

# Applying the TIMBUS Approach: Preserving More than Data

#### Carlos Coutinho

CMS, TIMBUS Tools Development Lead



INESC-ID, TIMBUS Tool Owner

Sara Day Thomson

DPC, TIMBUS Head of Dissemination



'Investing in Opportunity'
Wellcome Conference Centre
London
18 November 2014











### Outline





- The TIMBUS Approach
- Context Model
- Business Context Acquisition
- Risk Management
- Risk Solution: DP Expert Suite



### Outline





- The TIMBUS Approach
- Context Model
- Business Context Acquisition
- Risk Management
- Risk Solution: DP Expert Suite



### The TIMBUS Project

TIMELESS BUSINESS @ ③ ③



- TIMeless BUSiness Processes and Services
- IP FP7 Project (2011-2014)

























## TIMBUS Overview & Objectives

TIMELESS BUSINESS **④ ⑤ ▶** 



#### Primary Motivations

- Declining popularity of centralised in-house BPM
- Increasing popularity of external service providers
- Increased concern for dependability and trust

#### TIMBUS Approach

- Establish the set of activities for long-term preservation
- Explore DP from a Business Continuity perspective
- Risk assessment for cost effective preservation
- Dynamic and recurring maintenance





### The TIMBUS Approach

TIMELESS BUSINESS (4) ( )







**Real Business** 



Customer Environment PROCESSES,
not (only) data

Customer Environment





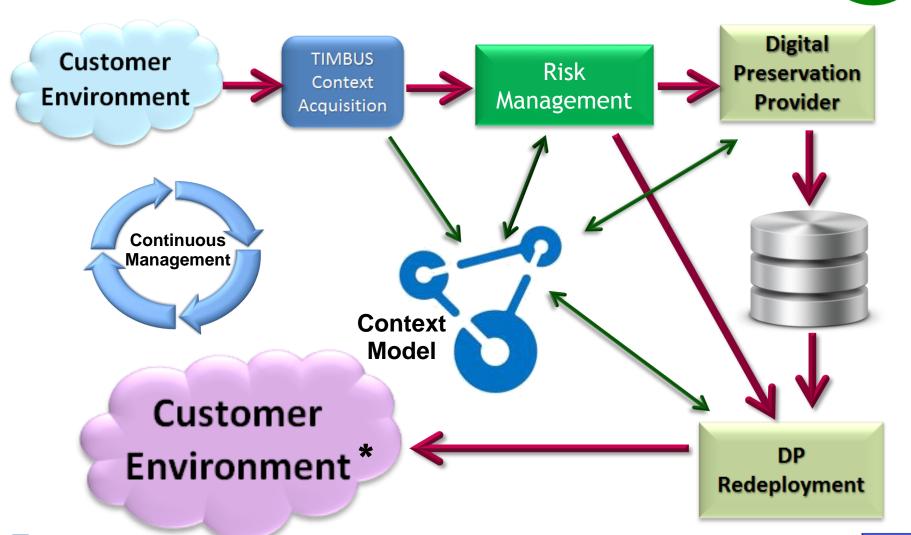


### **Business Context Preservation**













#### **Business Concerns**





- Businesses want to focus on Core Competences
- The Concern of Multiple/Tiered Dependencies
- Dynamic and Expanded Distributed Contexts

#### TIMBUS brings:

- Context: Awareness of Assets & Dependencies
- Risk-driven Management, Effective Cost Modelling
- Digital Preservation as Risk Mitigation Strategy
- Preservation of Data + Processes



### Results & Products of TIMBUS

TIMELESS BUSINESS (4) ( )



- Preserving the WHOLE PROCESS, not just data
- Set of METHODS & TOOLS to identify and capture relevant information
- Ability to capture and store THIRD-PARTY SERVICE INFORMATION
- Integrated LEGAL GUIDANCE and intensive RISK ASSESSMENT
- Safe and secure LONG-TERM DIGITAL PRESERVATION
- RELIABLE REDEPLOYMENT in the future



### TIMBUS Storyboard

TIMELESS BUSINESS @ ③ ③



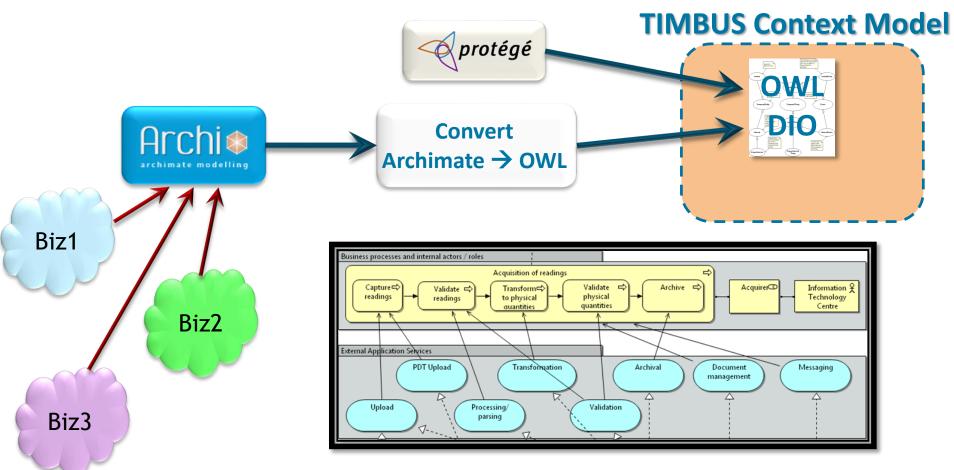
- TØ Context Model Definition/Design
- T1 Creation of Context Model Instance
- T2 Context Model Analysis
- T3 Digital Preservation
- T4 Preservation Maintenance
- T5 Business Redeployment



### TØ - Context Model Definition

TIMELESS BUSINESS (4) © (>)





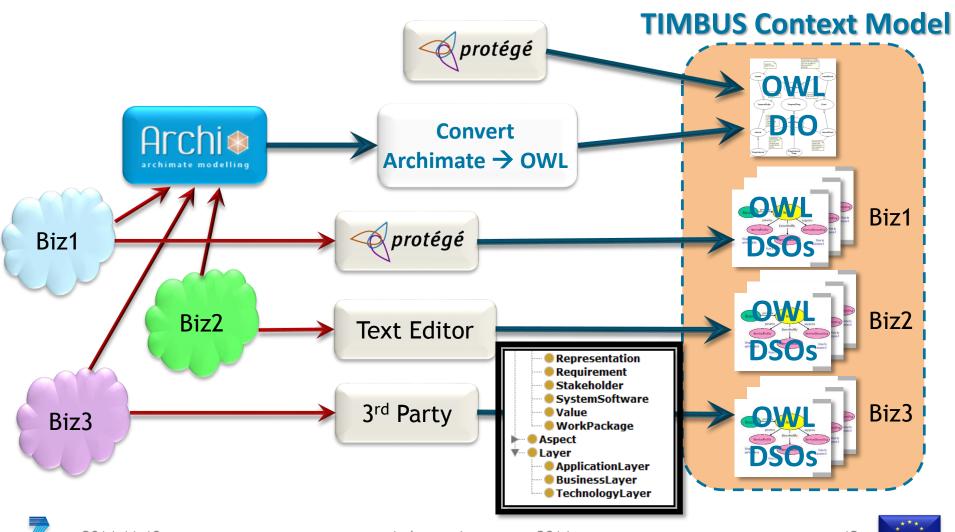




### TØ - Context Model Definition

TIMELESS BUSINESS (4) ( )





### T1 - Context Model Instance

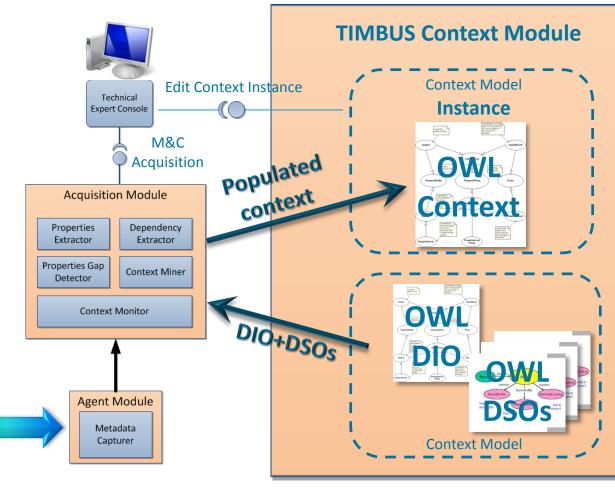
TIMELESS BUSINESS (4) ( )



#### **Extractors:**

- Linux HW
- Debian SW / Licences
- Red Hat SW / Licences
- Windows DLL / MSI
- Perl Modules
- LLM Surveys
- Process Extractors
- Network Information
- Network Topology
- Modified Code
- Process Monitoring
- Network Monitoring
- TACET / Sensors

**Customer Environment** 

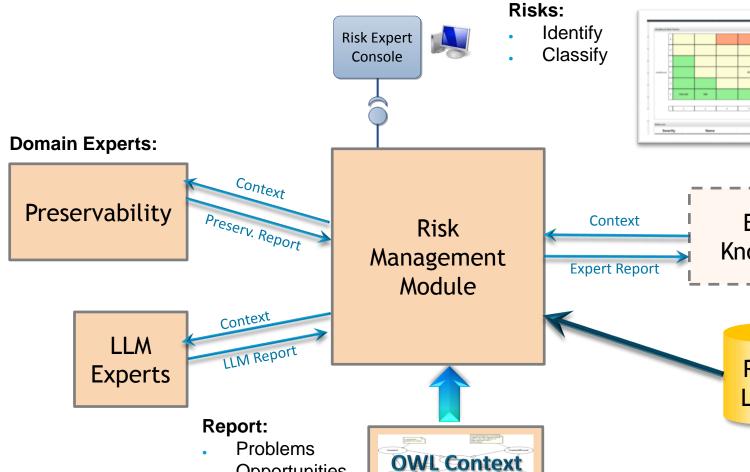






### T2 - Context Model Analysis





**Opportunities** 

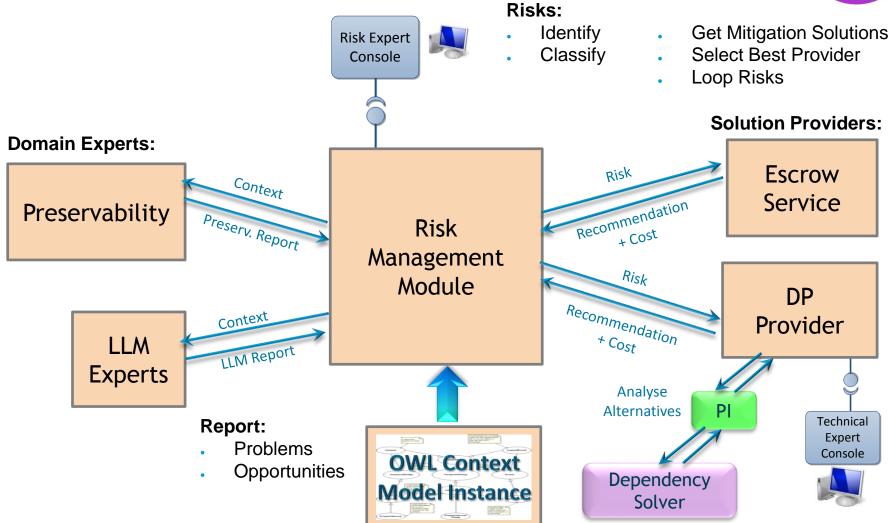


**Model Instance** 

### T2 - Context Model Analysis

TIMELESS BUSINESS (4) (©)







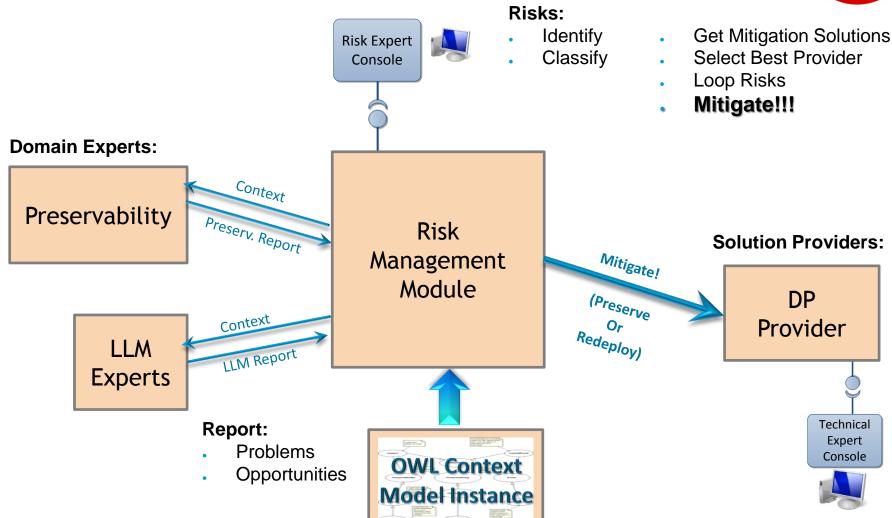
2014-11-18



### T3, T4, T5 - DP of Business

TIMELESS BUSINESS (4) (© (b)





### TIMBUS Use-case Validation



- WP7: Preserve Open Source
  - Open-Source Workflows
  - Phaidra Digital Library
  - RCAAP Aggregation of Digital Libraries
- WP8: Preserve Large Infrastructures
  - CAD/CAM
  - Sensors
  - Mathematical Simulations
- WP9: Preserve Distributed eHealth environment

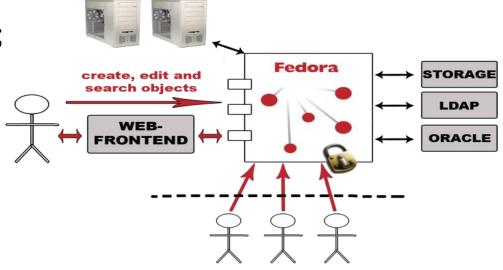




#### Phaidra - Structures



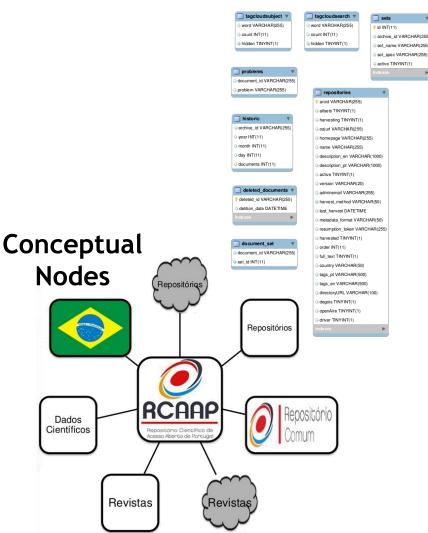
- Architecture
- Implementation
  - Debian 7.1;
  - Fedora Commons 3.3;
  - Tomcat 6;
  - Apache 2;
  - MySQL 5;
  - Catalyst 5.8;
  - Java 1.6;
  - Perl 5.14.
- Complex installation
  - requires multiple installators and custom modules



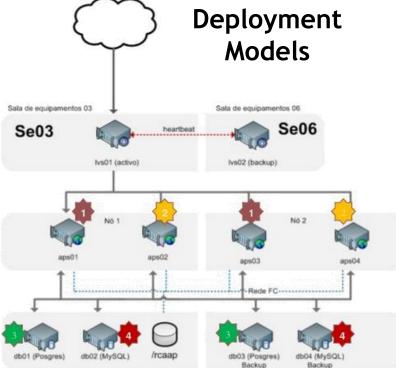
#### **RCAAP - Structures**

TIMELESS BUSINESS @ ©





#### **Database** harvested date DATETIME harvest event DATETIME archive id VARCHAR(255) **Schemas** titles LONGTEXT creators LONGTEXT subjects LONGTEXT descriptions LONGTEXT publishers LONGTEXT contributors LONGTEXT dates LONGTEXT types LONGTEXT o formats LONGTEXT identifiers LONGTEXT o sources LONGTEXT □ languages LONGTEXT o relations LONGTEXT







ooverages LONGTEXT

rights LONGTEXT

alitext LONGTEXT

parsed TINYINT(1)

document\_issue\_date VARCHAR(255
 document\_link VARCHAR(300)

odcument\_type VARCHAR(255)
document\_language VARCHAR(255)

document\_views INT(11)

## TIMBUS Tools & Prototypes (i)

TIMELESS BUSINESS **④ ⑤ ▶** 

Name	Responsible	WP7						WP8			WP9
		OS Workflows									
		DuplicateDetection (MatchBox)		Migration	Characterization	Phaidra	RCAAP	CAD/CAM	Sensors	MathSim	eHealth
		Server	Client	(ImageMagick)	(JHove)						
Context Population	CMS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
- Debian Software Extractor	CMS	□Raw □ OWL	□Raw □ OWL	□ Raw □ OWL	⊡ Raw ⊡ OWL	□Raw □ OWL	No	No	No	□ Raw □ OWL	□ Raw □ OWL
- Linux Hardware Extractor	CMS	□ Raw □ OWL	□ Raw □ OWL	⊡ Raw ⊡ OWL	⊡Raw ⊡ OWL	□ Raw □ OWL	□ Raw □ OWL	No	No	⊡ Raw ⊡ OWL	Raw OWL
- Windows DLL Extractor	INESCID	No	No	No	No	No	No	□Raw □ OWL	No	No	No
- RCAAP Network Topology Extractor	CMS	No	No	No	No	No	□ Raw □ OWL	No	No	No	No
- DSpace Metadata Extractor	CMS	No	No	No	No	No	□ Raw □ OWL	No	No	No	No
- Perl Modules Extractor	CMS	□Raw □ OWL	□Raw □ OWL	□ Raw □ OWL	□ Raw □ OWL		No	No	No	Raw OWL	□ Raw □ OWL
- Static Network Information Extractor	CMS	Yes(TODO)	Yes(TODO)	Yes(TODO)	Yes(TODO)	Yes(TODO)	Yes(TODO)	No	No	Yes(TODO)	Yes(TODO)
Taverna Extractor	SBA	No	₫ OWL	⊡ OWL	⊡ OWL	No	No	No	No	No	No
Dynamic Network xtractor (ApacheLog)	KIT	No	No	No	No	⊕Yes	No	No	No	No	No
Dynamic Network xtractor (SystemTap)	KIT	Yes	Yes	Yes	⊡Yes	No	Yes	No	No	Yes	Yes
Modified/Moved Managed Code	KIT	No	No	No	No	⊡Yes	No	No	No	No	No
Local Groups/Users	KIT	No	No	No	No	⊡Yes	No	No	No	No	No



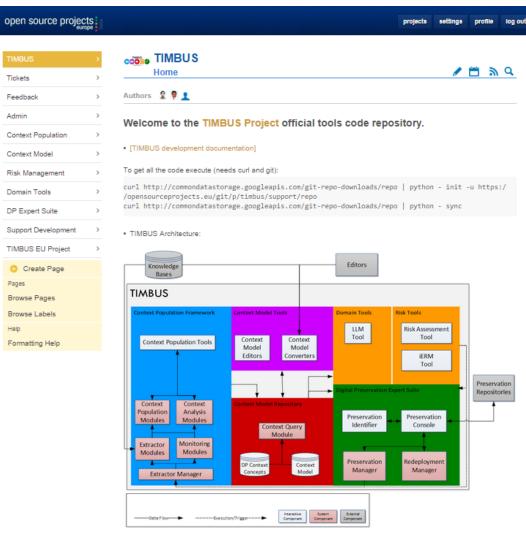
## TIMBUS Tools & Prototypes (ii)



Unix Permissions	KIT	No	No	No	No	⊕Yes	No	No	No	No	No
Added Files	KIT	No	No	No	No	⊕Yes	No	No	No	No	No
Modified Package	KIT	No	No	No	No	⊕Yes	No	No	No	No	No
Sensors Extractor	KIT	No	No	No	No	No	No	No	⊕Yes	No	No
TACET	KIT	No	No	No	No	No	No	No	⊕Yes	No	No
Process Extractor	SAP	No	No	No	No	No	No	No	Yes	No	Yes
LLM (Legal Life cycle Management) Tool	SAP/SQS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Risk Evaluation & Treatment	INESCID	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
iERM (Intelligent Enterprise Risk Management)	SAP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
iERM Extractor	CMS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Context Model DIO to BPMN Converter	CMS	⊕ BPMN	No	□ BPMN	⊡ BPMN	Yes(TODO)	□ BPMN	□ BPMN	No	□ BPMN	□ BPMN
Ontology Visualisation Tool	SBA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DPES (Digital Preservation Expert Suite)	INTEL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Preservation Alternative Identifier	SBA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependencies Reasoner	CMS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

### TIMBUS is Open Source





- ... almost totally
- Licence: Apache v2.0
- Reuses SOTA technology
- Uses standards:
  - Archimate®
  - ISO31000:2009
  - CUDF
- Code available on:

https://opensourceprojects.eu/p/timbus



### Outline





- The TIMBUS Approach
- Context Model

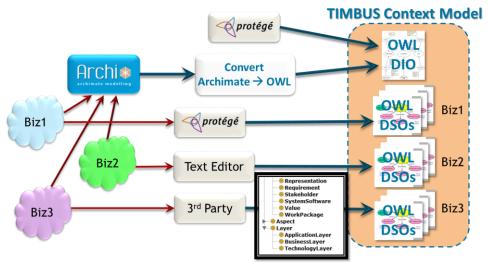
- Archimate®
  - Flexibility
- Business Context Acquisition
- Risk Management
- Risk Solution: DP Expert Suite

#### Context Model Overview

TIMELESS BUSINESS (4) ( )



- Information Systems Model
  - Ontology of concepts for semantic analysis & relation
  - Distributed databases
- Standards & Tools
  - Archi / Archimate®
  - Protégé
  - CUDF



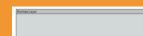
One DIO/Multiple DSOs

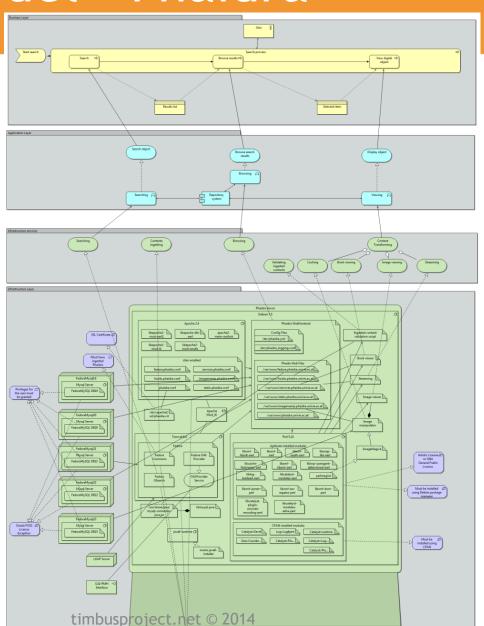


### Context Model - Phaidra









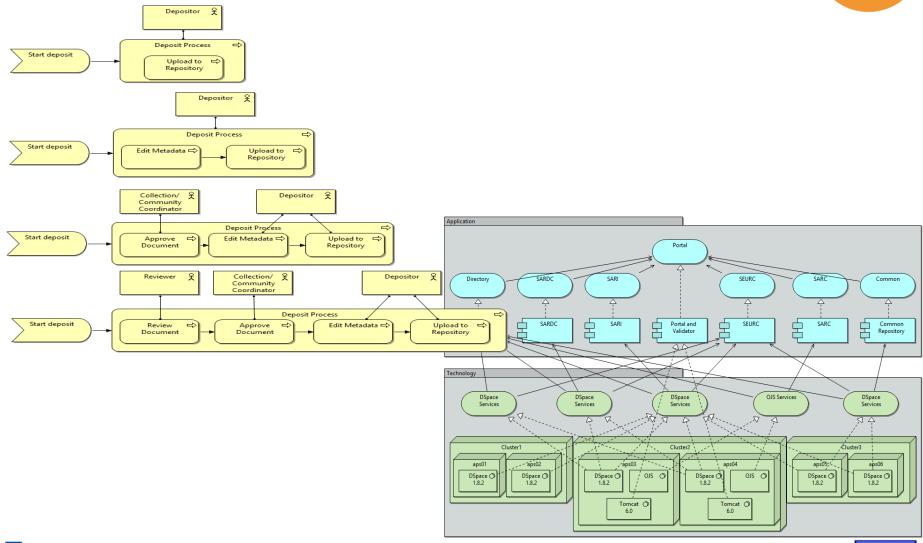




### Context Model - RCAAP

TIMELESS BUSINESS (4) ( )







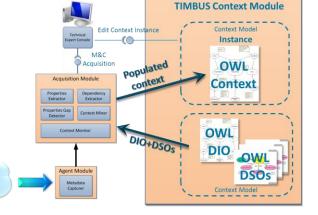


#### Outline



- The TIMBUS Approach
- Context Model
- Business Context Acquisition
- Risk Management
- Risk Solution: DP Expert Suite

- Challenges
- Aggregation







Customer

Environment

## Business Extraction: Objectives

TIMELESS BUSINESS **❸ ⑤ ⓒ** 



#### TIMBUS Phase 1: Capturing the customer environment





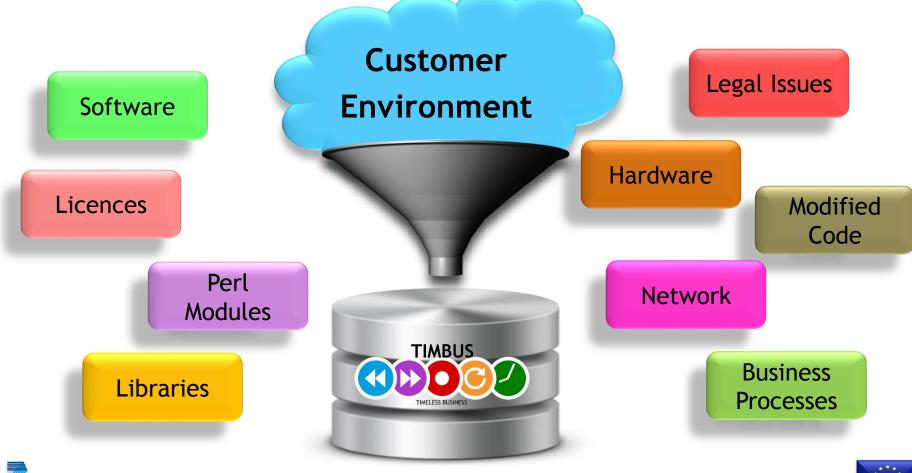


## Business Extraction: Objectives

TIMELESS BUSINESS @ © ©



#### TIMBUS Phase 1: Capturing the customer environment





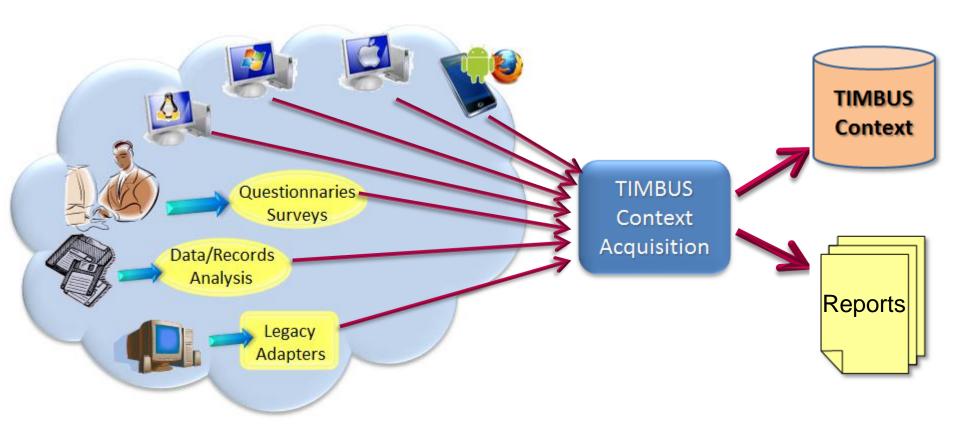


## Business Extraction: Challenges

TIMELESS BUSINESS @ © D



#### Heterogeneity



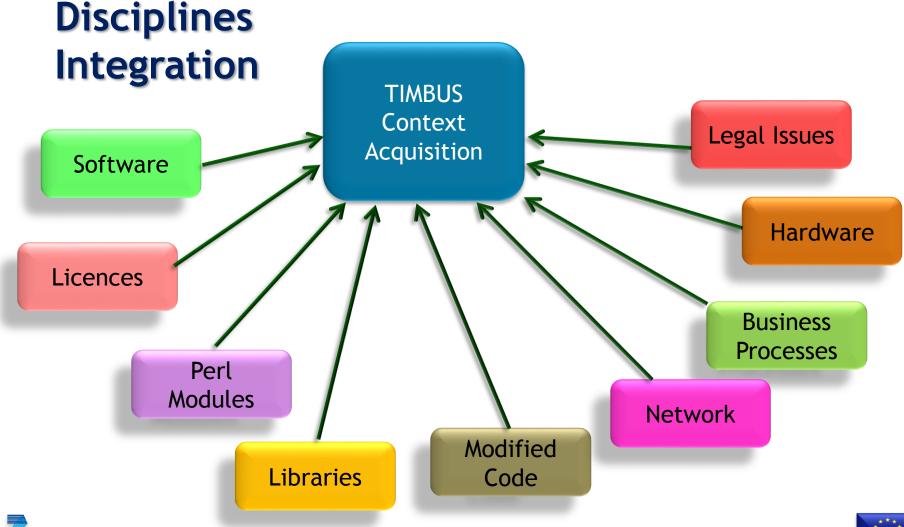




## Business Extraction: Challenges

TIMELESS BUSINESS (4) ( )









### Business Extraction: Requires

TIMELESS BUSINESS (4) ( )

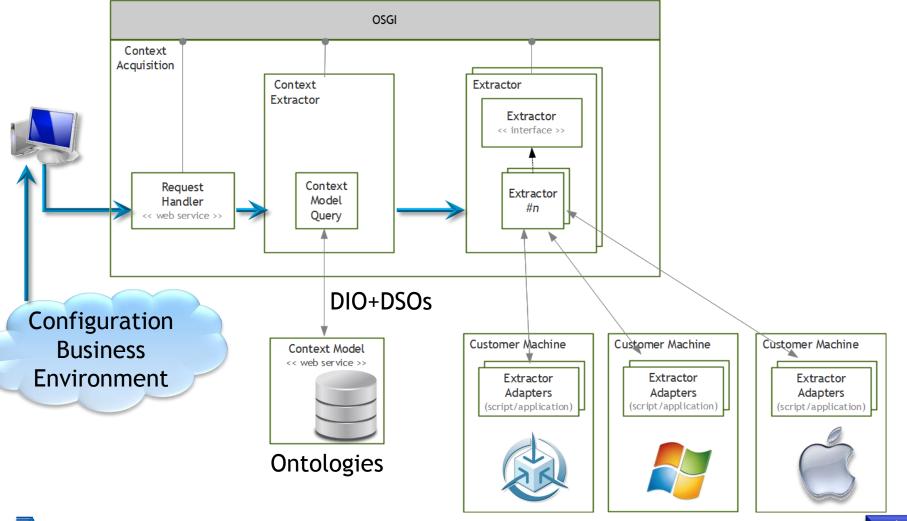


- Comprise multiple analyses
- Cover multiple disciplines
- Integrate various providers
- Reduce/avoid downtime
- Enable Modular & Dynamic extraction
- Handle heterogeneous environments
- Have a contributive philosophy



### **Business Extraction: Proposal**



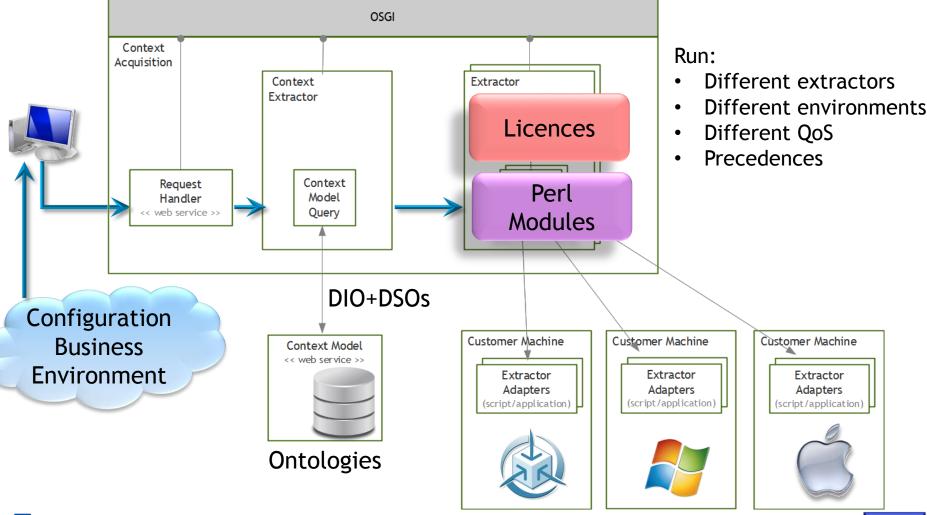






### **Business Extraction: Proposal**

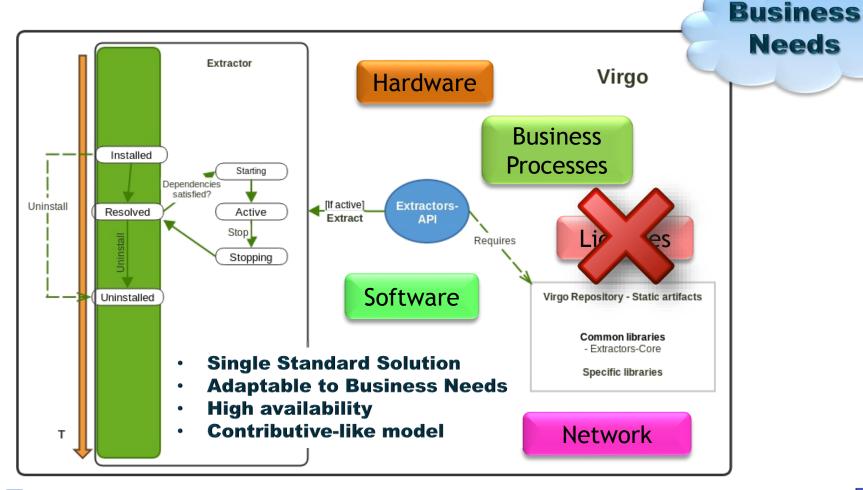




### OSGi in TIMBUS Extraction FW

TIMELESS BUSINESS (4) ( )







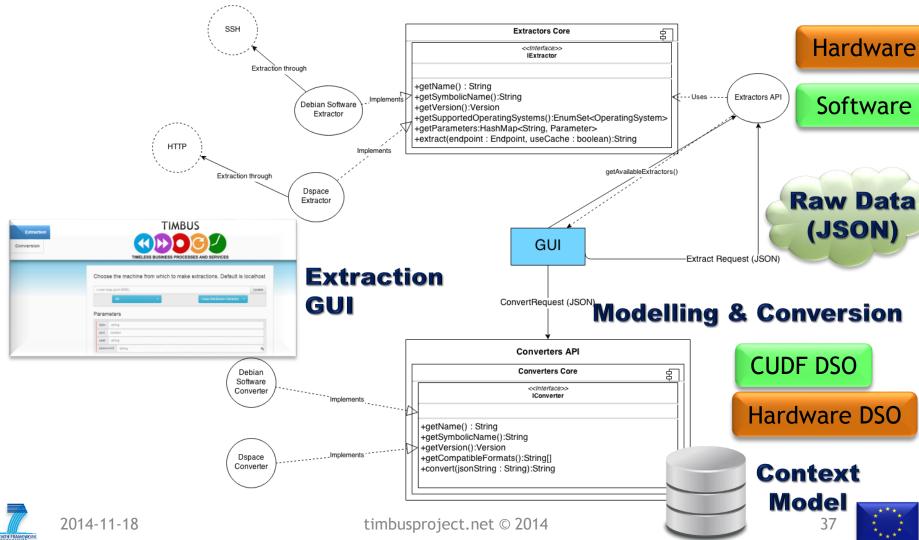


#### Extraction FW: Architecture

TIMELESS BUSINESS (4) (3) (4)



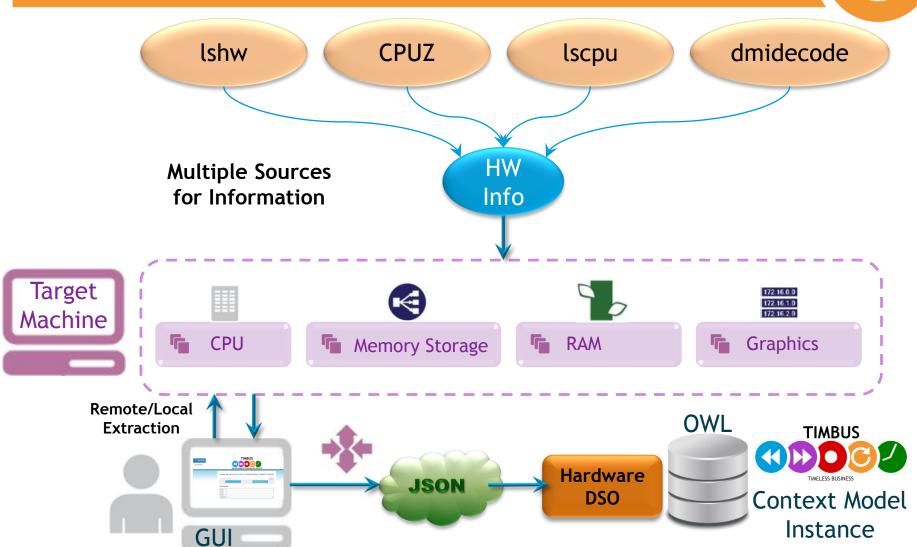
#### **Environment Extraction**





### Demo: Linux HW Extraction

TIMELESS BUSINESS (4) ( )



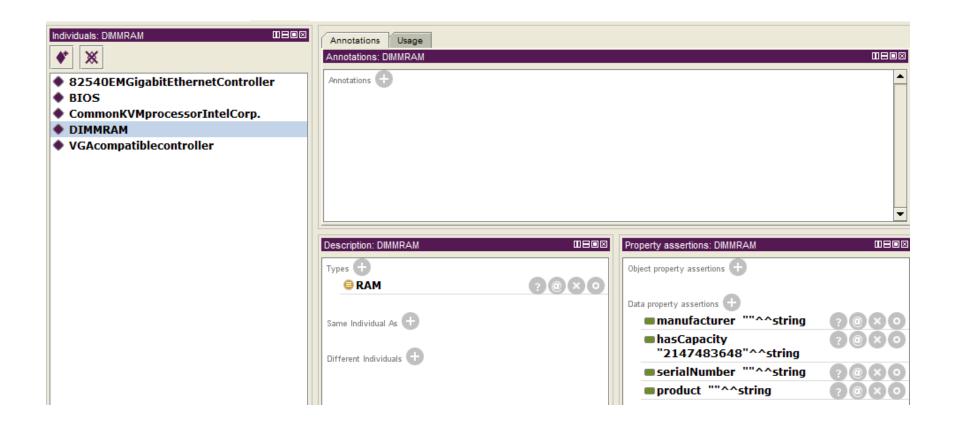




38

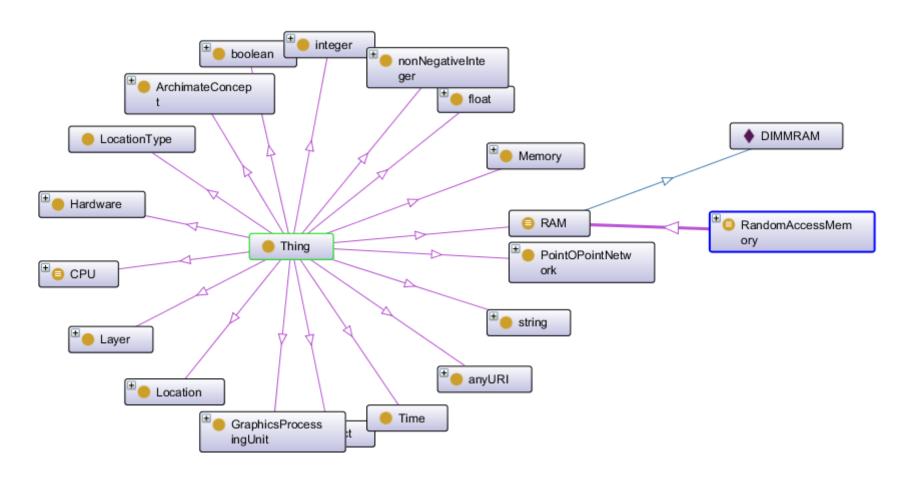
### Linux HW Result - Phaidra





## Linux HW Result - Phaidra

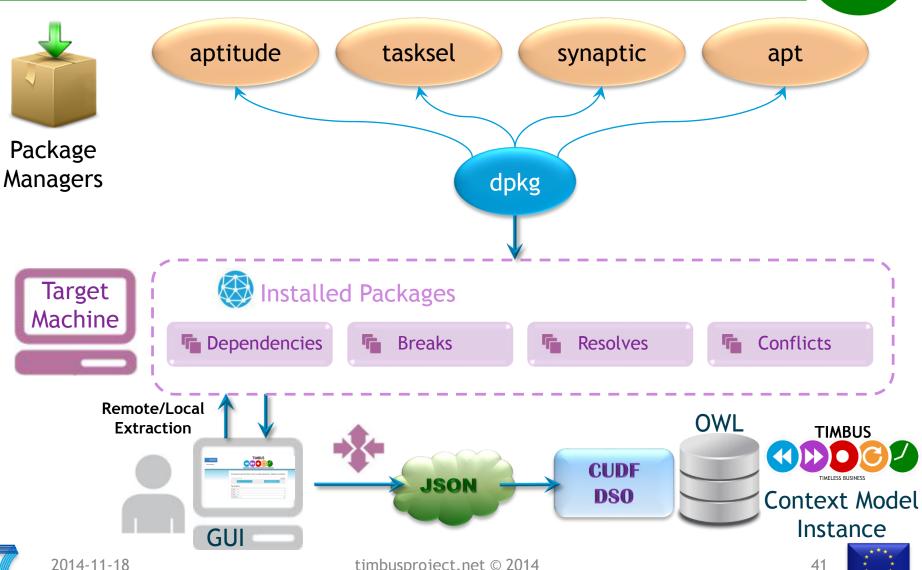








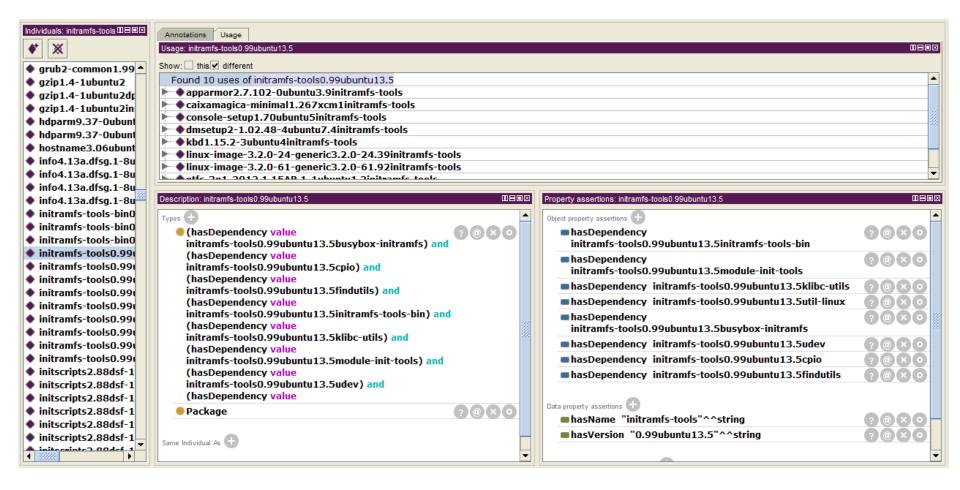
### Demo: Debian SW Extractor





### Debian SW Result - Phaidra



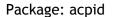






### Debian SW Result - Phaidra

#### TIMELESS BUSINESS 🐠 💿 🕟



Status: install ok installed

•••

Installed-Size: 156

Architecture: amd64

Version: 1:2.0.18-1ubuntu1

Depends: libc6 (>= 2.15), upstart-job,

lsb-base (>= 3.2-14), kmod

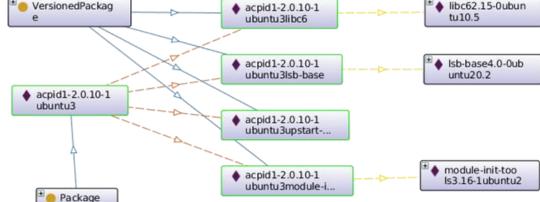
#### Conffiles:

•••

/etc/acpi/events/powerbtn 57f021ab2369f211018d6e7d11434314





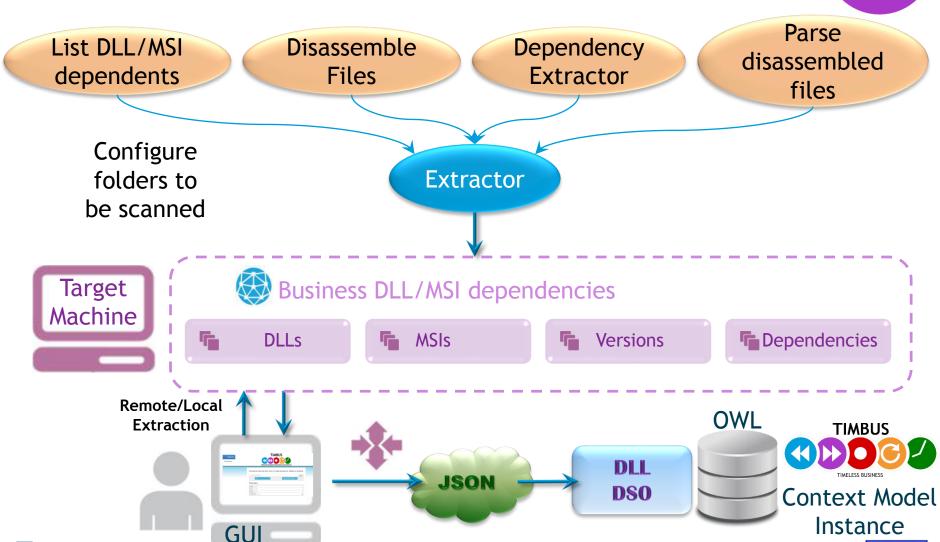






## **Demo: Windows DLL Extraction**





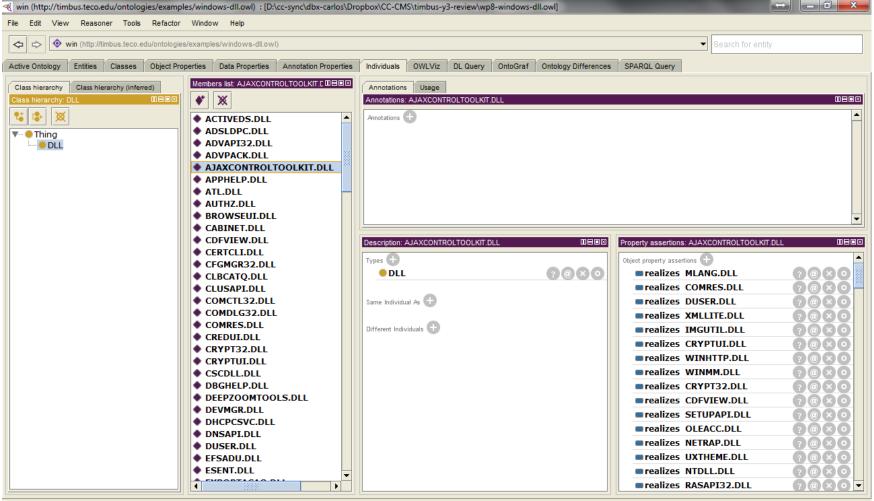




### Windows DLL Result

TIMELESS BUSINESS (4) (©)





timbusproject.net © 2014

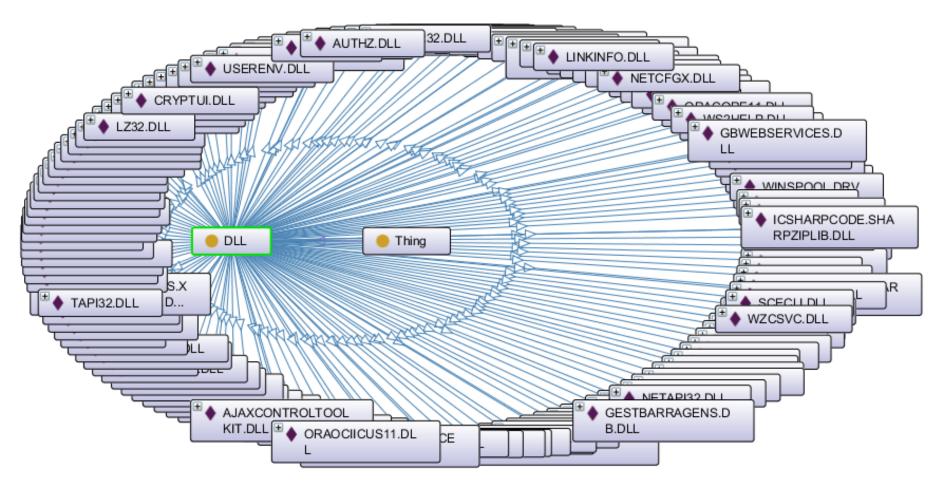




No Reasoner set, Select a reasoner from the Reasoner menu Show Inferences

### Windows DLL Result







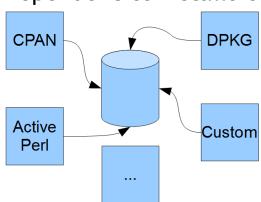


### Demo: Perl Modules Extractor

TIMELESS BUSINESS **④ ◎ ▶** 



#### Dependencies installers



- System specific installers
  - DPKG, RPM, Emerge, ...



- Perl Community
  - CPAN, Active Perl, ...

Package Managers

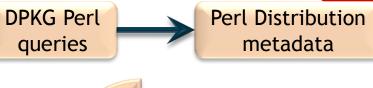
- Diversity and no unified view
  - No super installer
  - No communication among them
  - Each has their set of Perl Distributions
  - They are complementary

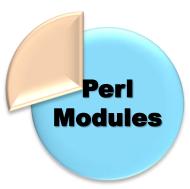
- Capturing ALL Perl dependencies:
  - Query different sources
  - Use Meta information
  - Use Perl VM dependency checker

### Demo: Perl Modules Extractor

TIMELESS BUSINESS **④ ⑤ ▶** 









Peri

**Modules** 

Perl VM
Dep Check

Perl Modules
dependencies

Perl Modules
dependencies

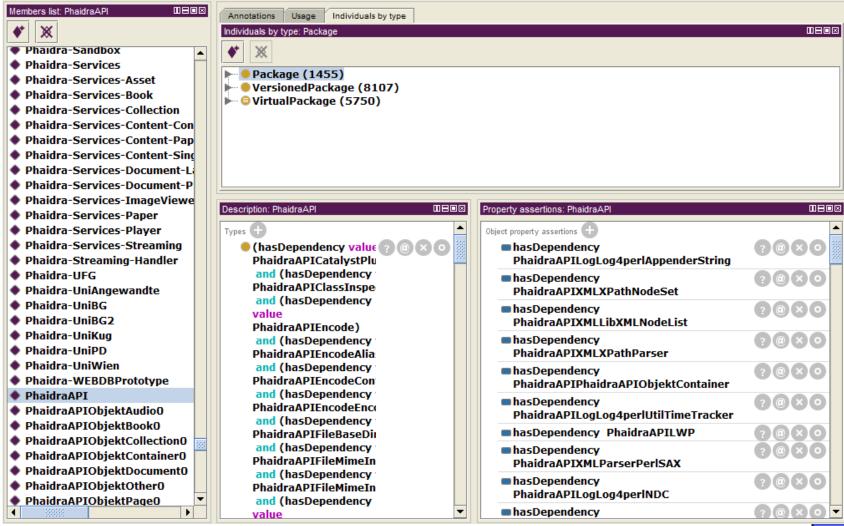




### Perl Modules Result - Phaidra

TIMELESS BUSINESS **④ ⓒ №** 

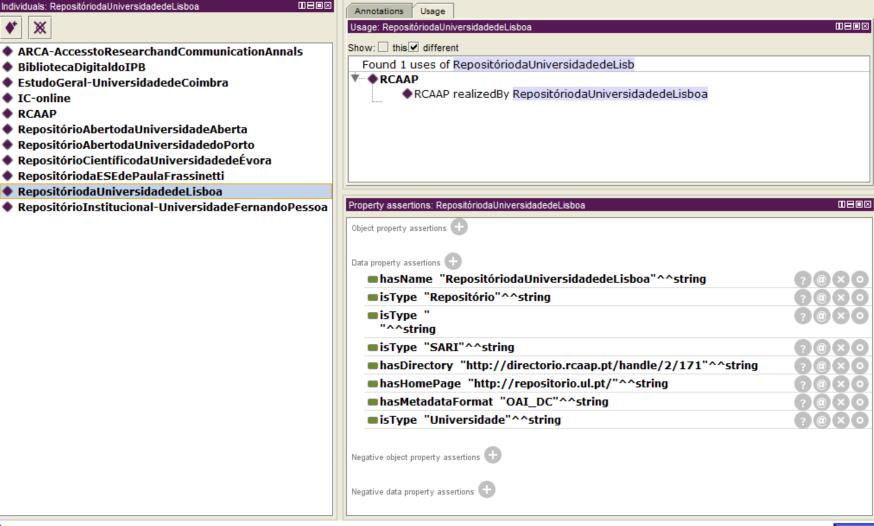








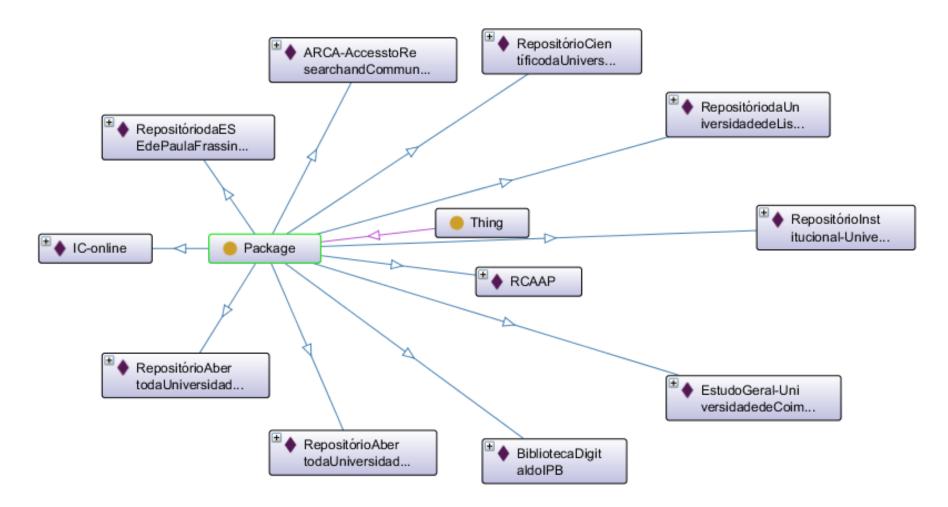
# Network Topology Result - RCAAP







# Network Topology Result - RCAAP







# Context Extraction Framework

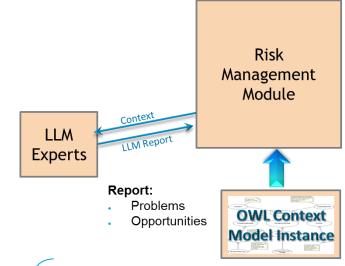


- The Extraction Framework allows:
  - Flexibility: One solution that fits all needs
  - Distribution: The solution is upgradeable online
  - Customised: Solution customised for each business
  - Contribution-driven: Can receive contributions
  - The creation of new business models

### Outline



- The TIMBUS Approach
- Context Model
- Business Context Acquisition
- Risk Management
- Risk Solution: DP Expert Suite



- Provider: LLM
- Risk Management Tool





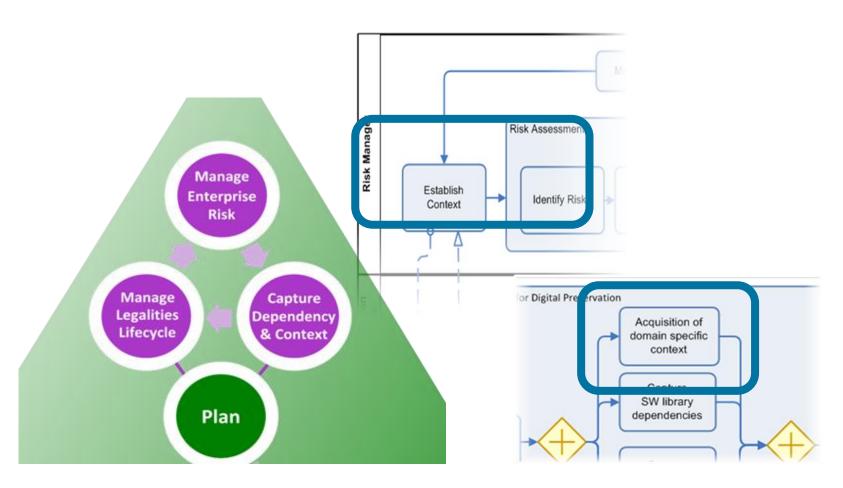
# LLM in the DP process

TIMELESS BUSINESS **④ ⊙ ⊚** 











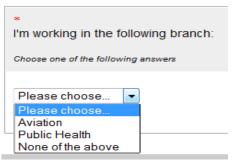


# LLM-Tool - possible legal issues

TIMELESS BUSINESS **④ ◎ №** 



- Guidance and support in four relevant areas:
  - Intellectual property management
  - IT-contracting
  - Data protection
  - Preservation obligations
- Different requirements and scenarios
- Increase Business Legal Awareness

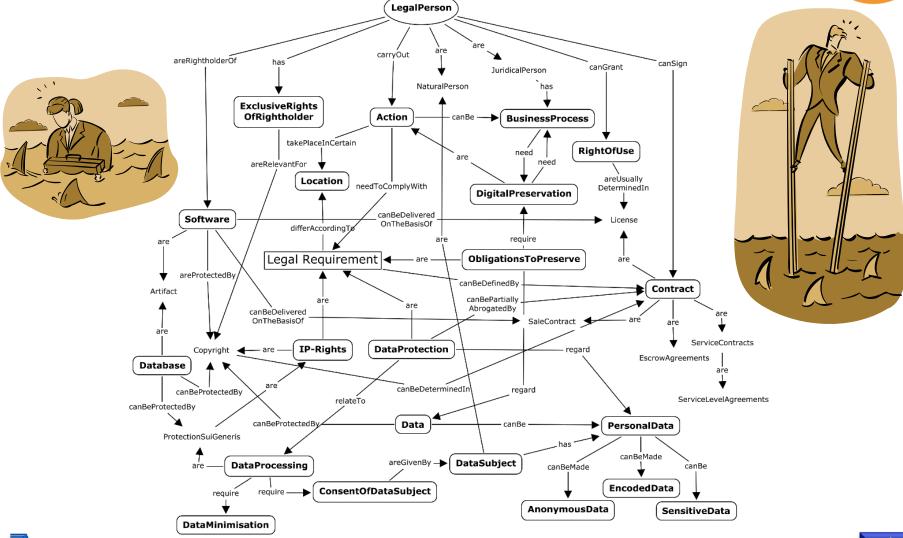






# TIMBUS Legal Ontology



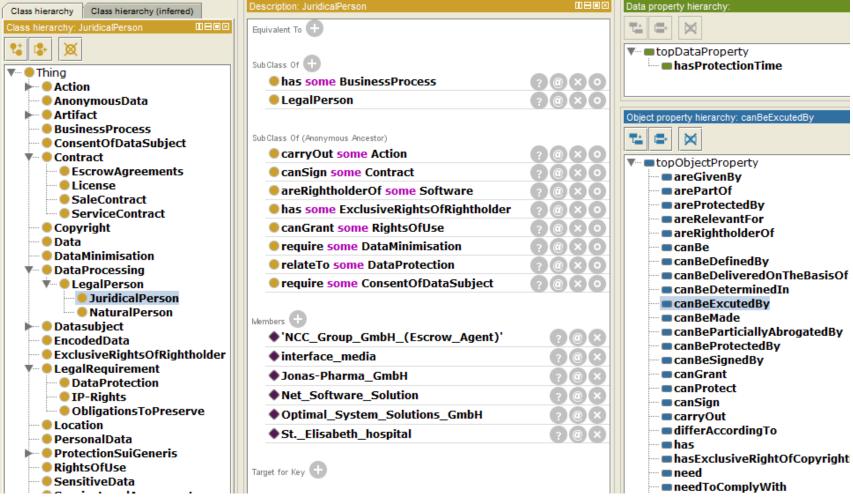






# TIMBUS Legal Ontology



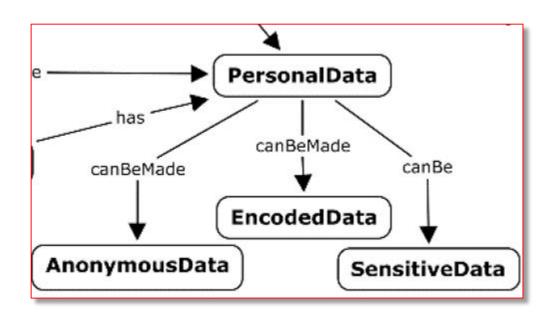


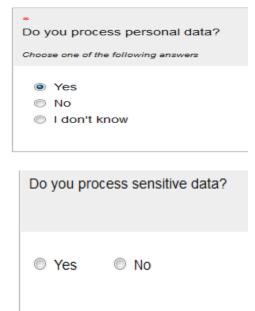




# Ontology ←→ Survey







Is it possible to anonymize the data you want to preserve?

<u>Explanation:</u> One way to circumvent the regulations of data privacy laws is to make the personal data you are preserving anonymous.

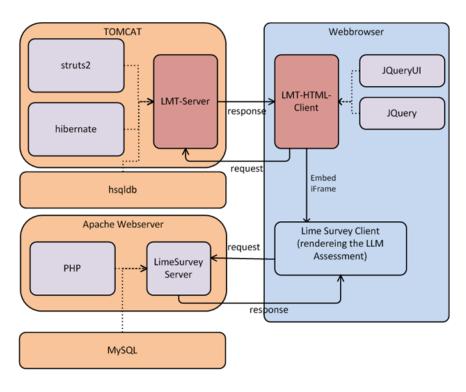




### **LLM Architecture**

TIMELESS BUSINESS **④ ⊙ ⊙** 











# LLM Legal Monitor



- Integration component
- Accessible under Creative Commons License
- Implements relevant jurisdiction and legislation
- Adds a legal monitor to the project
- Automatic notification in case of legal amendments
- Access to legal guidelines

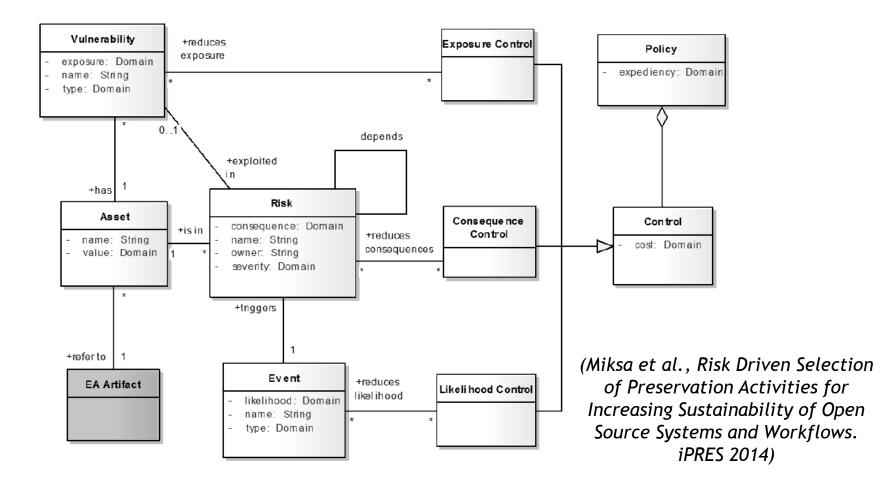






# Risk Management Tool Concepts



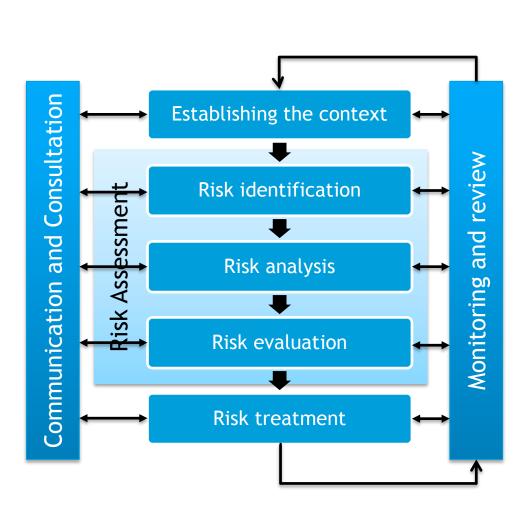


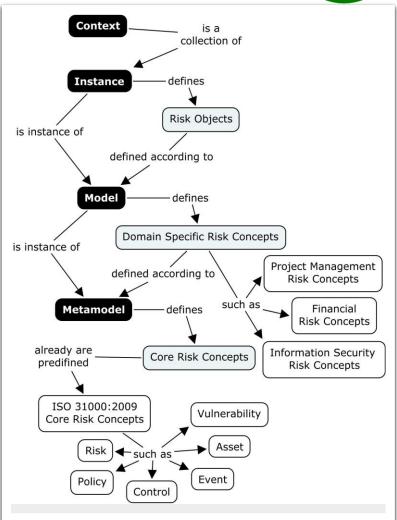




# Risk Management Lifecycle











62

### Risk Assessment - Phaidra

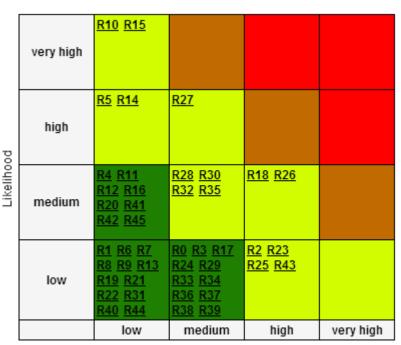


### Without TIMBUS

#### R10 R15 very high R16 R20 R5 R14 R27 high Likelihood R4 R11 R42 R45 R32 R35 R41 R12 R13 medium R18 R26 R28 R30 R1 R19 R3 R17 R2 R6 R7 R0 R21 R22 R33 R34 R8 R9 R23 R31 R44 R36 R37 R24 R25 low R38 R39 R29 R43 R40 low medium high very high

#### Consequences

### With TIMBUS



Consequences



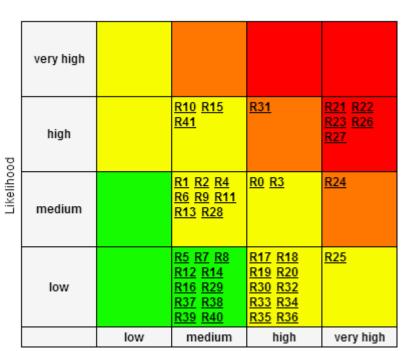


### Risk Assessment - RCAAP

TIMELESS BUSINESS @ ③ •

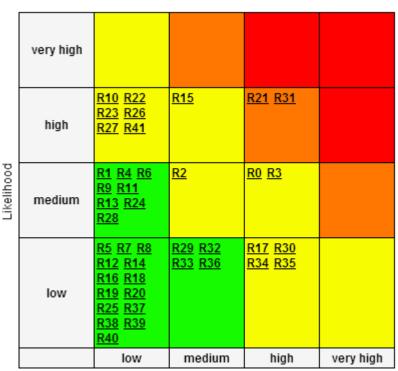


### Without TIMBUS



Consequences

### With TIMBUS



Consequences

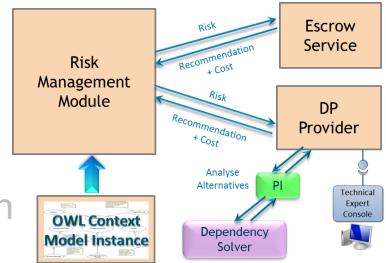




### Outline



- The TIMBUS Approach
- Context Model
- Business Context Acquisition
- Risk Management
- Risk Solution: DP Expert Suite



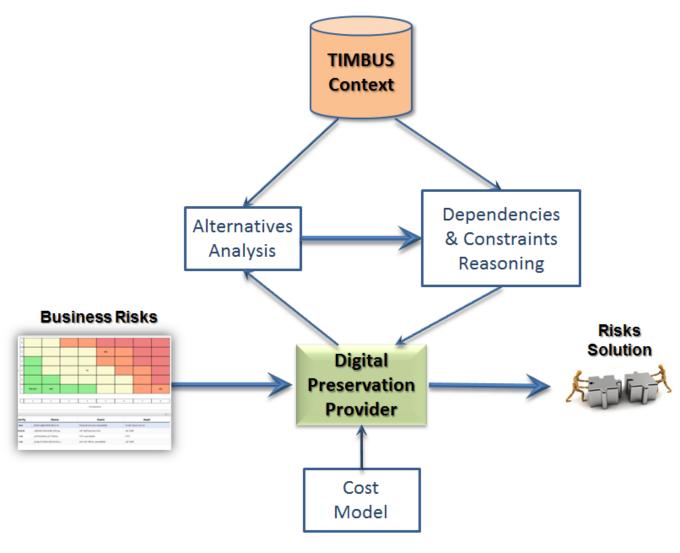
- Digital Preservation
- Alternatives: PI+DRS
- Redeployment





# TIMBUS Preservation Approach







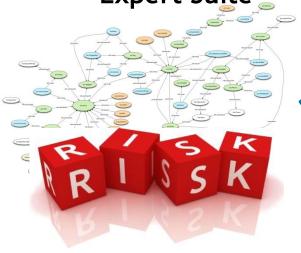


### Preservation Identifier

TIMELESS BUSINESS **④ ⊙ ⊚** 



Digital Preservation Expert Suite



Preservation Alternative Identification











# PI - Knowledge Bases

TIMELESS BUSINESS (4) (©)



Information on alternatives for file formats & software

- Based on existing registries / knowledge bases
- Integrates three registries / KBs
  - PRONOM
  - Freebase
  - Linux Software packages



# PI - Knowledge Bases

TIMELESS BUSINESS (4) (5)



- File format registry: PRONOM (National Archives)
  - Centrally curated database, ~1,100 formats
  - Unique, persisten identifier for formats
  - http://www.nationalarchives.gov.uk/pronom
- Files identified & matched to PRONOM Ids
  - E.g. with DROID



(<a href="http://digital-preservation.github.com/droid">http://digital-preservation.github.com/droid</a>)



# PI - Knowledge Bases

TIMELESS BUSINESS @ ③ ③



- File format & software applications: Freebase
  - Community maintained linked-open data base
    - http://www.freebase.com/ Freebase



- File Formats: 3,500 entries
- Software applications: 9,000 entries
  - Including different versions of the same software
- Software ←→ File Format read/write associations
- Software & File Formats categorised in genres
  - E.g. Audio file formats, word processor



### PI - Alternatives Identification

TIMELESS BUSINESS @ © ©



- Input: Process context model
- Optional input: list of formats & SW at risk
- Output: list of all technically possible alternatives

- Alternatives need to be reviewed by an expert for their feasibility
- Based on Ontology reasoning / DL queries



### DP Alternatives - Phaidra

TIMELESS BUSINESS (4) ( )



Substitution of packages at risk

oracle-java7installer

openjdk-6-jdk

gcj-4.6-jdk

ImageMagick

graphicsmagickimagemagick-compat (1.3.18-2)

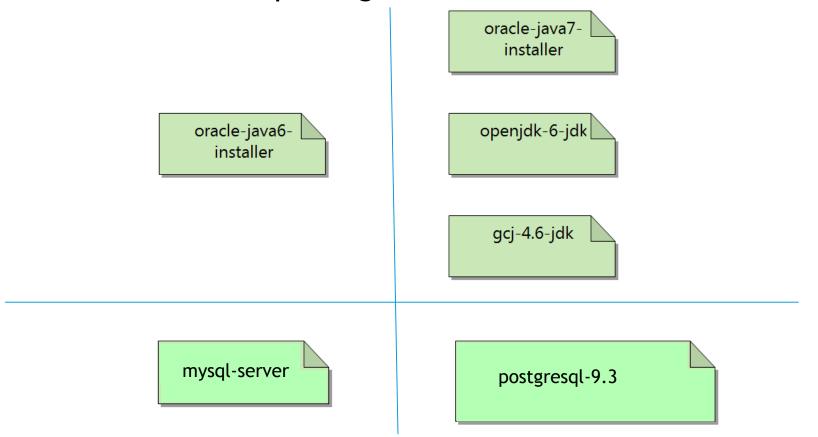


### DP Alternatives - RCAAP

TIMELESS BUSINESS ⊕ ⊙ ⊙



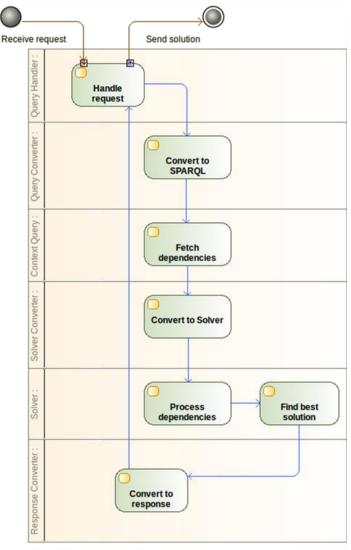
Substitution of packages at risk



# Dependencies Reasoner/Solver

TIMELESS BUSINESS 🕢 🌀 😥





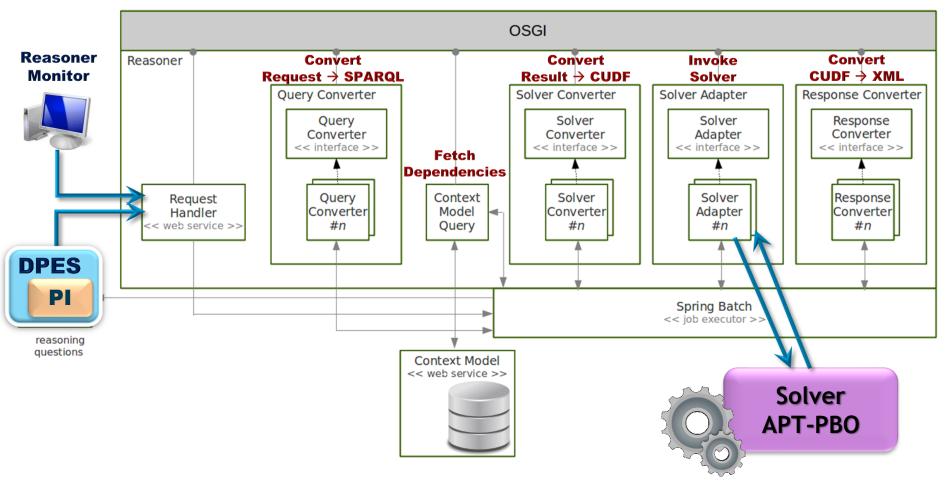




# Dependencies R/S Architecture

TIMELESS BUSINESS **④ ⓒ №** 



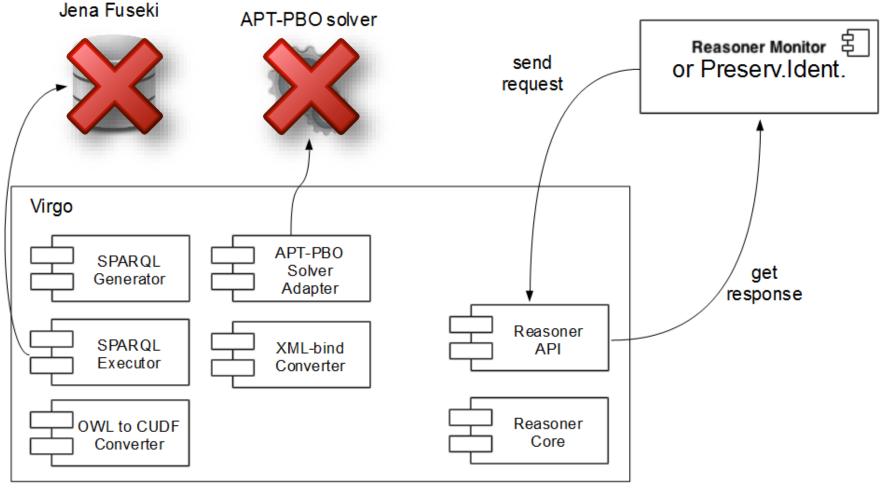






# Dependencies Reasoner/Solver





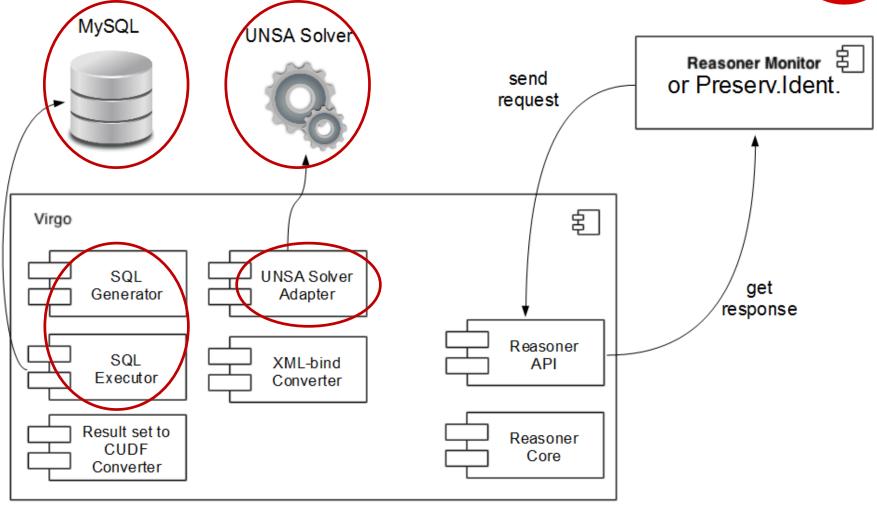




# Dependencies Reasoner/Solver

TIMELESS BUSINESS **④ ⑤ ▶** 







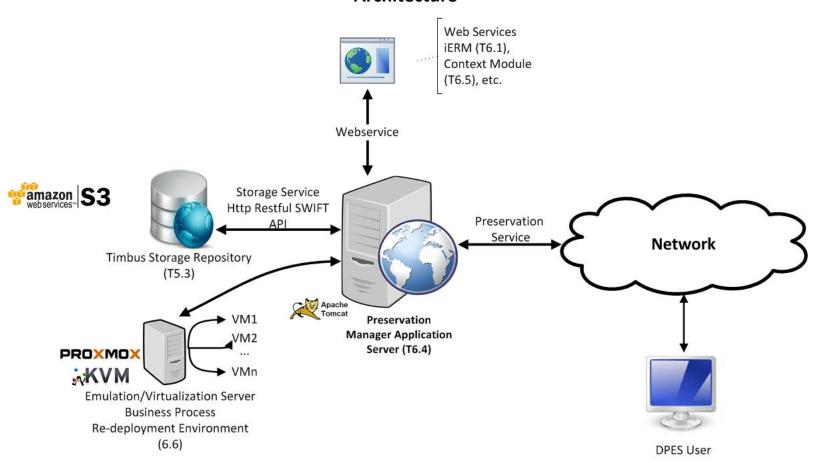


### **DPES - Architecture**

TIMELESS BUSINESS (4) ( )



### Digital Preservation Expert Suite Architecture







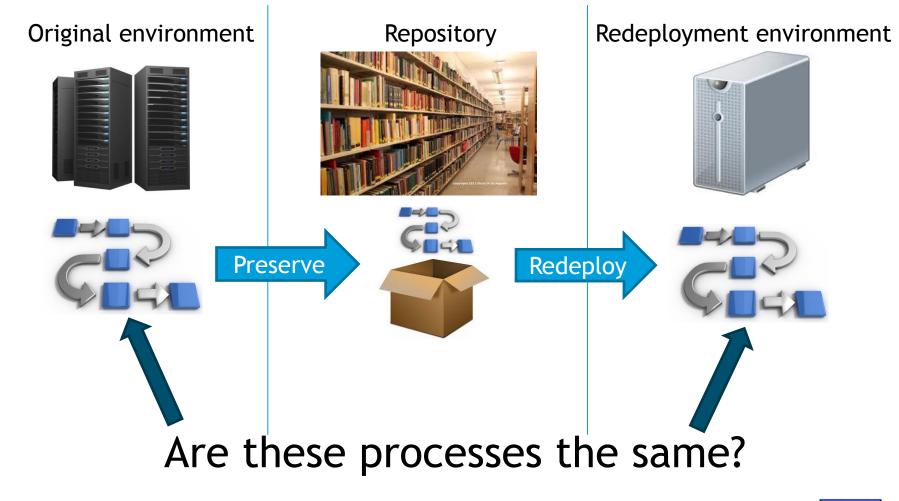
78

# Preservation V&V - VFramework









# Preservation V&V - VFramework











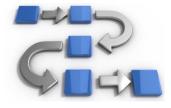




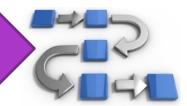


#### Redeployment environment





DIRECT COMPARISON NOT POSSIBLE







**COMPARE EVIDENCE** 





# Preservation V&V - VFramework

TIMELESS BUSINESS (4) ( )











Prepare system for preservation



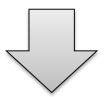
Design verification setting



Capture verification data



### **VPlan**



Prepare system for redeployment



Capture redeployment performance data



Compare and assess





# Thank YOU!!!







