TIMBUS

TIMELESS BUSINESS

Introduction to the digital preservation of processes and contexts

Angela Dappert

Digital Preservation Coalition

From Preserving Data to Preserving Research: Curation of Process and Context

TIMELESS BUSINESS

TIMBUS

TIMELESS BUSINESS PROCESSES AND SERVICES

Workflow 4Ever

SCAPE

SCAlable Preservation Environments

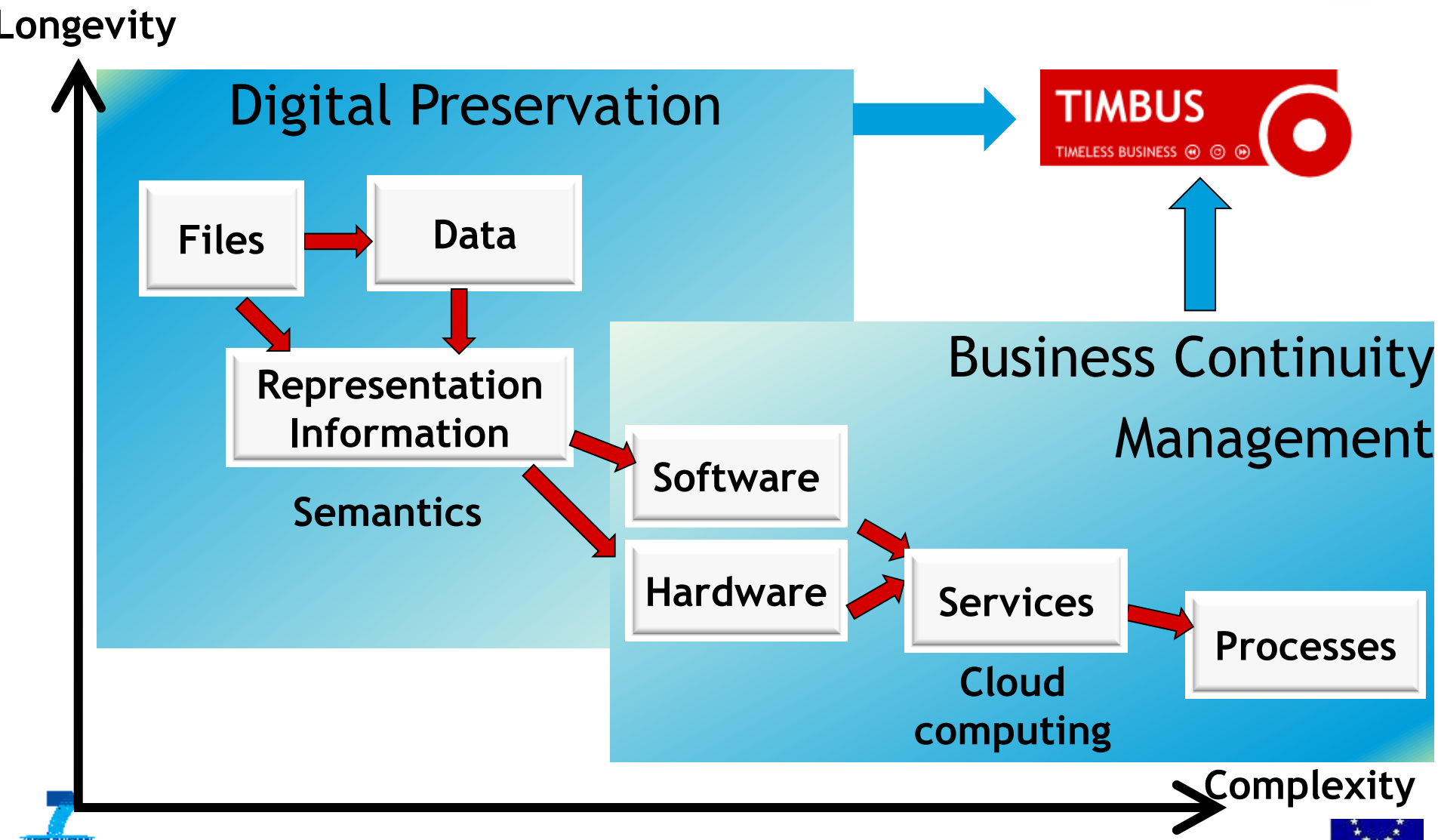
Agenda

TIMELESS BUSINESS

14:30 – 15:00 Angela Dappert	Welcome and House Keeping Introduction to the digital preservation of processes and contexts
15:00 – 15:45 Rudolf Mayer	Session 1: Preservation of Processes
15:45– 16:05 Kevin Page	Session 2a: Describing Experiments via Research Objects Introduction to Research Objects
16:05 – 16:20	Coffee Break
16:20 – 16:45 Raul Palma	Session 2b: Research Object Demo Research Object management tools
16:45– 17:30 Stefan Pröll	Session 3: Data Citation
17:30 – 17:45 Catherine Jones	Guest speaker: Research Context Preservation in SCAPE
17:45-18:00 William Kilbride	Panel discussion

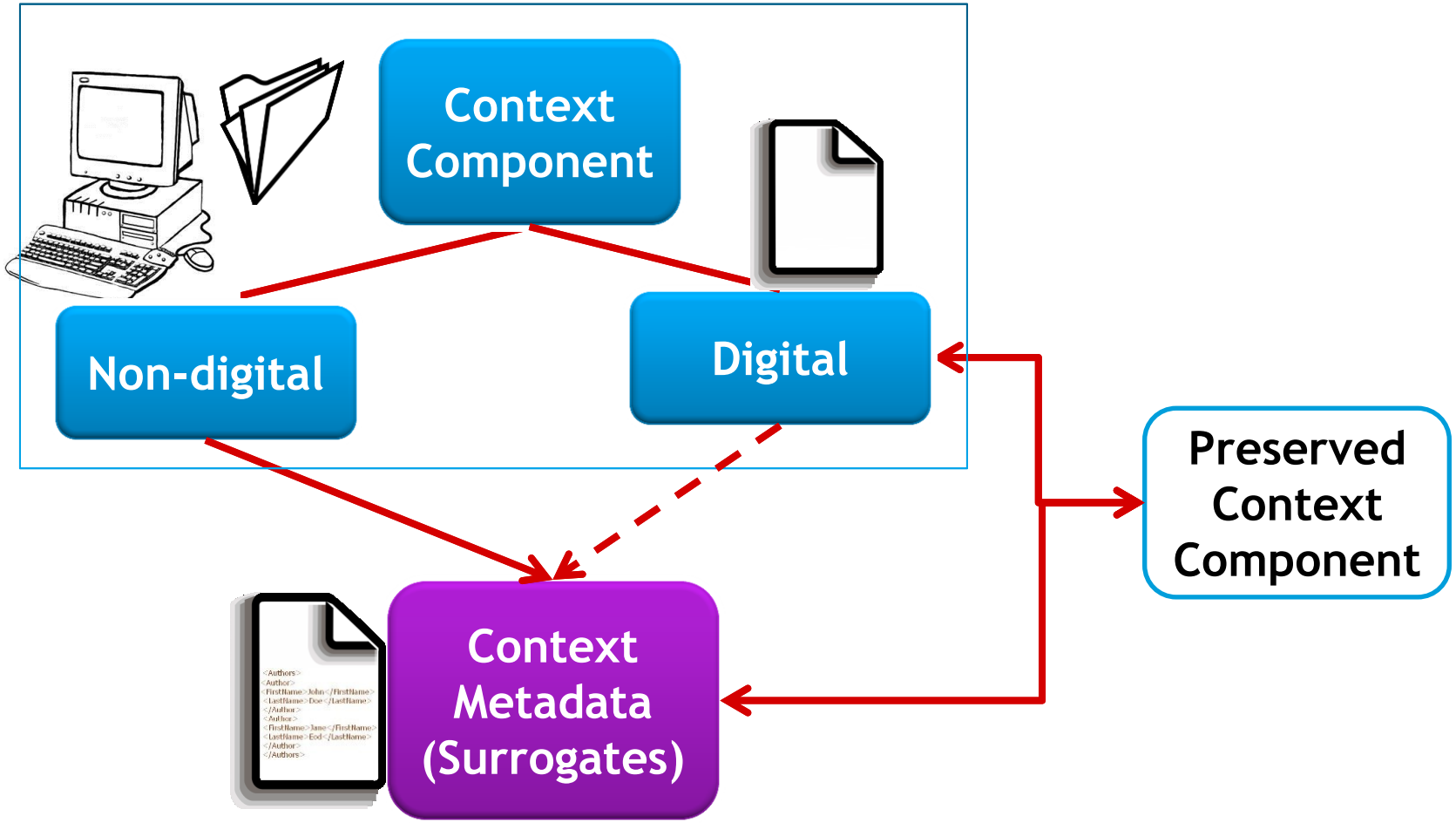
A Preservation Continuum

TIMELESS BUSINESS



Context Components and Surrogates

TIMELESS BUSINESS



TIMBUS Approach

TIMELESS BUSINESS

- Methods, processes, architecture and tools
  - ensure continued access to
  - processes and
  - supporting services and infrastructure.

- Align
  - digital preservation (DP) with
  - enterprise risk management (ERM) and
  - business continuity management (BCM).

- Explore Digital Preservation from a BCM perspective.

ations 1

e information:

industries: fully document processes

software and processes that produce data

- **Analysis of processes:**

ations 1

- **Provenance information:**
  - Evidence for litigation
  - Prove authenticity
  - Prove quality of process products

ations 1

industries: fully document processes
liance
, reproduce, or diagnose

ations 1

software and processes that produce data

- Assess validity of data & derived scientific claims
- Credit assignment

ations 1

- **Analysis of processes:**
  - Continuous improvement

Motivations 2

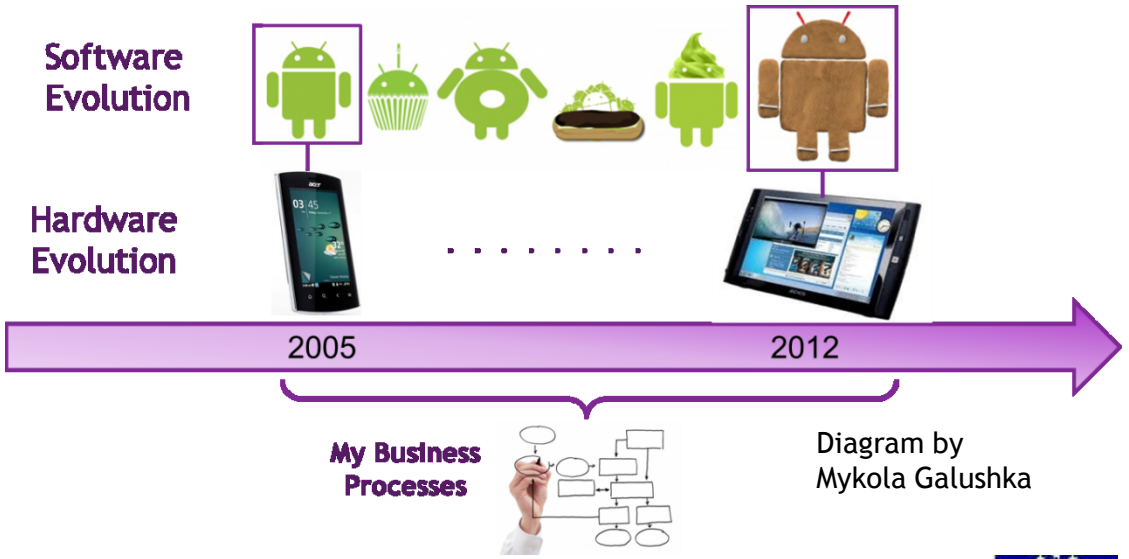
- **Service and licensing models:**
- **Changes in technical environments:**
- **Staff changes:**

Motivations 2

- **Service and licensing models:**
  Declining popularity of centralized in-house services.
  Increasing popularity of SaaS, PaaS, (*aaS), and IoS.
  - ➢ Protect data and functionality - escrow service

Motivations 2

- **Changes in technical environments:**
  - Manage services across platforms



Motivations 2

- **Staff changes:**
  - Knowledge retention

Vision

Slide by Mykola Galushka

TIMELESS BUSINESS

Processes

Preservation of Processes

Redeployment of Processes

New OS

New Device

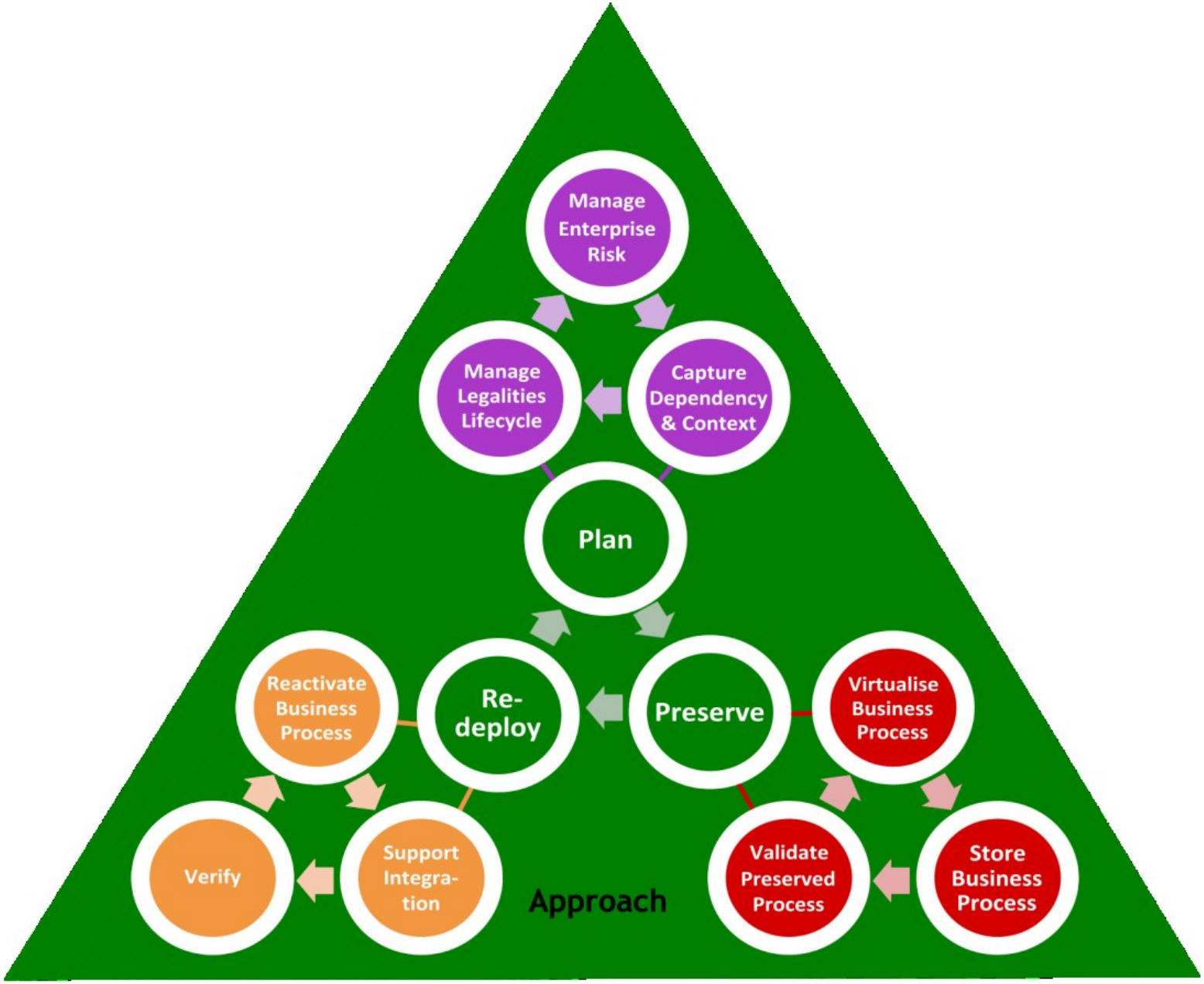
Processes

Diagram by Mykola Galushka

Goals

- **Process preservation:**
  URIs/DOIs for referencing resources
- **Reusability** of any part of the process
- **Repeatability /Reproducibility:**
  Redeployment of the process
- **Traceability** and error detection
- **Attribution:**
  Able to cite data and publications of the process
- **Understandability:**
  Links between data, results and annotations
- **Curation:**
  By explicitly exposing the methods used in the process

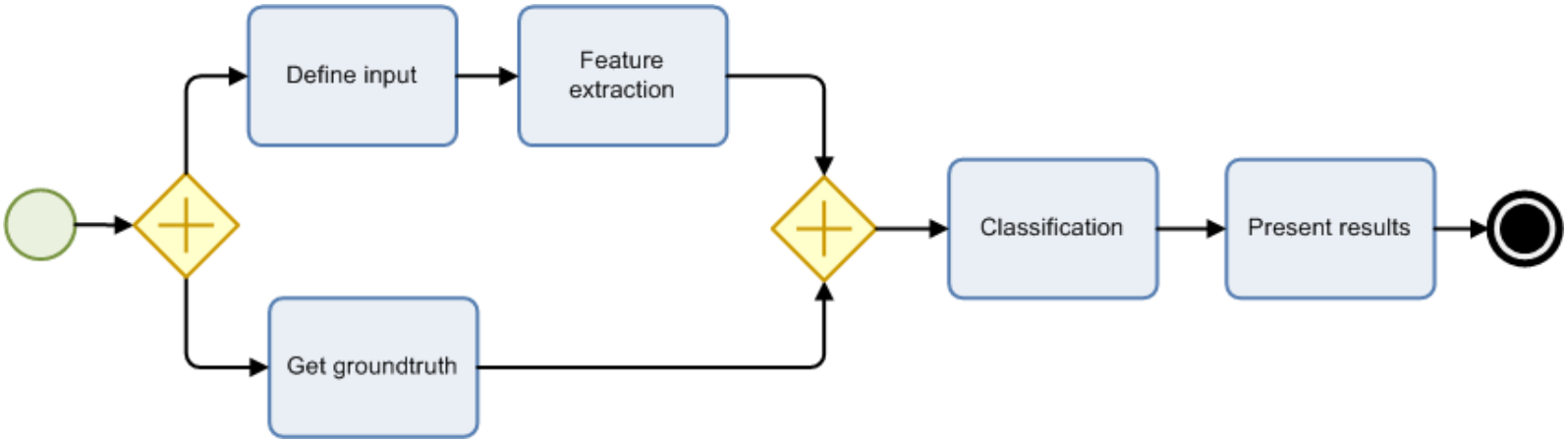
Daniel Garijo


Manage Enterprise Risk
Capture Dependency & Context
Manage Legalities Lifecycle
Plan
Preserve
Virtualise Business Process
Store Business Process
Validate Preserved Process
Re-deploy
Reactivate Business Process
Support Integra-tion
Verify
Approach

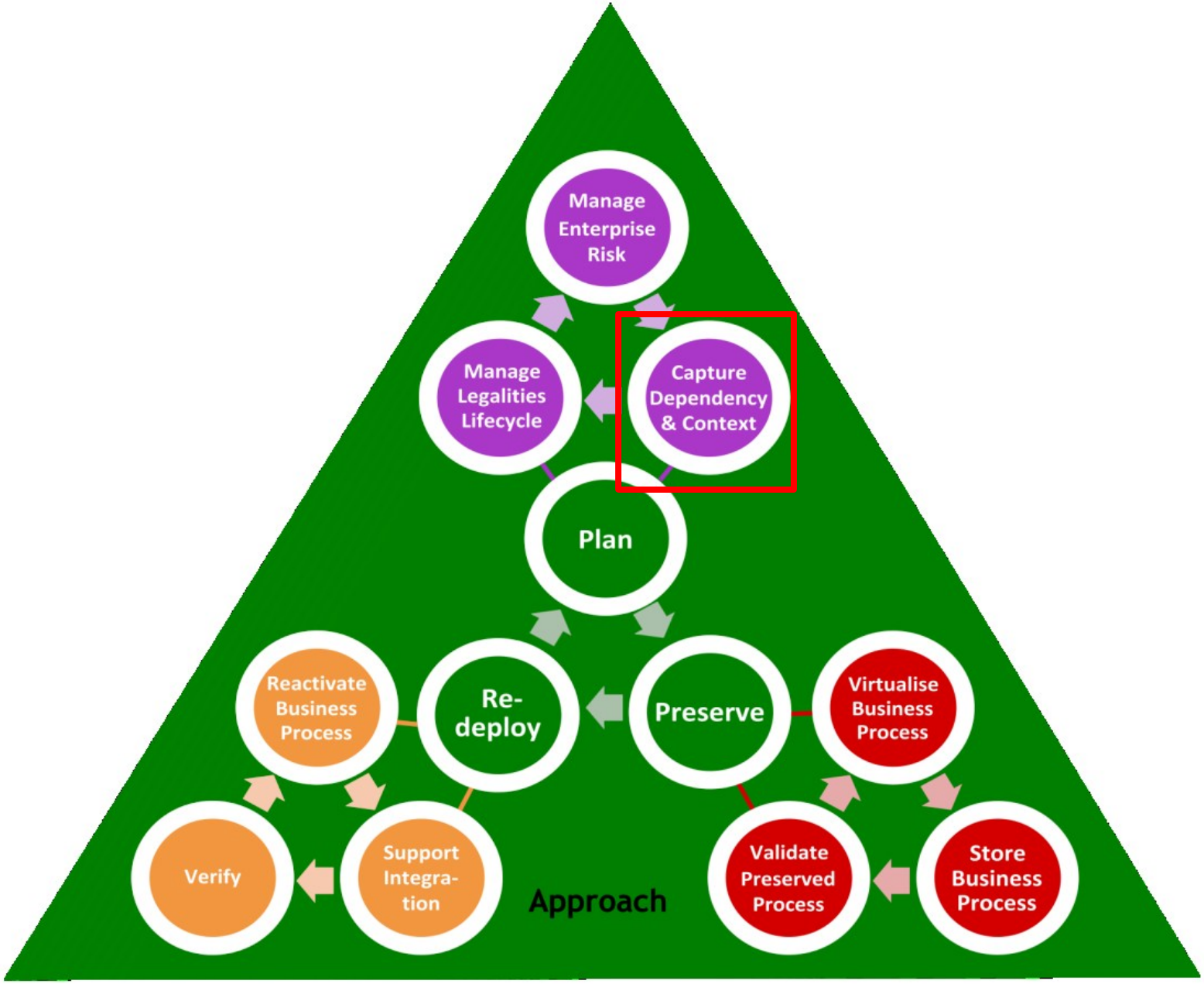
Example: Music Classifcation Workflow

Slide by Rudolf Mayer

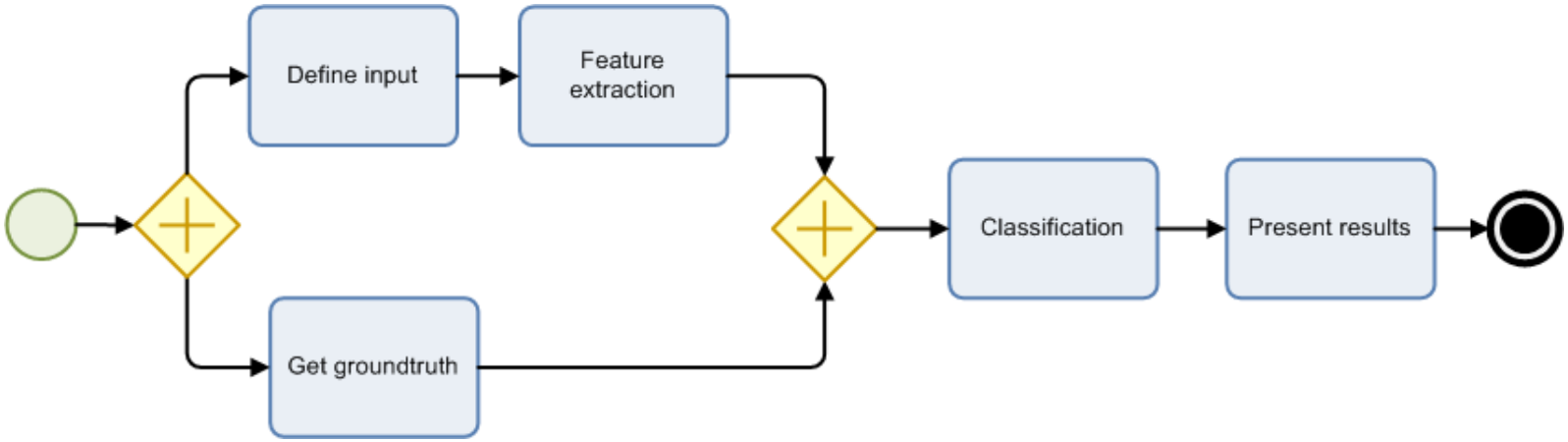
TIMELESS BUSINESS



- Workflow: classification of music into predefined set of genres
- Learns a machine-learning model from given training data (i.e. data with manually assigned class/genre)
- Predicts genre for previously unseen data


Manage Enterprise Risk
Manage Legalities Lifecycle
Capture Dependency & Context
Plan
Reactivate Business Process
Re-deploy
Preserve
Virtualise Business Process
Verify
Support Integra-tion
Approach
Validate Preserved Process
Store Business Process

Example: Music Classifcation Workflow

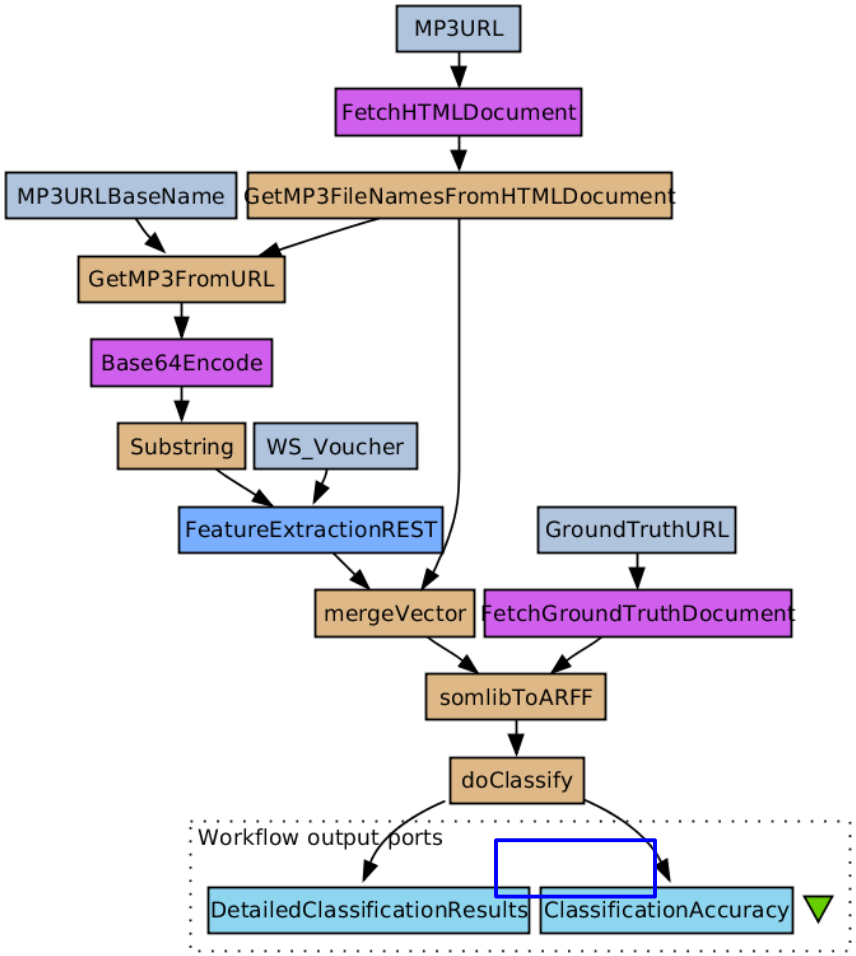


- ArchiMate
- BPMN
- Petri Nets
- Workflow Engines – for purely digital workflows

Process Context Capture

Slide based on Alex Neumann's slide

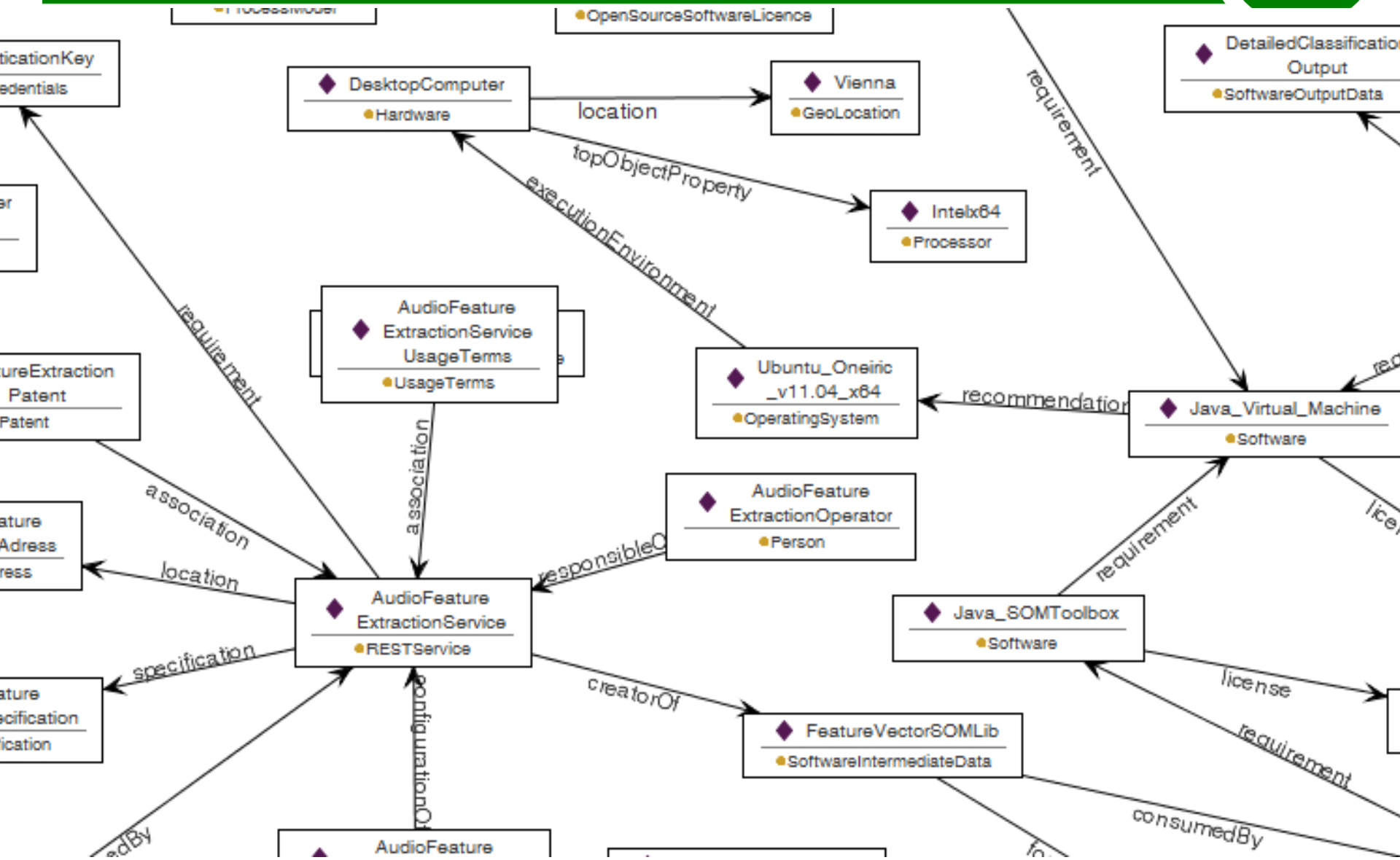
TIMELESS BUSINESS



- Process Specification
  - Inputs and Outputs
  - Dependencies
- Data
  - URLs, Files, Documents, Streams, Constants, Classifier, Classifications
- Services (internal, external)
- Software
  - Version, run-time environment
  - Platform (Taverna)
  - Libraries (WEKA, SOMLib), Scripts, DLLs
- Specifications
  - ARFF, REST, HTTP, HTML, MP3, Algorithms

Dependencies

TIMELESS BUSINESS



ations 1

e information:
nce for litigation
authenticity
quality of process products

industries: fully document processes
liance
, reproduce, or diagnose

software and processes that produce data
- ➢ Assess validity of data & derived scientific claims
- ➢ Credit assignment

- **Analysis of processes:**
  - ➢ Continuous improvement

Process Context Capture

TIMELESS BUSINESS

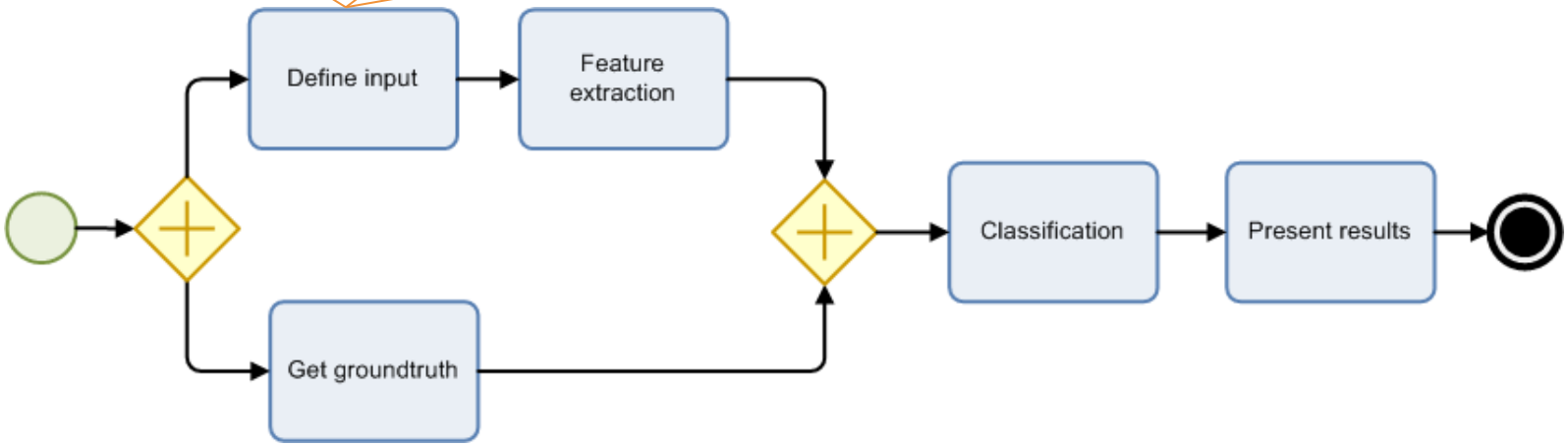
Sampling methods
Test sets used
Test methods
Exception treatment
Processes and logs
Quality criteria

Requirements

Policy

Goal: Process Diagnosis

R3:
Trainings sets were poorly chosen



Process Context Model

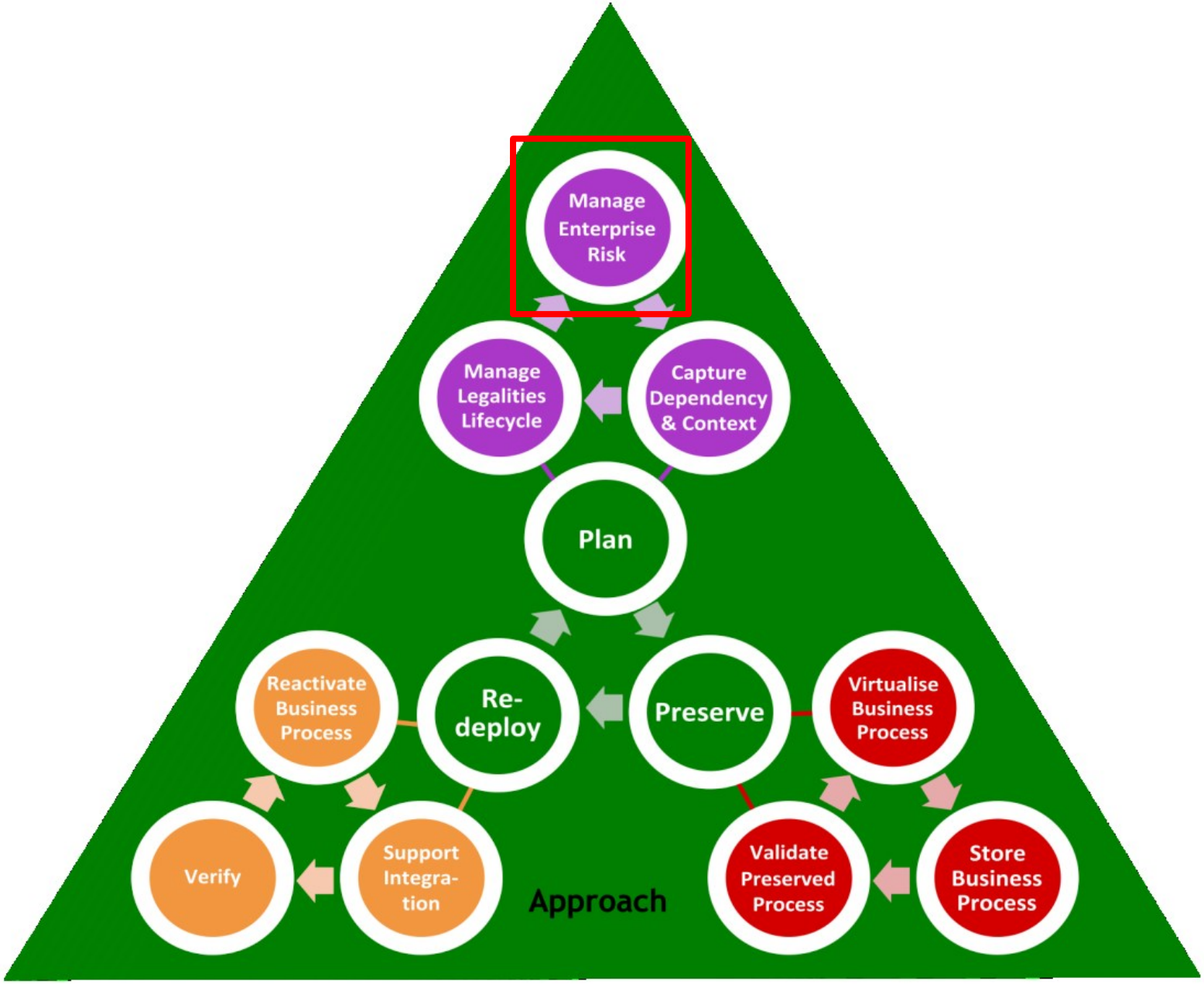
- Process Context is based on a Process Context Model
  - A formal meta-model
  - Can be instantiated
  - Enables process redeployment
  - Captures the relevant aspects of a process and supporting software/technology
  - Relevance is established through risk, impact, value assessment -> significant characteristics
  - Sufficiency relative to significant characteristics

Zachman Framework

TIMELESS BUSINESS

Data Function Systems People Time Motivation

ENTERPRISE ARCHITECTURE : ZACHMAN FRAMEWORK						
Focus / Perspective	What	How	Where	Who	When	Why
Scope of Business						
Business Model						
System Model						
Technology Model						
Out of Context						
Executable System						

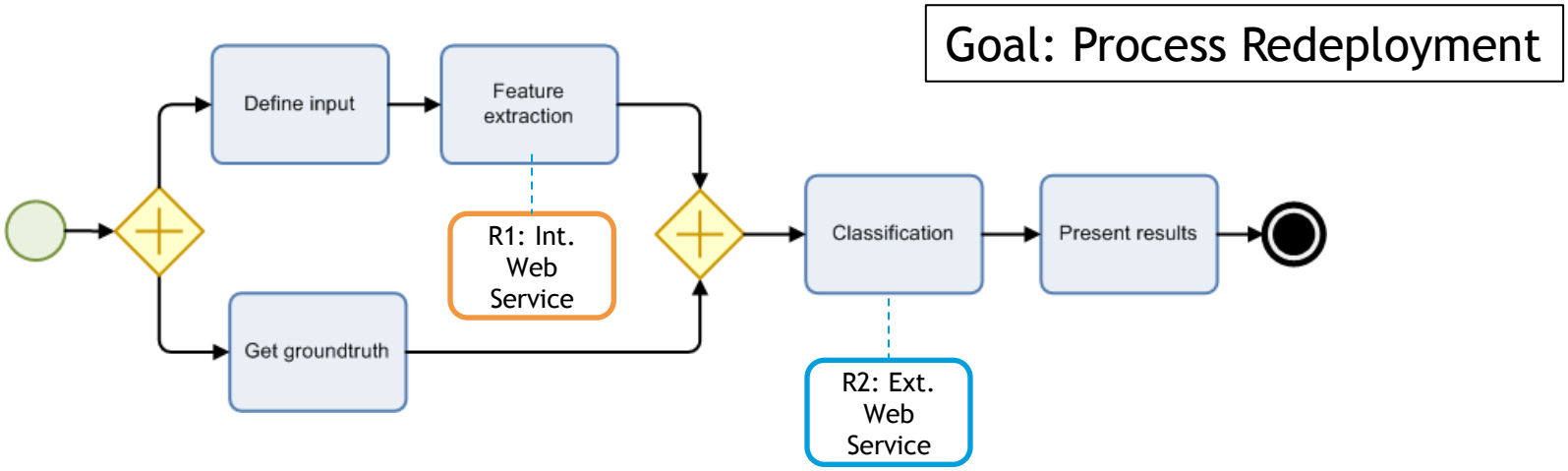

Manage Enterprise Risk
Manage Legalities Lifecycle
Capture Dependency & Context
Plan
Reactivate Business Process
Re-deploy
Preserve
Virtualise Business Process
Verify
Support Integra-tion
Approach
Validate Preserved Process
Store Business Process

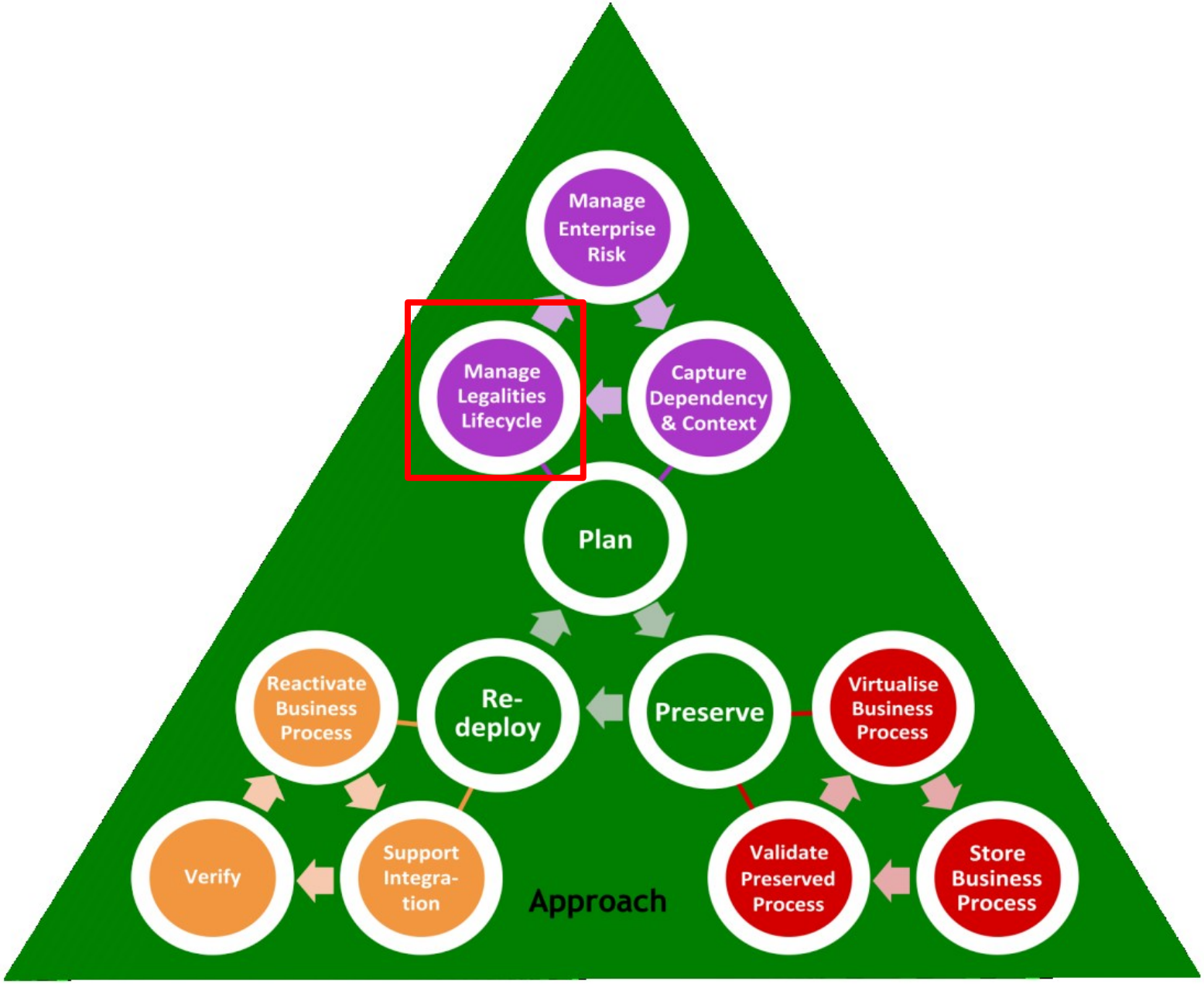
Enterprise Risk Management

Slide by Rudolf Mayer

TIMELESS BUSINESS

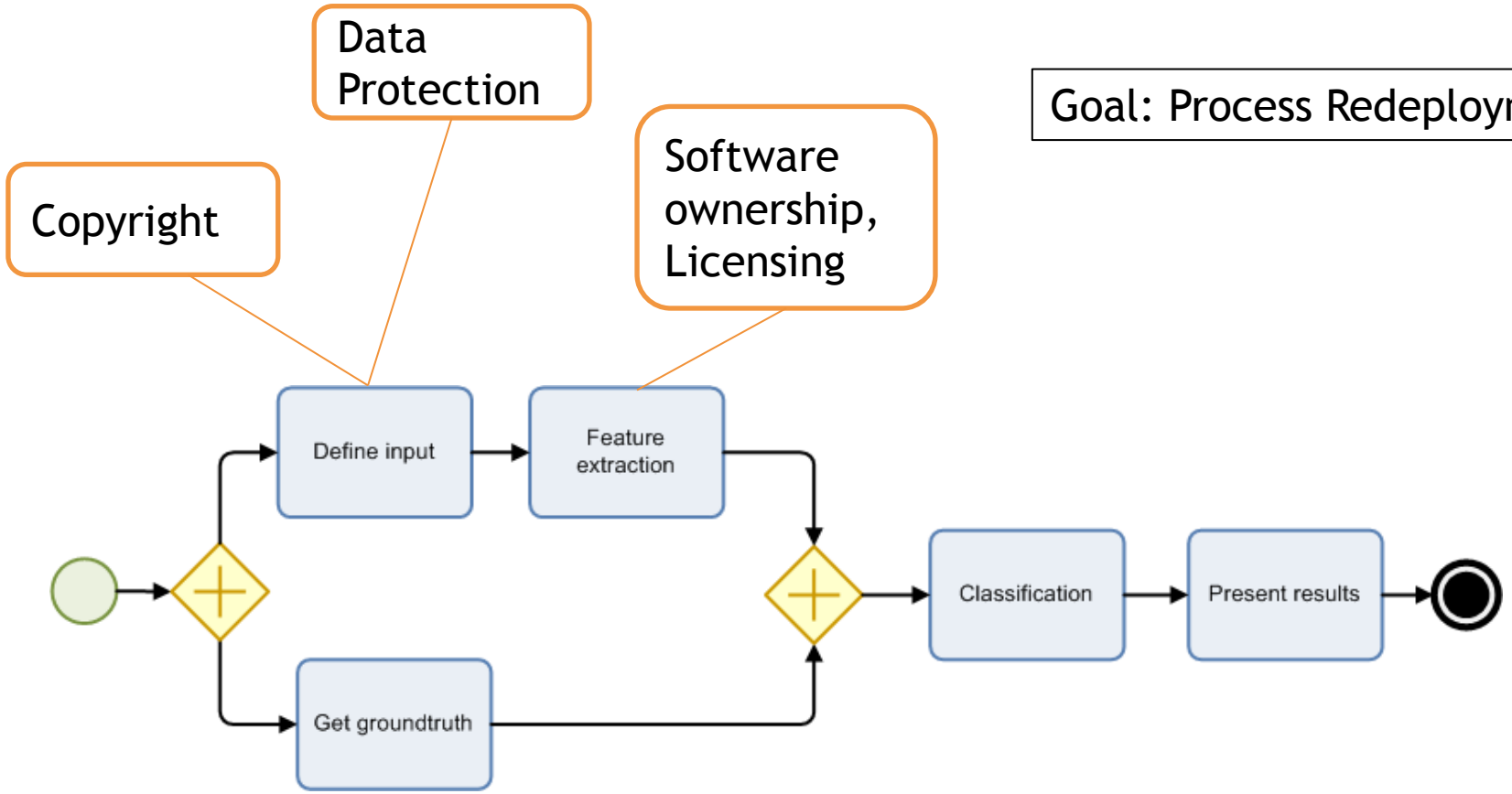
- Identification of risk that might affect the IT applications of the “Music Classification Process”:
  - R1: Obsolescence of application providing internal web service
  - R2: Bankruptcy of external web service provider

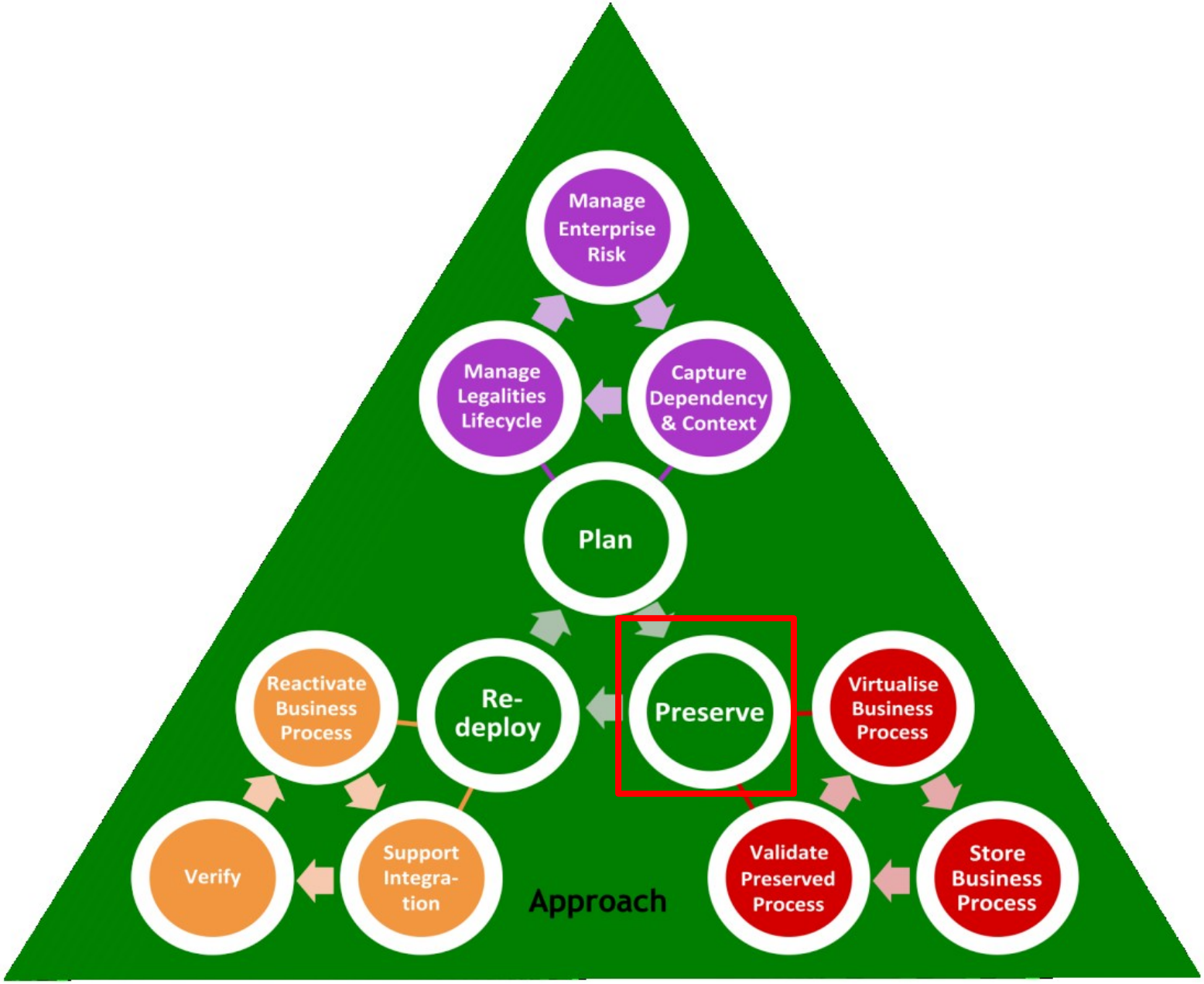



Manage Enterprise Risk
Manage Legalities Lifecycle
Capture Dependency & Context
Plan
Reactivate Business Process
Re-deploy
Preserve
Virtualise Business Process
Verify
Support Integra-tion
Approach
Validate Preserved Process
Store Business Process

Legalities Lifecycle Management

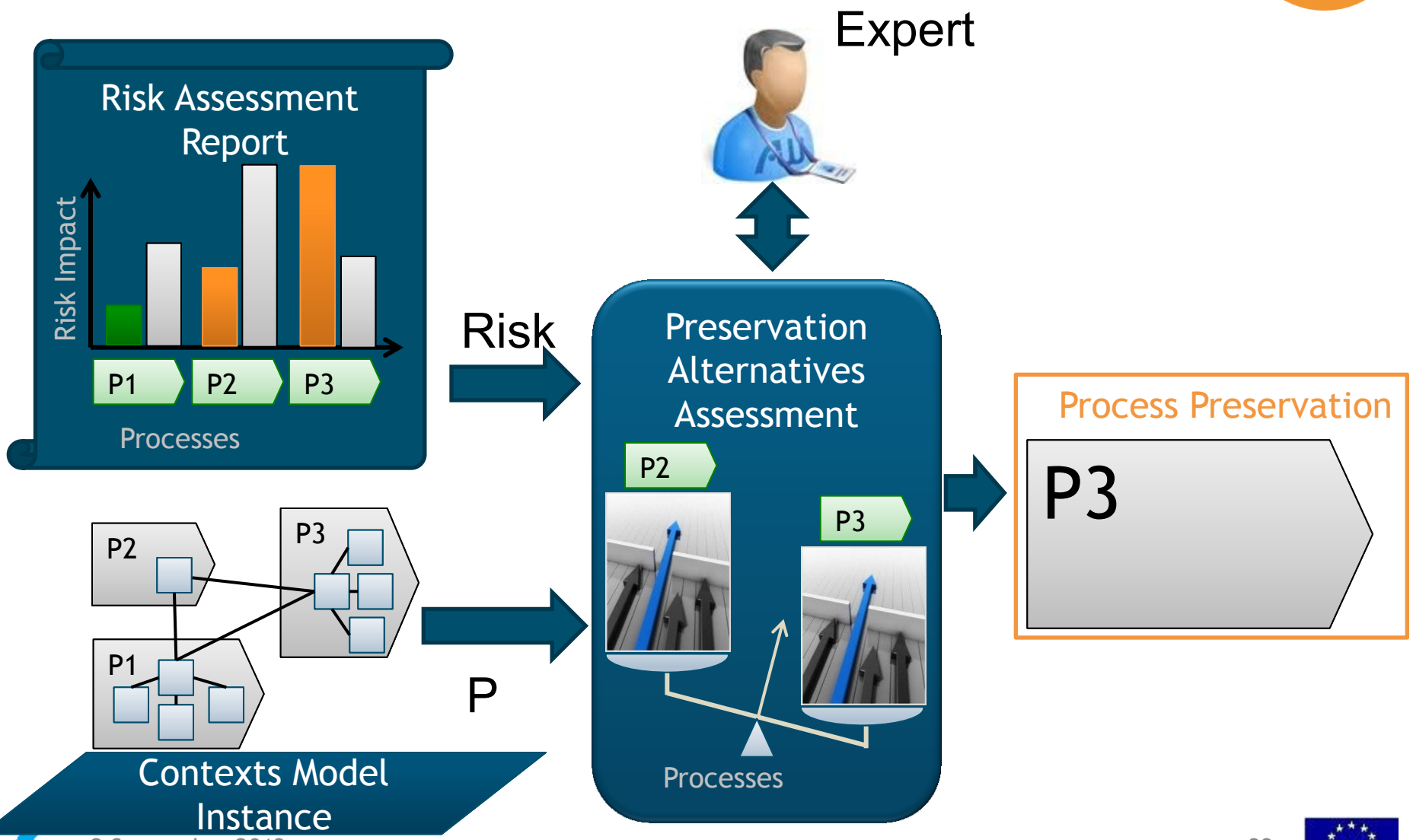
TIMELESS BUSINESS

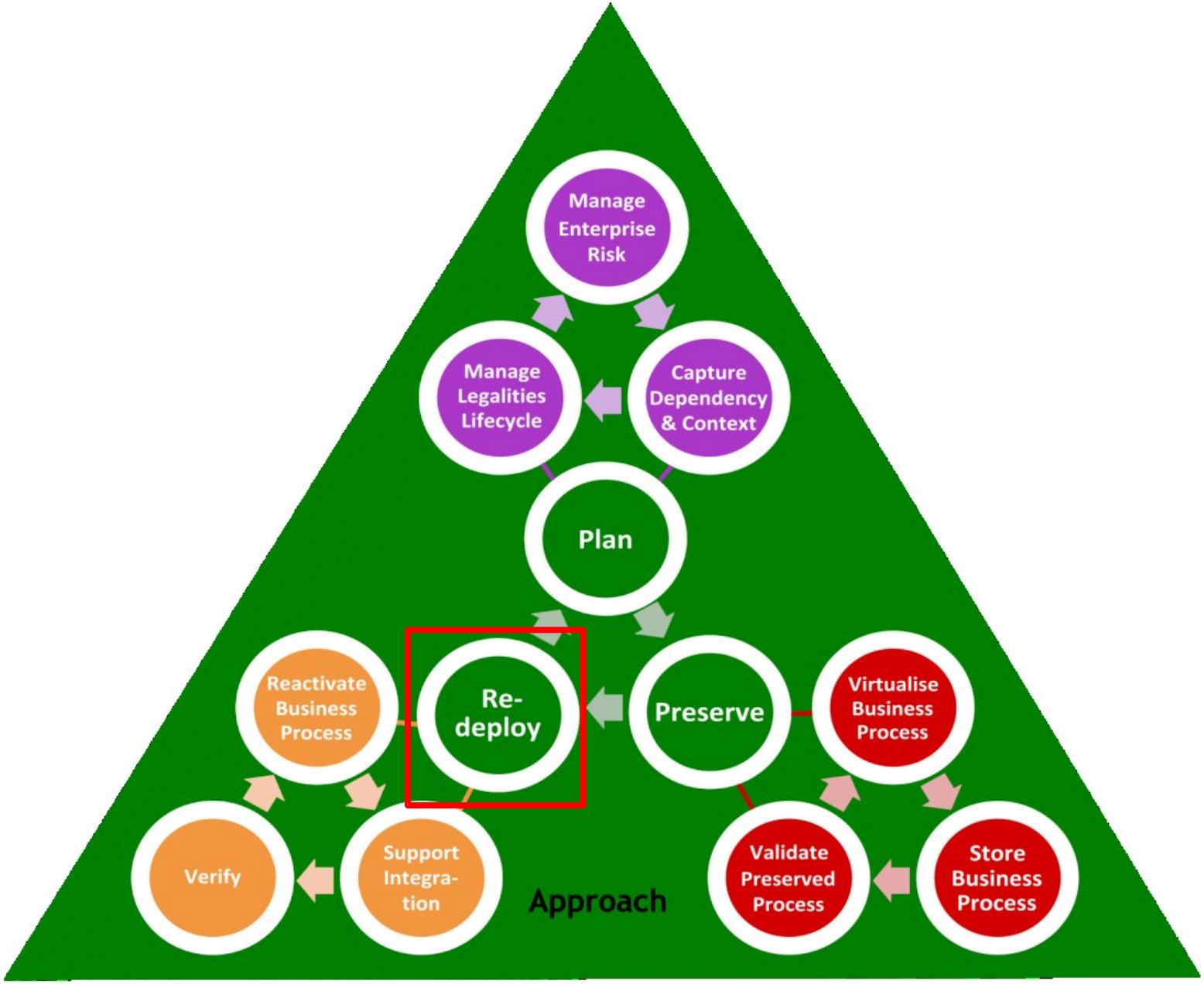



Manage Enterprise Risk
Capture Dependency & Context
Manage Legalities Lifecycle
Plan
Preserve
Virtualise Business Process
Store Business Process
Validate Preserved Process
Re-deploy
Reactivate Business Process
Support Integra-tion
Verify
Approach

DP Engine: Alternatives Assessment

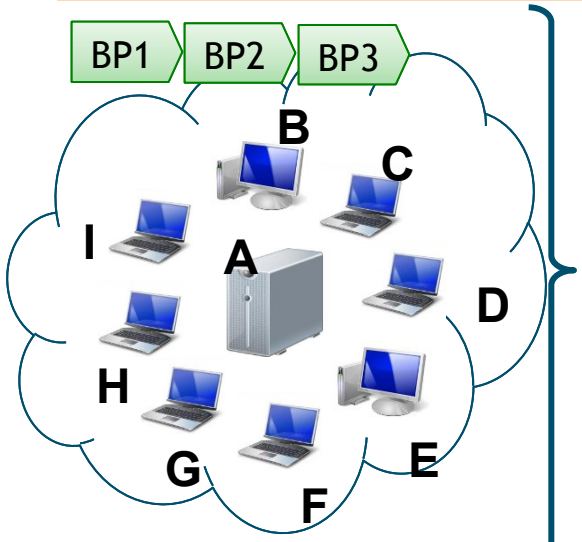
TIMELESS BUSINESS




Manage Enterprise Risk
Capture Dependency & Context
Manage Legalities Lifecycle
Plan
Preserve
Re-deploy
Reactivate Business Process
Virtualise Business Process
Verify
Support Integra-tion
Validate Preserved Process
Store Business Process
Approach

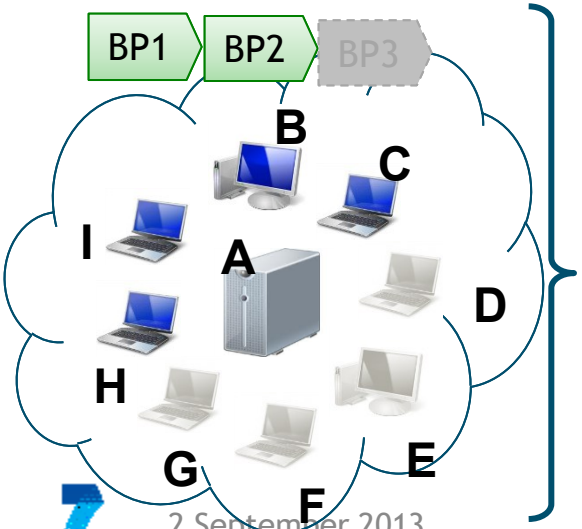
DP Engine: Alternatives Assessment

TIMELESS BUSINESS



From: DP System Archive

Properties:
A,B,C,D,E,F,
G,H,I

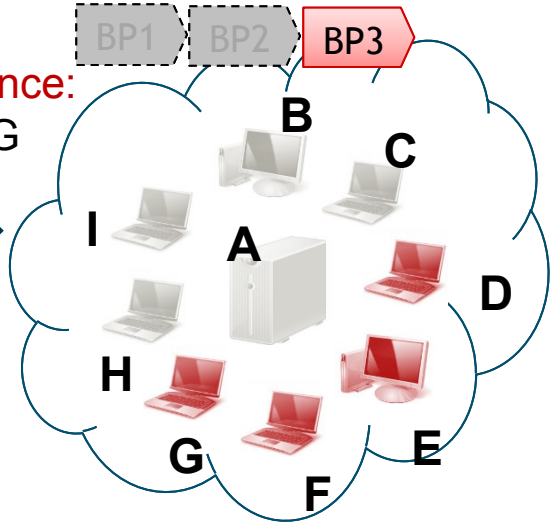


Properties:
A,B,C, H,I

From: Redeployment Environment

Redeployment Alternatives Assessment

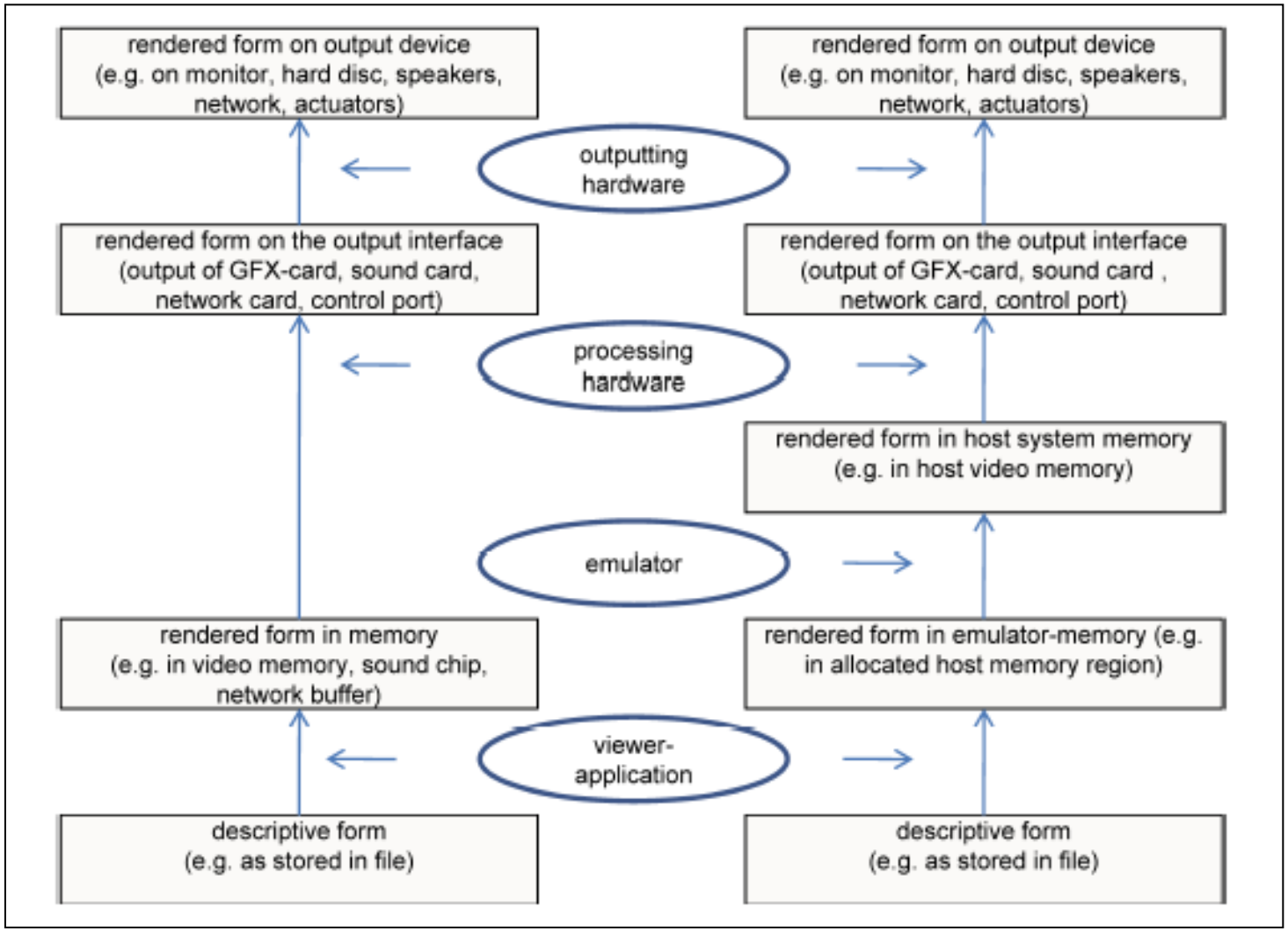
Difference:
D,E,F,G



Redeployment Environment:
missing components

Validation

TIMELESS BUSINESS



Thank you!

TIMBUS

TIMELESS BUSINESS