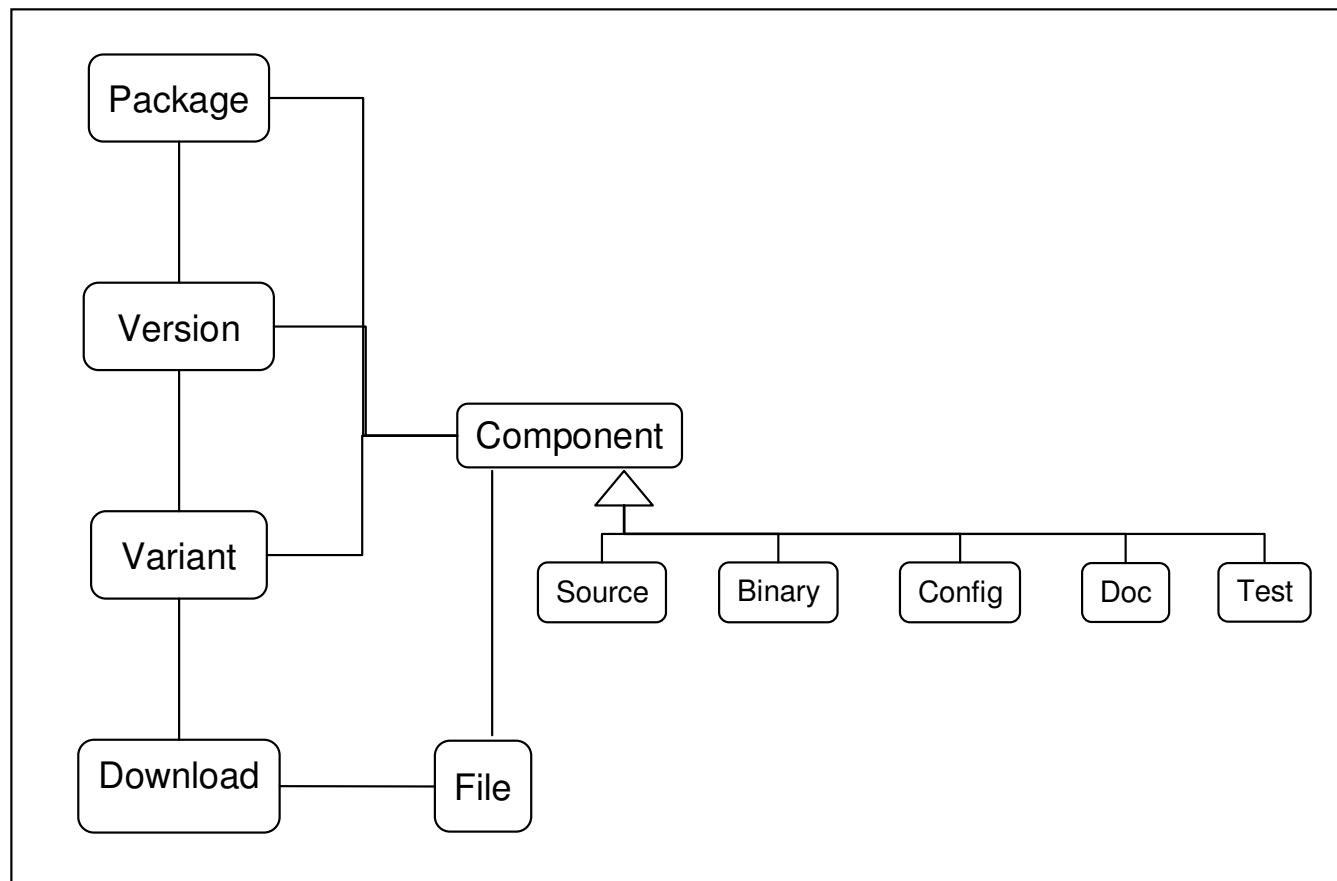


Property Category	Software Property	
Functionality	-	
Provenance and Ownership	licensee	Named licensee of the download
	conditions	Local conditions of use of this download.
	licence_code	Licence key value
Software Environment	environment_variable	Specific settings for environmental variables.
	IP_address	Specific IP address
	hardware_address	Specific MAC address (or equivalent) identifying a specific machine.
Software Architecture	-	
Operating Performance	-	
Software Composition	file	Names and addresses of specific files in the download.



The Significant Properties of Software: A Study

TIMELESS BUSINESS ◀ ◉ ▶



The Software Component Conceptual Model

Overview: Current Research Activities

TIMELESS BUSINESS   



□ A vision:

❖ Processes and their execution environments:

- 🖨️ The TIMBUS Project

□ The building blocks:

❖ OAIS repository objects:

- 🖨️ PREMIS data dictionary;
- 🖨️ PREMIS Environment working group;
- 🗄️ file format work (Jhove, Pronom, UDFR, LoC, etc.);

❖ Software preservation:

- 🖨️ The Significant Properties of Software: A Study;
- 🖨️ Preserving Virtual Worlds;
- 🖨️ SWOP; 🖨️ DOAP;
- 🗄️ NSRL National Software Reference Library; 🗄️ AMINET

❖ Virtualized infrastructure:

- 🖨️ VRDF; 🖨️ TIMBUS Dependency Analysis; 🖨️ CDMI; 🖨️ WSDL

❖ Computing environments: 🗄️ KEEP Totem;

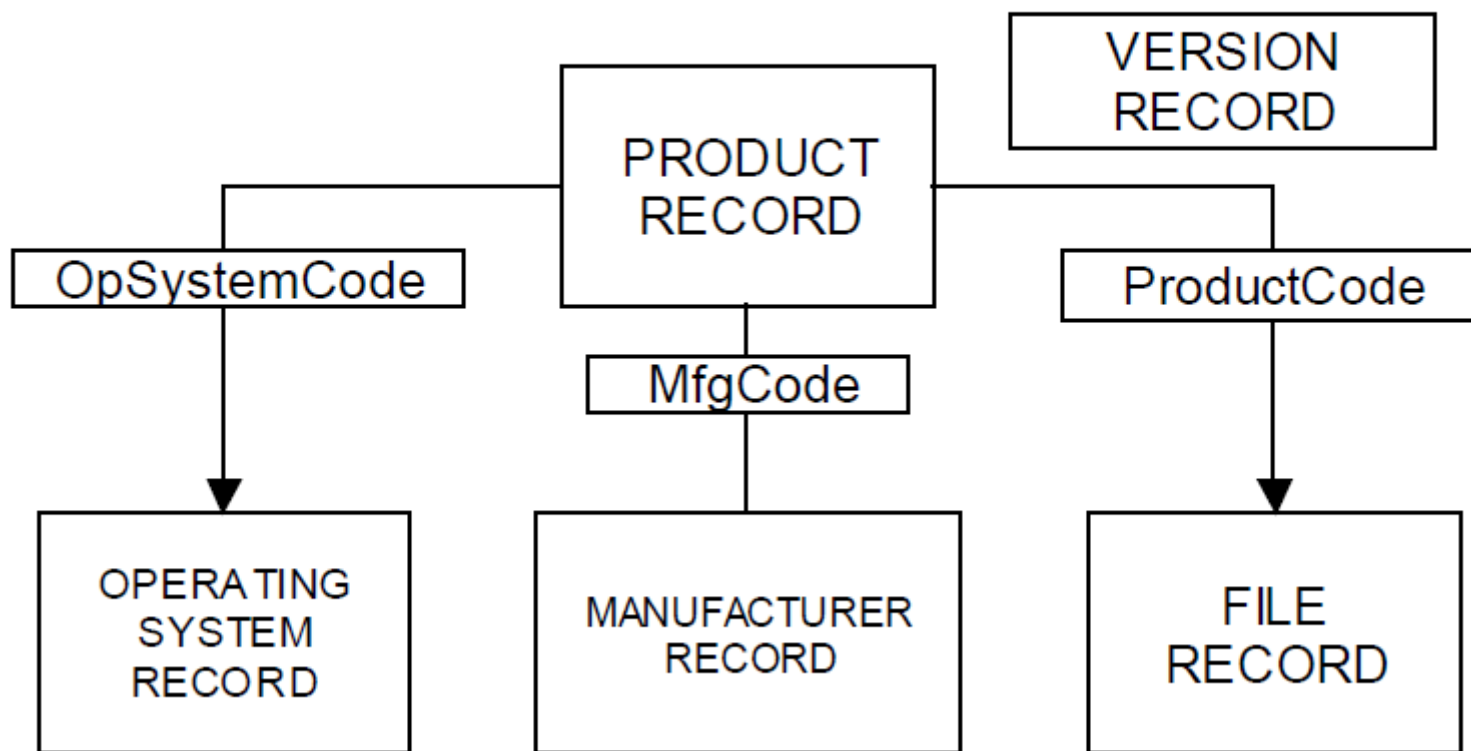


<http://www.nsrl.nist.gov/Documents/Data-Formats-of-the-NSRL-Reference-Data-Set-16.pdf>

- Typical uses for this data set include the forensic examination of computers and stored data seized
- For law enforcement agencies, copyright infringement investigations, and similar types of functions.



Figure 1. NSRL RDS Logical Record Relationships



Overview: Current Research Activities

TIMELESS BUSINESS   



□ A vision:

❖ Processes and their execution environments:

- 🖨️ The TIMBUS Project

□ The building blocks:

❖ OAIS repository objects:

- 🖨️ PREMIS data dictionary;
- 🖨️ PREMIS Environment working group;
- 🗄️ file format work (Jhove, Pronom, UDFR, LoC, etc.);

❖ Software preservation:

- 🖨️ The Significant Properties of Software: A Study;
- 🖨️ Preserving Virtual Worlds;
- 🖨️ SWOP; 🖨️ DOAP;
- 🗄️ NSRL National Software Reference Library; 🗄️ AMINET

❖ Virtualized infrastructure:

- 🖨️ VRDF; 🖨️ TIMBUS Dependency Analysis; 🖨️ CDMI; 🖨️ WSDL

❖ Computing environments: 🗄️ KEEP Totem;

Virtual Resource Description

Framework (VRDF)

TIMELESS BUSINESS 



http://www.caida.org/workshops/wide-casfi/1004/slides/wide-casfi1004_ykadobayashi.pdf

Context: cloud computing

- ❖ Virtualized infrastructure
- ❖ Sheer scale and extensive use of virtualization will make it almost intractable

By Youki Kadobayashi, NAIST

Virtual Resource Description Framework (VRDF) Objectives

TIMELESS BUSINESS 



- Develop a framework to describe and analyze complex dependency of services, virtual machines, virtual routers and VLANs
 - ❖ ... in order to improve the availability of Data Centers and maintain their security level
- Describe dependency of VM to physical machine
- Describe dependency of virtual nets to networking devices
- Assess the impact of failure
- Analyze dependency

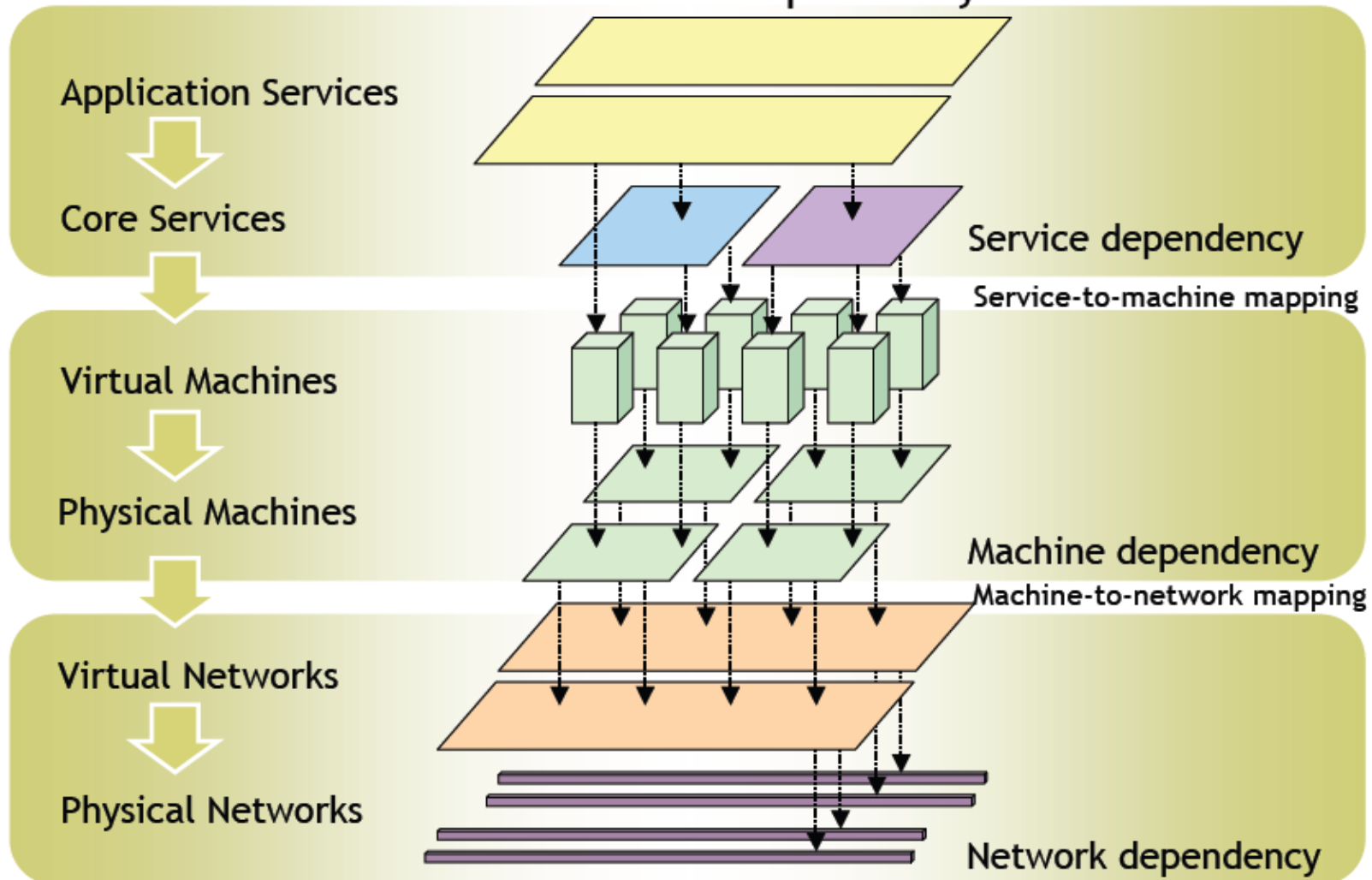
By Youki Kadobayashi, NAIST

Virtual Resource Description Framework (VRDF)

TIMELESS BUSINESS



VRDF: 3 tiers of dependency



Virtual Resource Description Framework (VRDF) Objectives

TIMELESS BUSINESS 



- Enumeration of virtual & physical resources
- Description of dependency
- RDF schema
- Initial attempt to use rule engines e.g., SWRL

By Youki Kadobayashi, NAIST



27 January 2012

timbusproject.net © 2011

45



Overview: Current Research Activities

TIMELESS BUSINESS   



□ A vision:

❖ Processes and their execution environments:

- 🖨️ The TIMBUS Project

□ The building blocks:

❖ OAIS repository objects:

- 🖨️ PREMIS data dictionary;
- 🖨️ PREMIS Environment working group;
- 🗄️ file format work (Jhove, Pronom, UDFR, LoC, etc.);

❖ Software preservation:

- 🖨️ The Significant Properties of Software: A Study;
- 🖨️ Preserving Virtual Worlds;
- 🖨️ SWOP; 🖨️ DOAP;
- 🗄️ NSRL National Software Reference Library; 🗄️ AMINET

❖ Virtualized infrastructure:

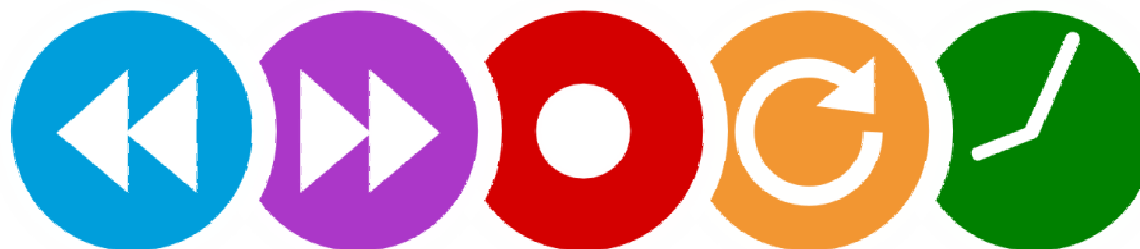
- 🖨️ VRDF; 🖨️ TIMBUS Dependency Analysis; 🖨️ CDMI; 🖨️ WSDL

❖ Computing environments: 🗄️ KEEP Totem;



Thank you!

TIMBUS



TIMELESS BUSINESS

