



RODA: repositório de objetos digitais autênticos - passado, presente e futuro...

José Carlos Ramalho jcr@keep.pt

KEEP SOLUTIONS www.keep.pt

Miguel Ferreira mferreira@keep.pt

KEEP SOLUTIONS www.keep.pt

Encontro Internacional de Arquivos
Évora, Portugal, 2014-10-03

Luís Faria lfaria@keep.pt

KEEP SOLUTIONS www.keep.pt

Hélder Silva hsilva@keep.pt

KEEP SOLUTIONS www.keep.pt

Sebastien Leroux sleroux@keep.pt

KEEP SOLUTIONS www.keep.pt

KEEP SOLUTIONS: Projetos



- DigitArq, CRAV (2003..[2008-2012])
- RODA (2006..[2008-...])
- RCAAP (2008-...)
- PPA (2009)
- Open source: RODA, KOHA, DSpace, Moodle, etc.
- Scientific research
 - **SCAPE**: Preservação digital em larga escala
 - **4C**: previsão de custos na preservação digital
 - **e-ark**: desenvolvimento de um modelo de referência europeu baseado no OAIS



<http://www.keep.pt>



O passado: 2006 - 2009

RODA: Repositório de Objectos Digitais Autênticos

José Carlos Ramalho
jcr@di.uminho.pt

Francisco Barbedo
frbarbedo@iantt.pt

Miguel Ferreira
mferreira@dsi.uminho.pt

Cecília Henriques
chenriques@iantt.pt

Rui Castro
Rcastro@iantt.pt

Luis Corujo
lcorujo@iantt.pt

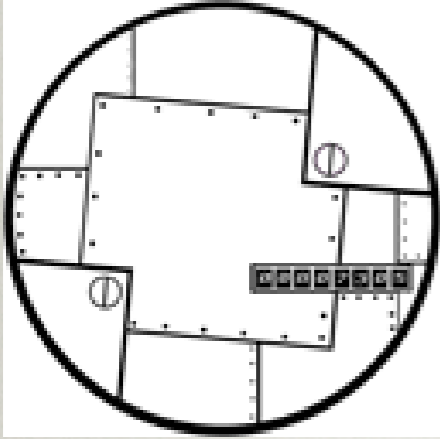
Luis Faria
lfaria@iantt.pt

O que é o RODA?

Um repositório digital concebido para arquivos, com as seguintes características:

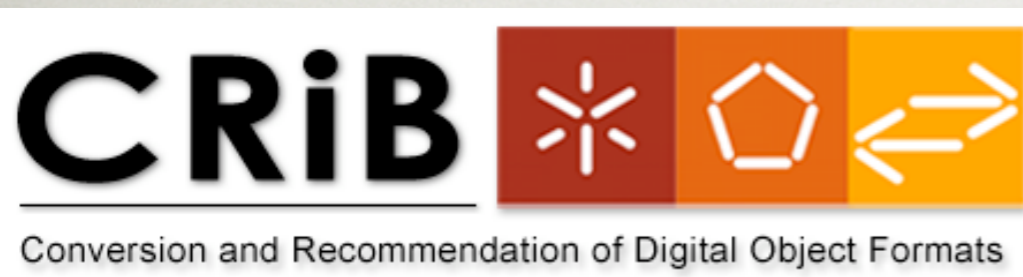
- Preservação a longo prazo e Autenticidade
- Basedo em standards (OAIS, EAD, PREMIS, METS, etc.)
- Seguindo os requisitos do TRAC (Trustworthy Repositories Audit & Certification)
- Com requisitos de segurança
- Service Oriented Architecture (SOA)
- Design simples e apelativo
- Open source
- DGArc e Universidade do Minho

CONTEXTO



RODA (2006-2009)

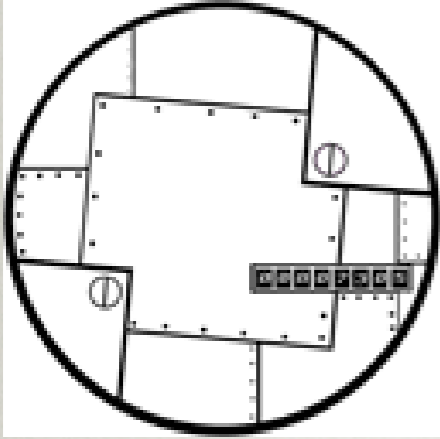
- Gestão de Metainformação (baseada em EAD)
- Gestão de Objectos Digitais (...)
- Políticas e protocolos de Preservação Digital
- Projecto nacional (Direcção Geral de Arquivos)



CRiB: Serviços de Preservação para Repositórios Digitais (2005-2008)

- Serviço de Migração Distribuído
- Serviço de Aconselhamento e Suporte à Migração
- Tese de Doutoramento / U. Minho (Miguel Ferreira)

CONTEXTO



RODA (2006-2009)

- Gestão de Metainformação (baseada em EAD)
- Gestão de Objectos Digitais (...)
- Políticas e protocolos de Preservação Digital
- Projecto nacional (Direcção Geral de Arquivos)



CRiB: Serviços de Preservação para Repositórios Digitais (2005-2008)

- Serviço de Migração Distribuído
- Serviço de Aconselhamento e Suporte à Migração
- Tese de Doutoramento / U. Minho (Miguel Ferreira)

REQUISITOS/QUESTÕES?

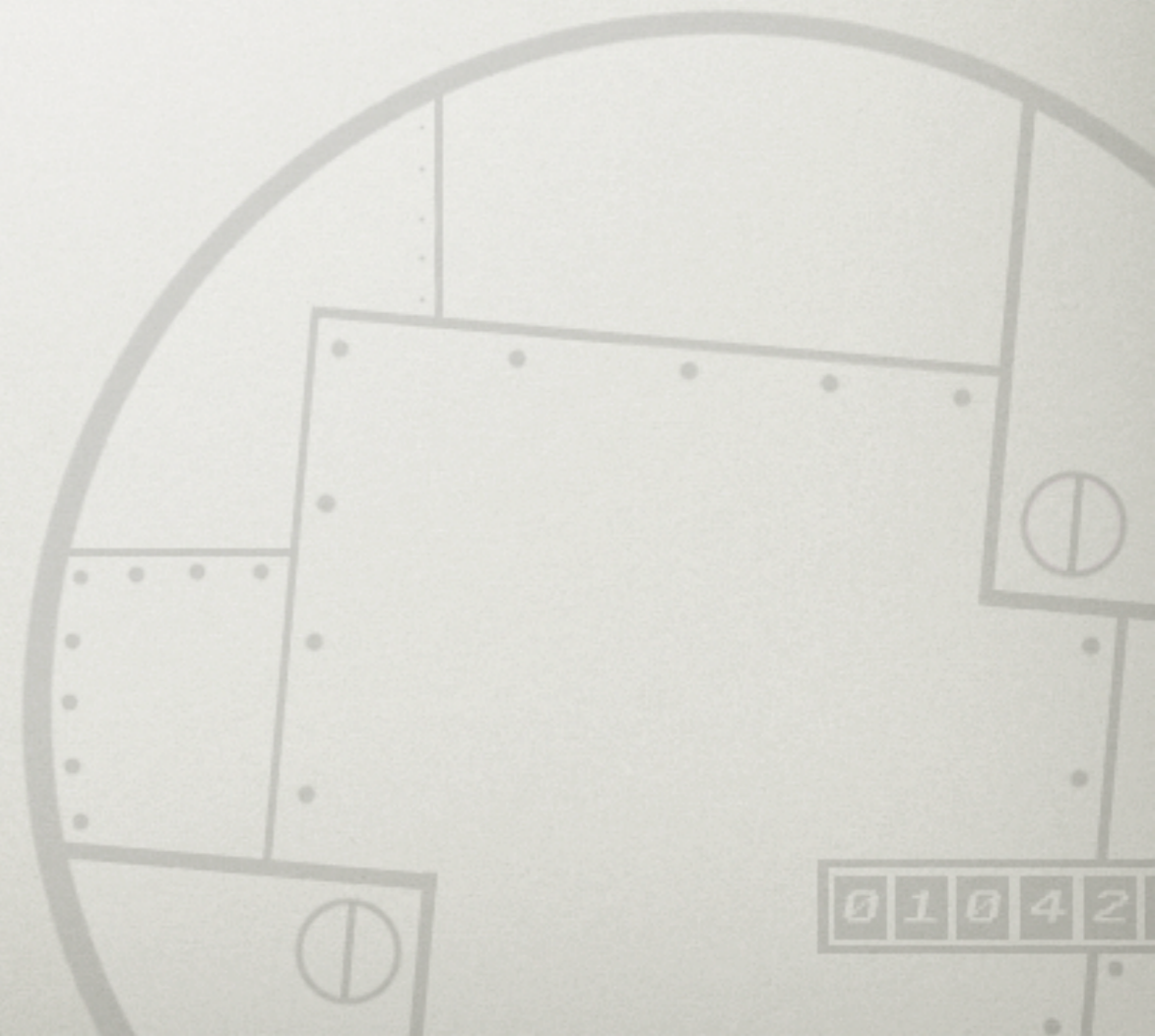
- Como garantir a **autenticidade**?
- Como **descrever e classificar** ODs?
- Como **implementar** a Preservação Digital?



AUTENTICIDADE

Temos de confiar nas nossas fontes: na História não há discurso directo ou prova.

EX: a Bíblia



AUTENTICIDADE

Temos de confiar nas nossas fontes: na História não há discurso directo ou prova.

EX: a Bíblia

Como nos tornamos “alguém de confiança”?

AUTENTICIDADE

Temos de confiar nas nossas fontes: na História não há discurso directo ou prova.

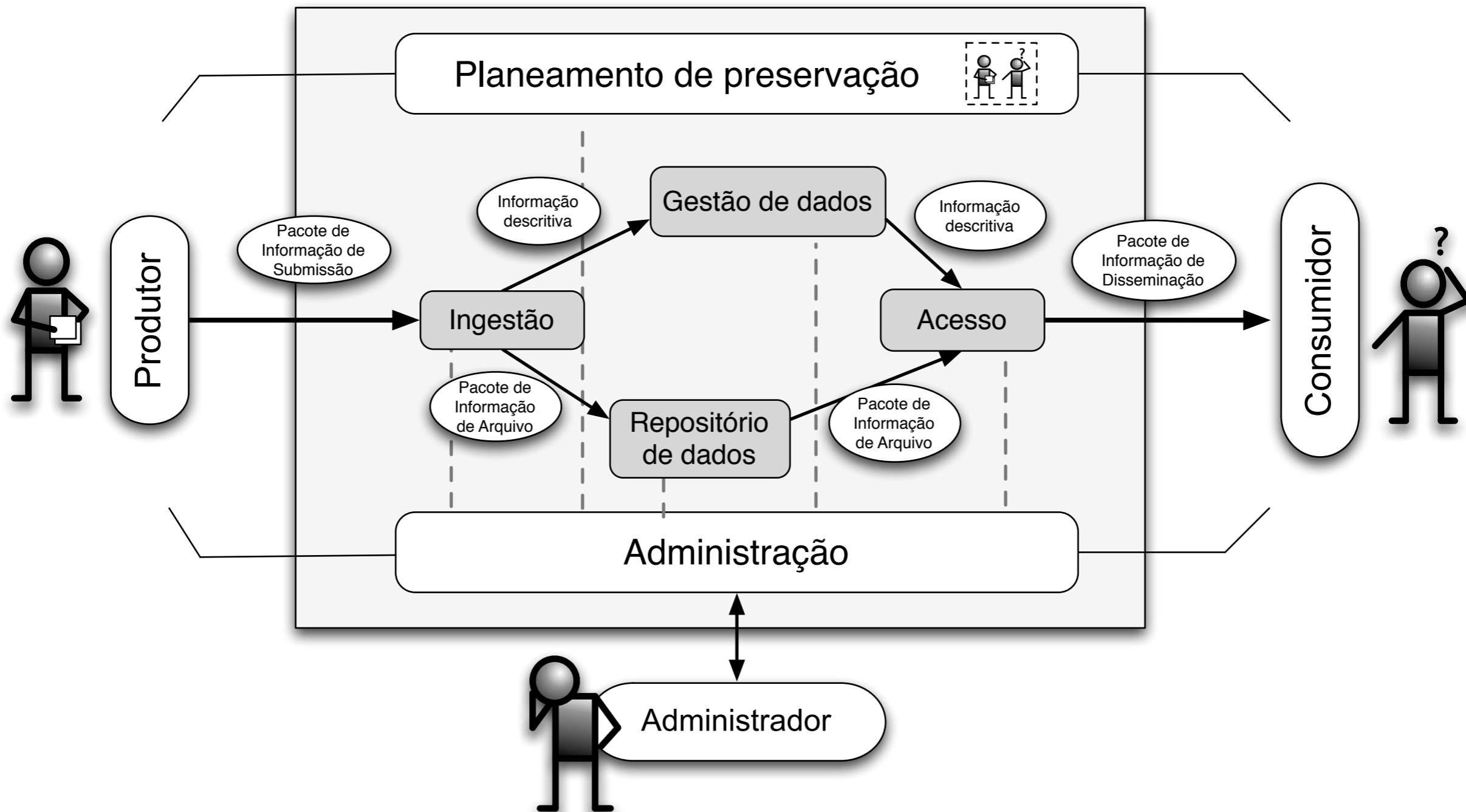
EX: a Bíblia

Como nos tornamos “alguém de confiança”?

- Reputação
- Documentando toda e qualquer acção realizada sobre os ODs

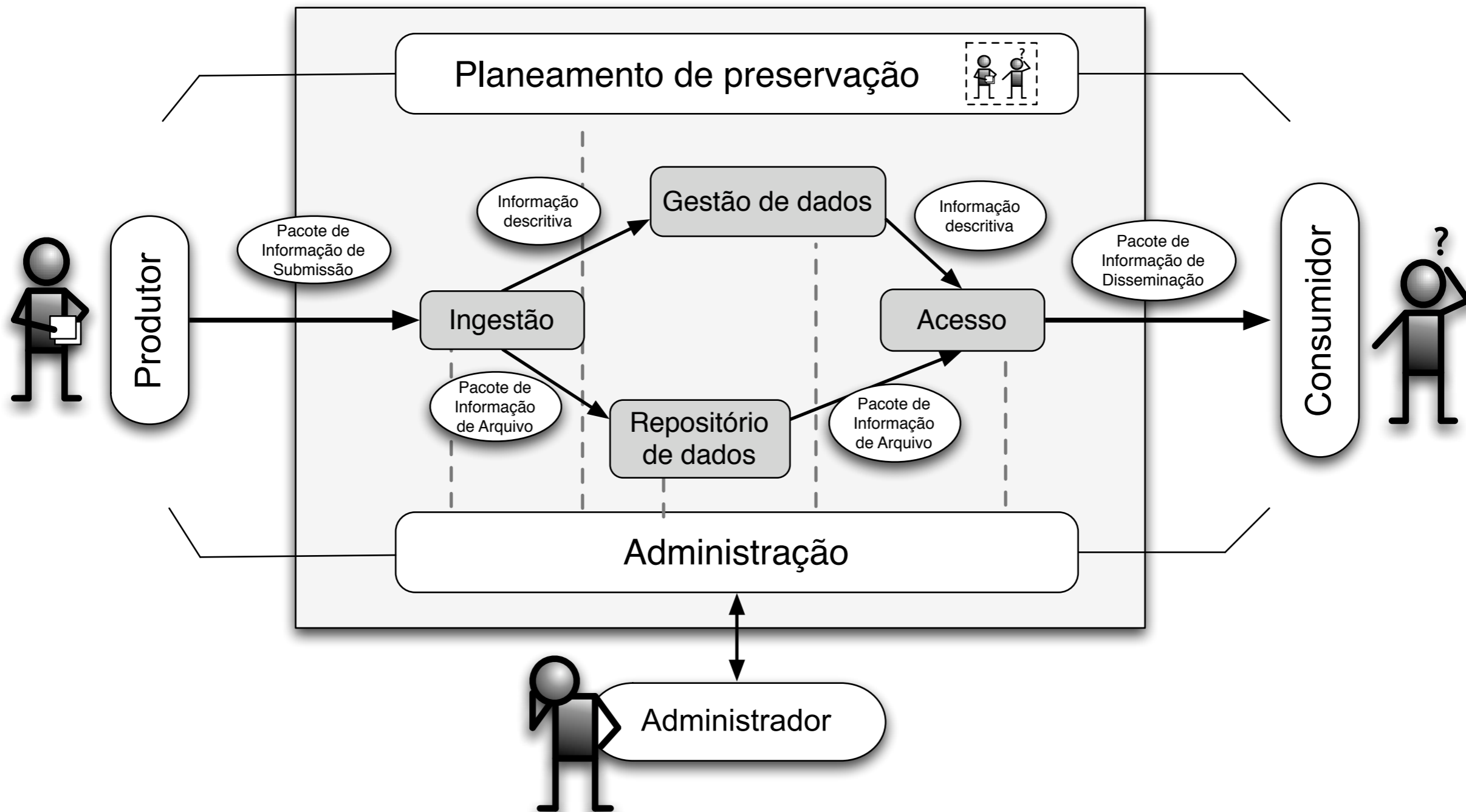
Open Archival Information System - OAIS

ISO 14721:2003



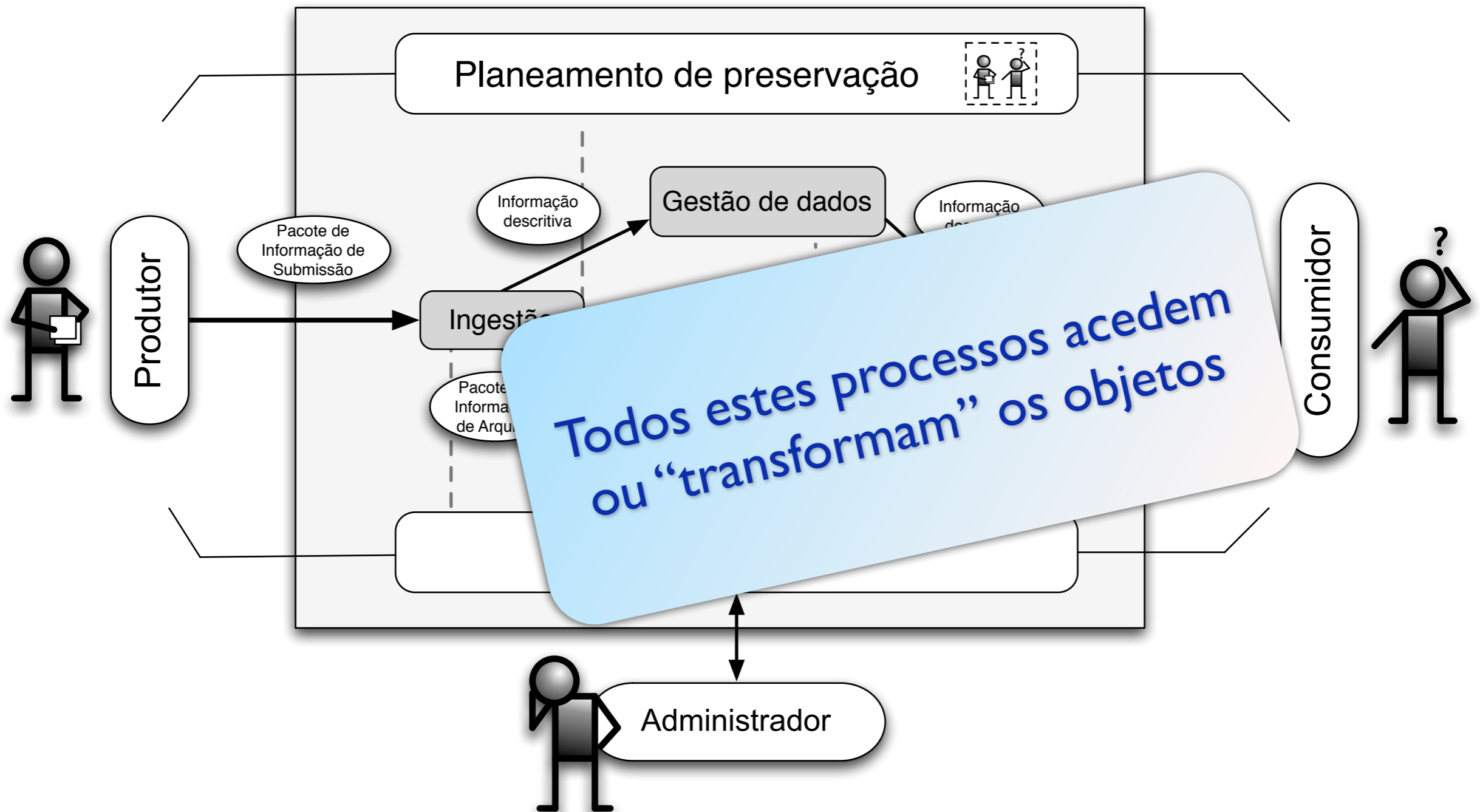
Open Archival Information System - OAIS

ISO 14721:2003



Open Archival Information System - OAIS

ISO 14721:2003



O presente: 2012 - ...

ISO 16363

Audit and Certification of Trustworthy Digital
Repositories

O ISO 16363 deriva do TRAC, documento publicado em 2007 pela

- ▶ RLG (Research Library Group)
- ▶ NARA (National Archives and Records Administration)

Enumera um conjunto de **requisitos** que vão desde a gestão organizacional às infraestruturas **visam** aferir a **confiabilidade de um repositório**

O TRAC tornou-se uma norma ISO em 2012

- ▶ ISO/DIS 16363 – Audit and certification of **trustworthy digital repositories = TRAC**
- ▶ ISO/DIS 16919 – **Requirements for bodies providing audit and certification of candidate trustworthy digital repositories**

Os **objetivos** da norma são:

- ▶ Fornecer uma ferramenta que permita **auditar, avaliar**, e potencialmente **certificar** repositórios digitais
- ▶ Estabelecer a **documentação necessária** para realizar uma auditoria
- ▶ **Delinear o processo de certificação**
- ▶ Estabelecer metodologias apropriadas para **determinar a robustez e a sustentabilidade de um repositório digital**

A aplicação da norma **potencia a confiança** junto dos utilizadores do repositório pois

- ▶ Estabelece um **clima de transparência** relativamente aos processos implementados
- ▶ **Auxilia na realização de auditorias internas e externas**

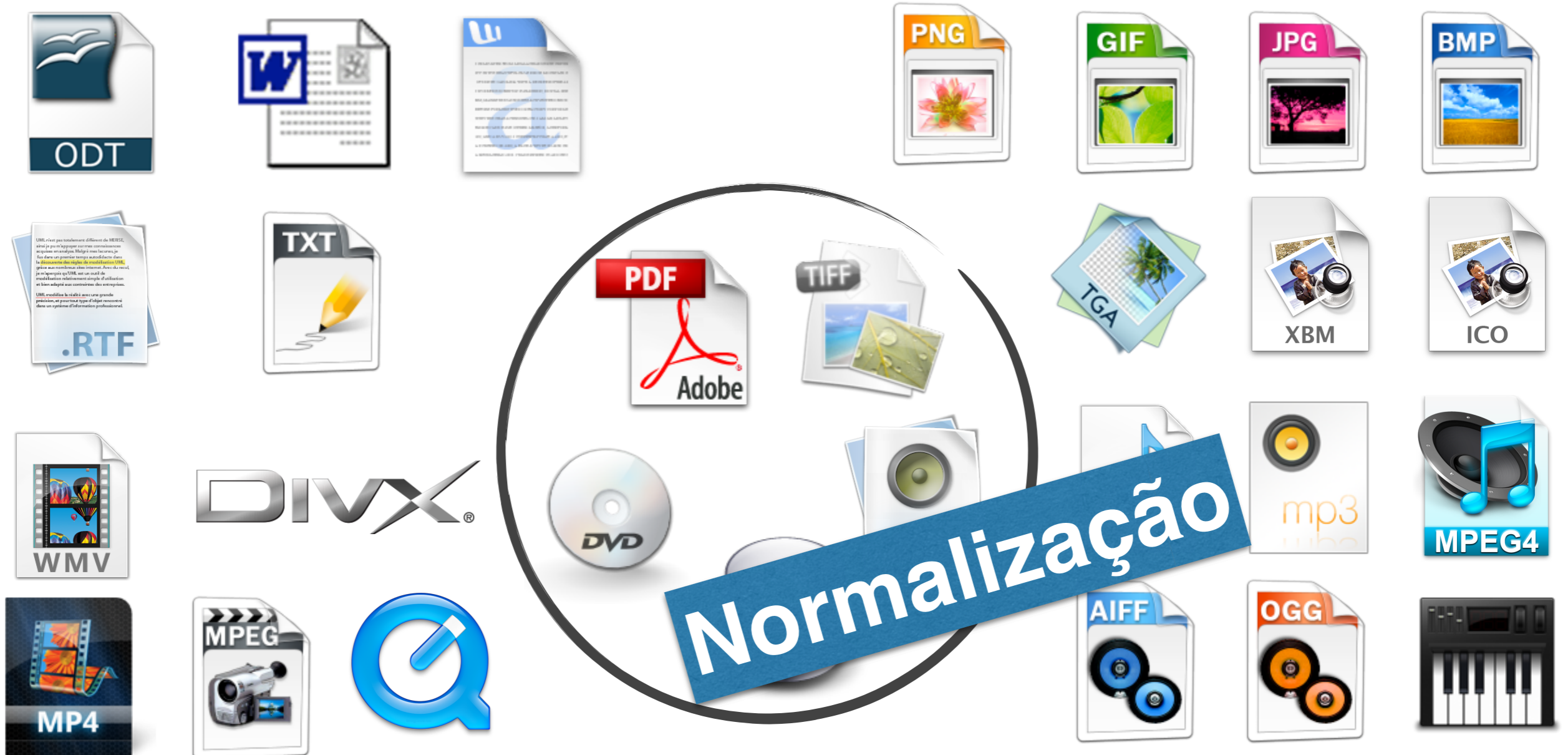
Classes de ODs



Classes de ODs



Classes de ODs



DIVX



PostgreSQL



Classes de ODs



DIVX



Normalização
Propriedades Significativas



Anatomia de um OD

Nível
Conceptual

Nível
Lógico

Nível
Físico

Base de
Dados

Doc. Texto
Imagem

SQL Server

Access

PDF Doc.

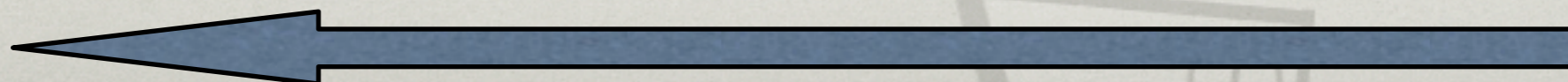
Ms Word Doc.

PNG image

Disco Duro

Tape

...

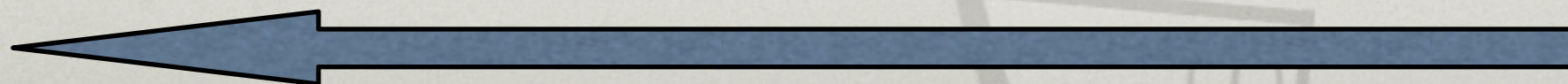
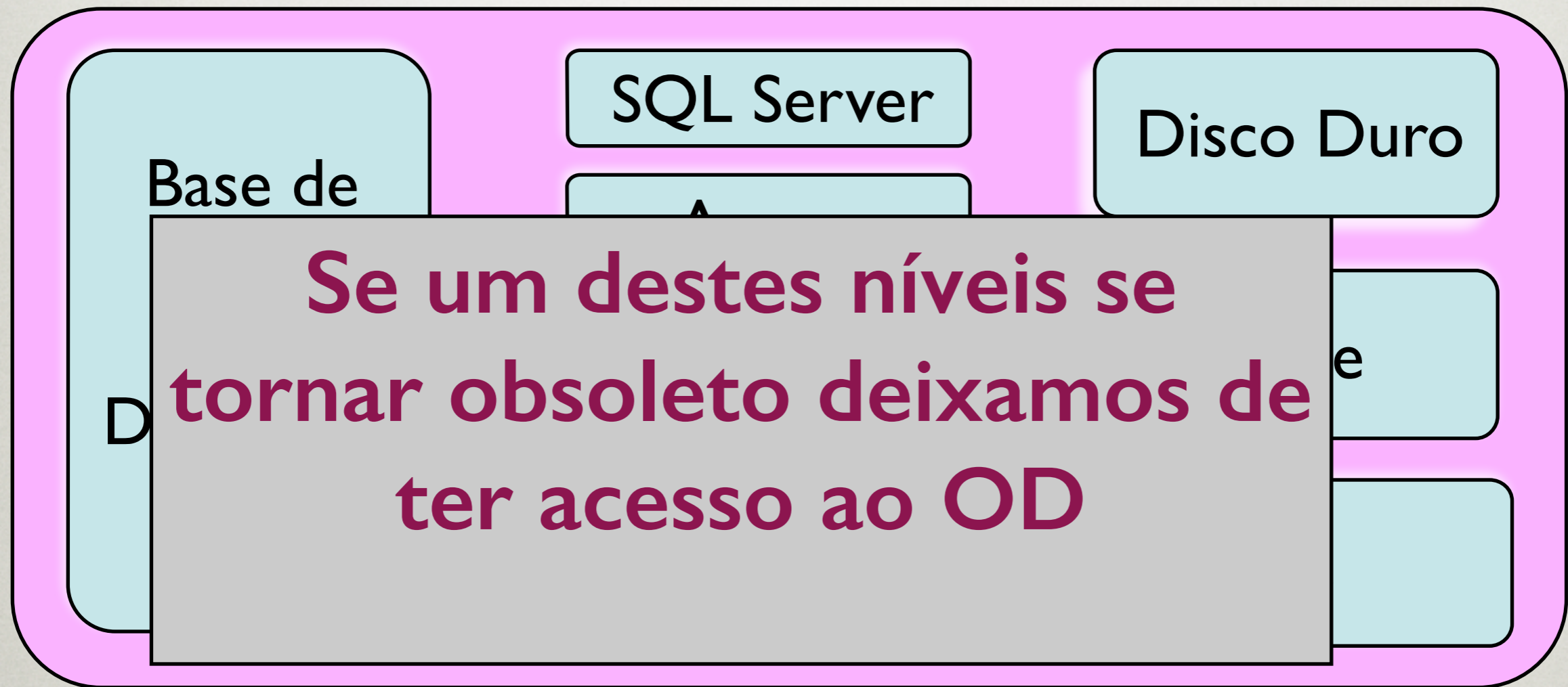


Anatomia de um OD

Nível
Conceptual

Nível
Lógico

Nível
Físico



O presente: 2012 - ... (tese mestrado)

Depois do encontro de Copenhaga (Arq.Nac.Dinamarca 2012):

O presente: 2012 - ... (tese mestrado)

Depois do encontro de Copenhaga (Arq.Nac.Dinamarca 2012):

Database Preservation Toolkit

O presente: 2012 - ... (tese mestrado)

Depois do encontro de Copenhaga (Arq.Nac.Dinamarca 2012):

Database Preservation Toolkit

Developed within RODA project

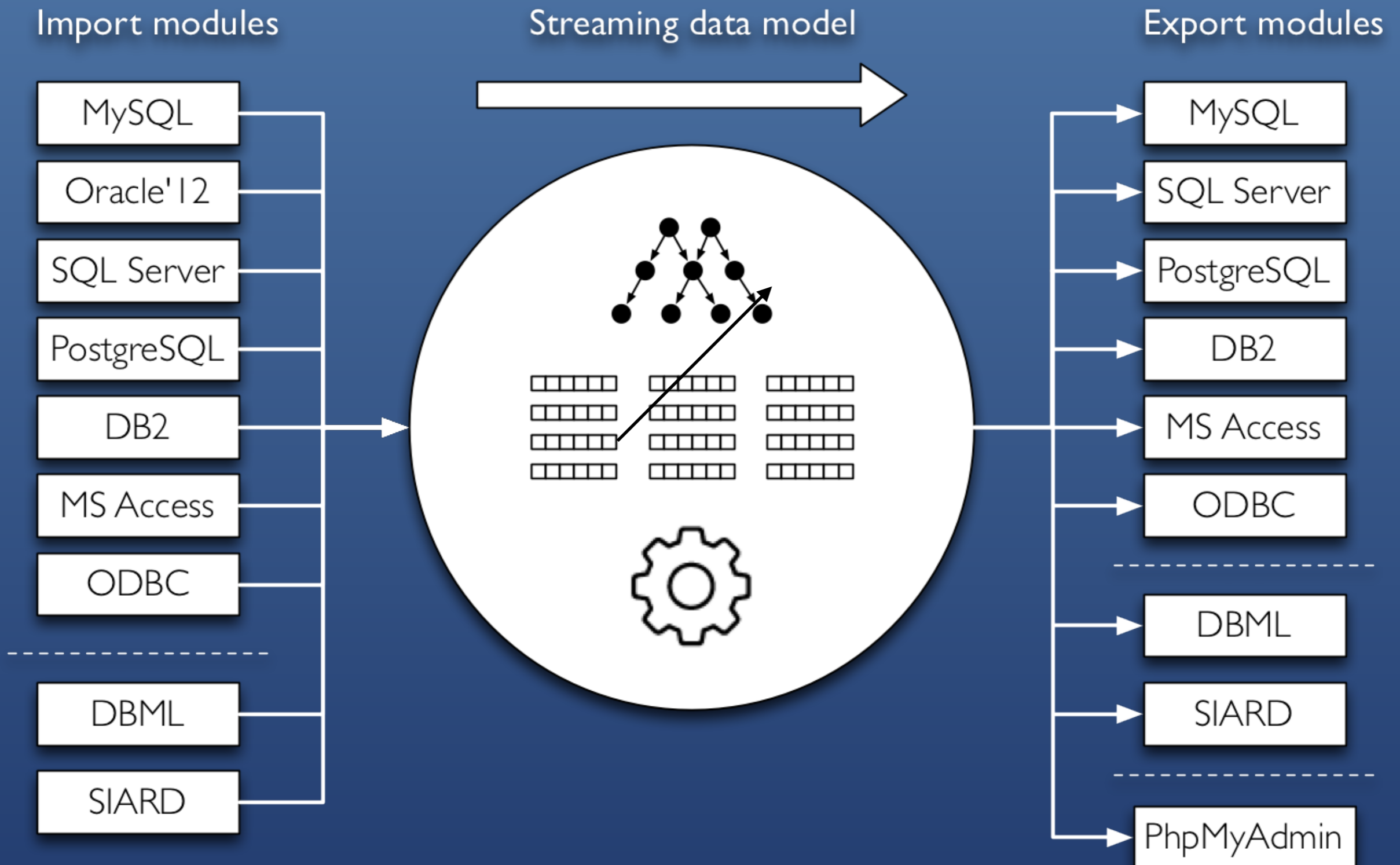
Now stand-alone open-source project

<http://keeps.github.io/db-preservation-toolkit/>

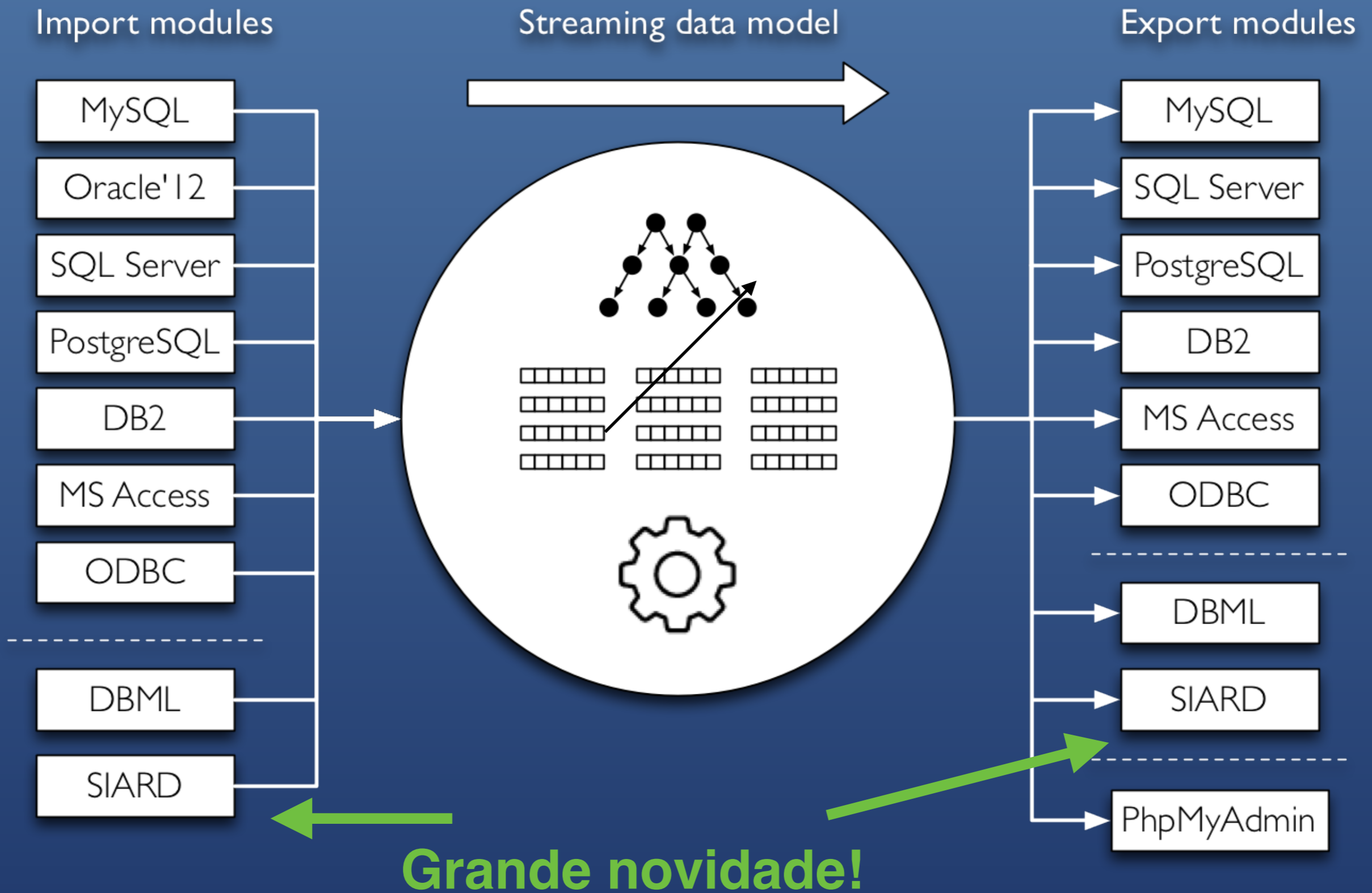
Imports and exports between DBMS and DB formats

Supports preservation formats: DBML, SIARD

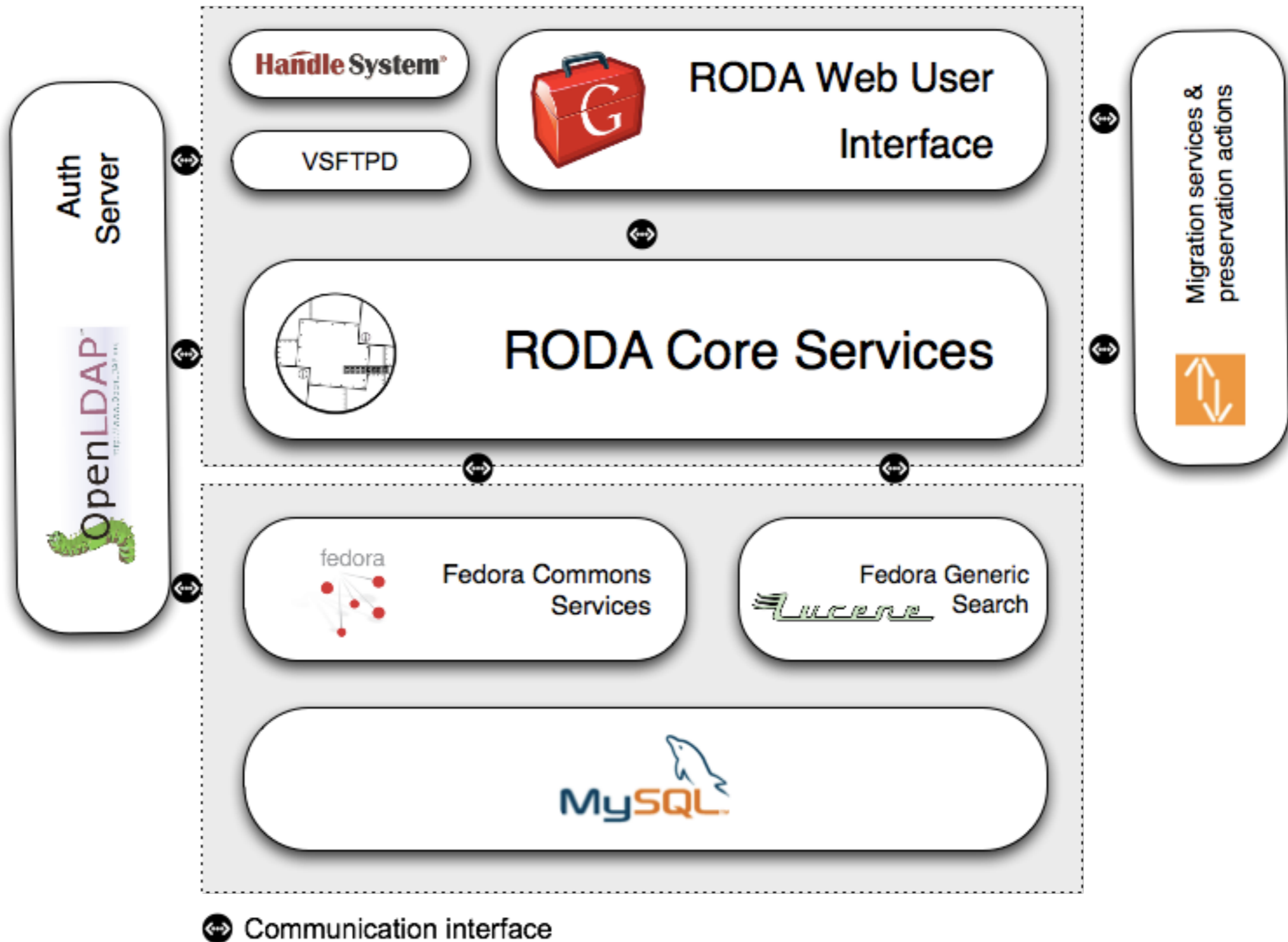
db-preservation-toolkit architecture



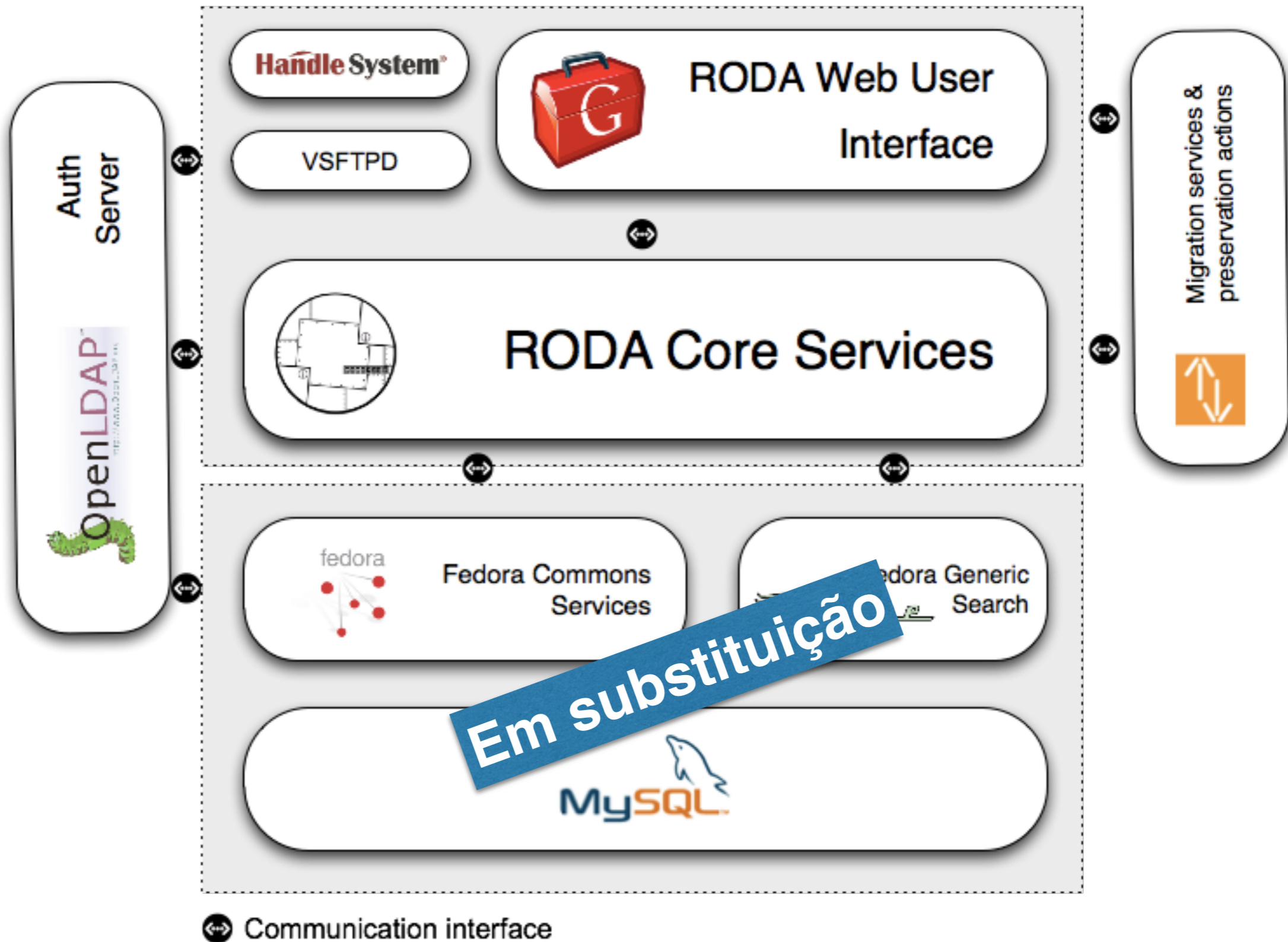
db-preservation-toolkit architecture



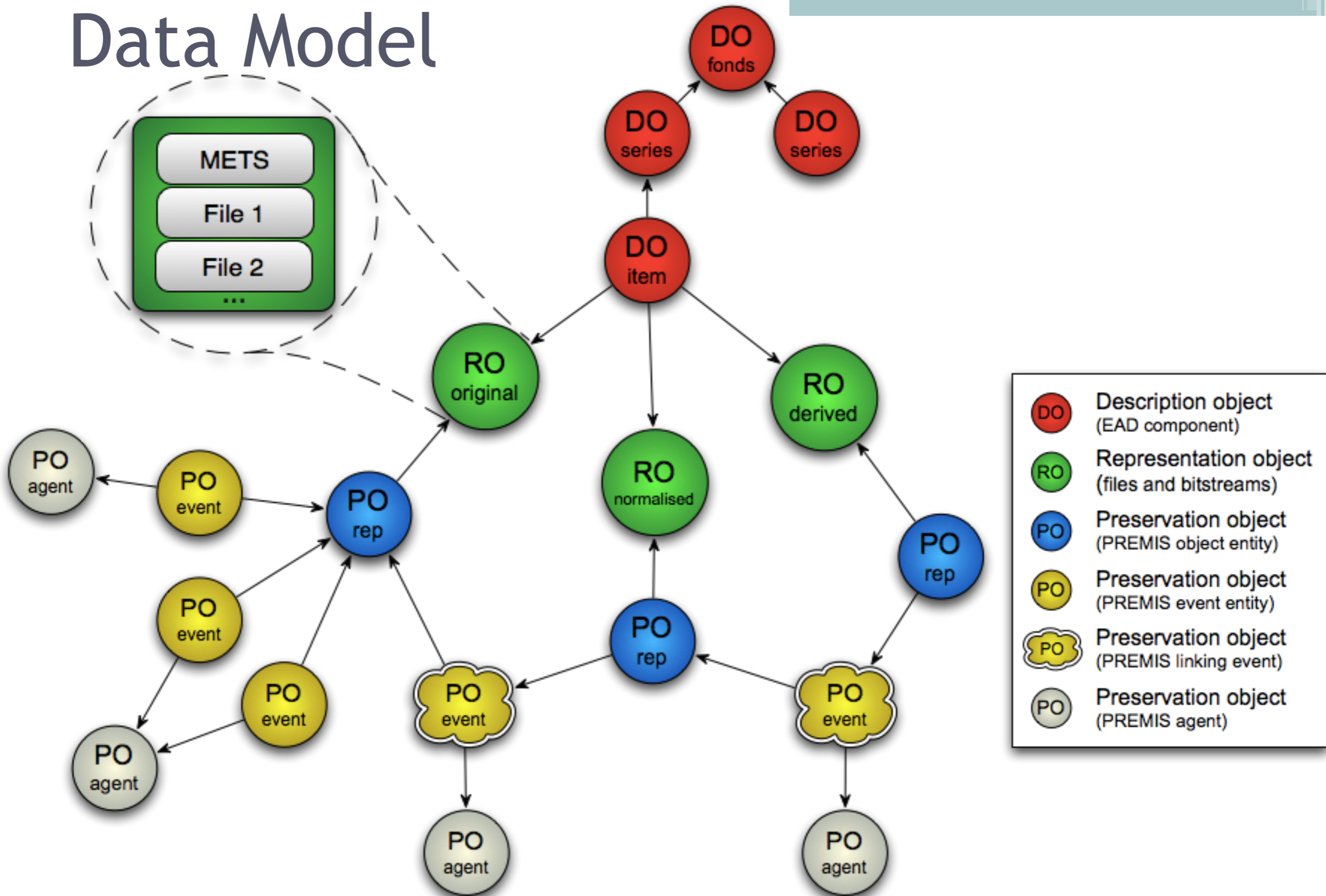
Architecture



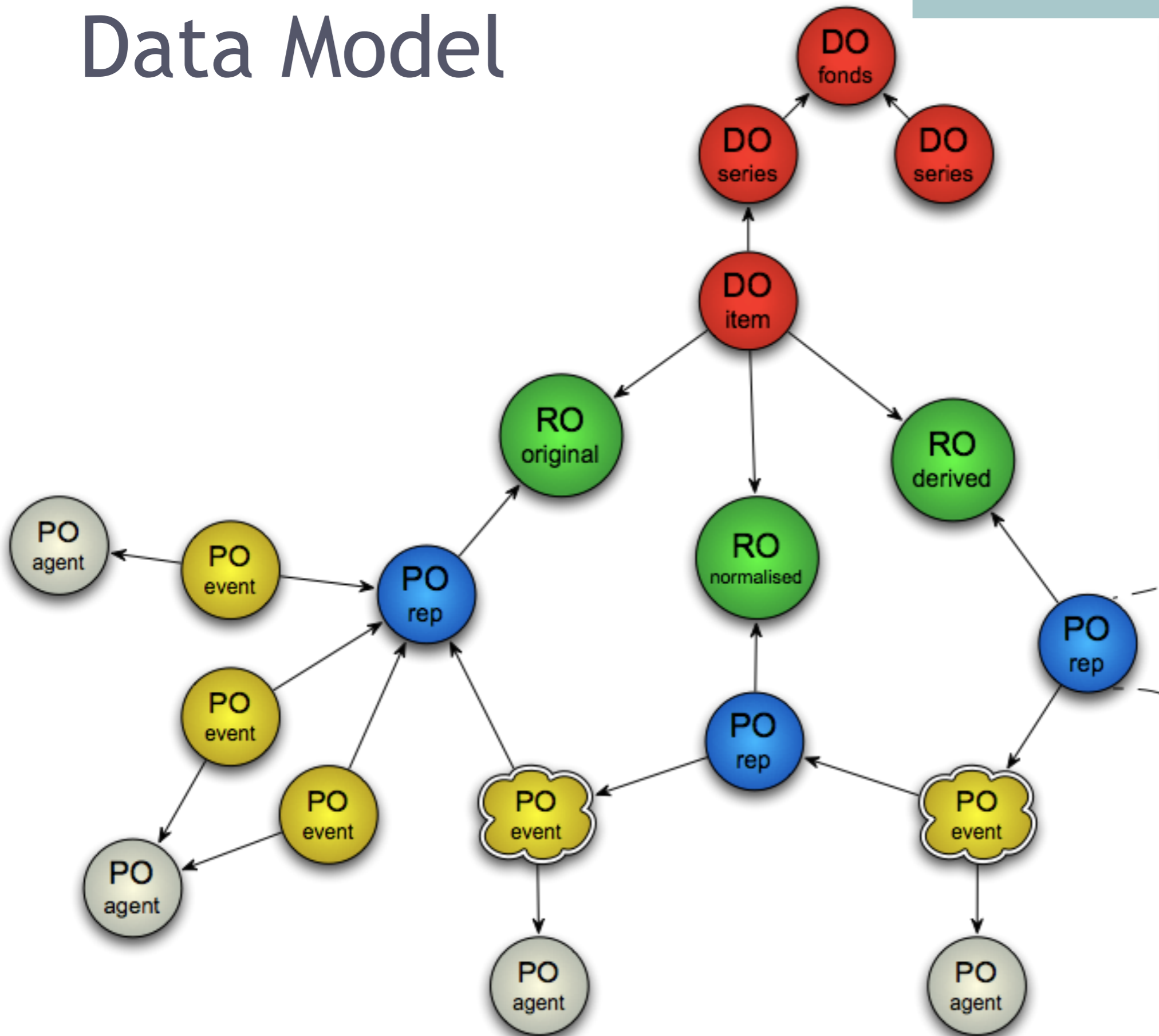
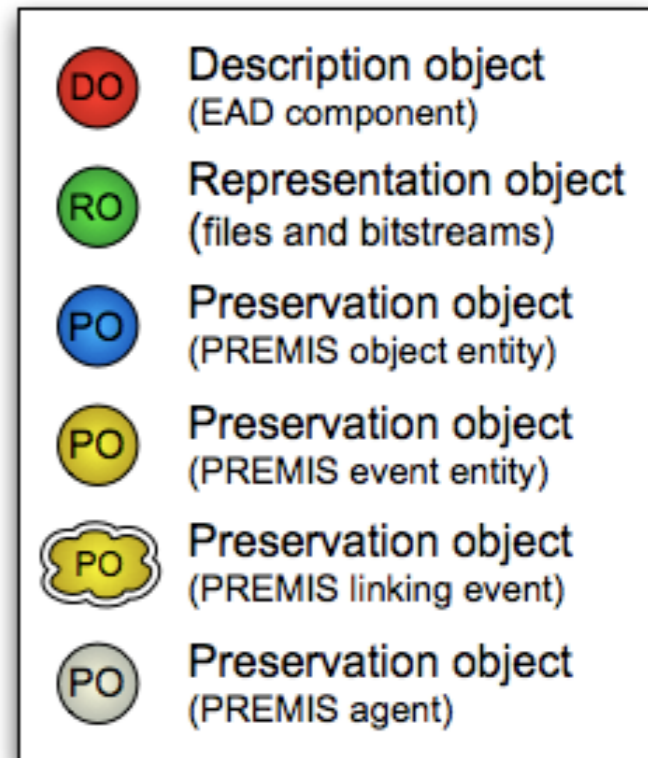
Architecture



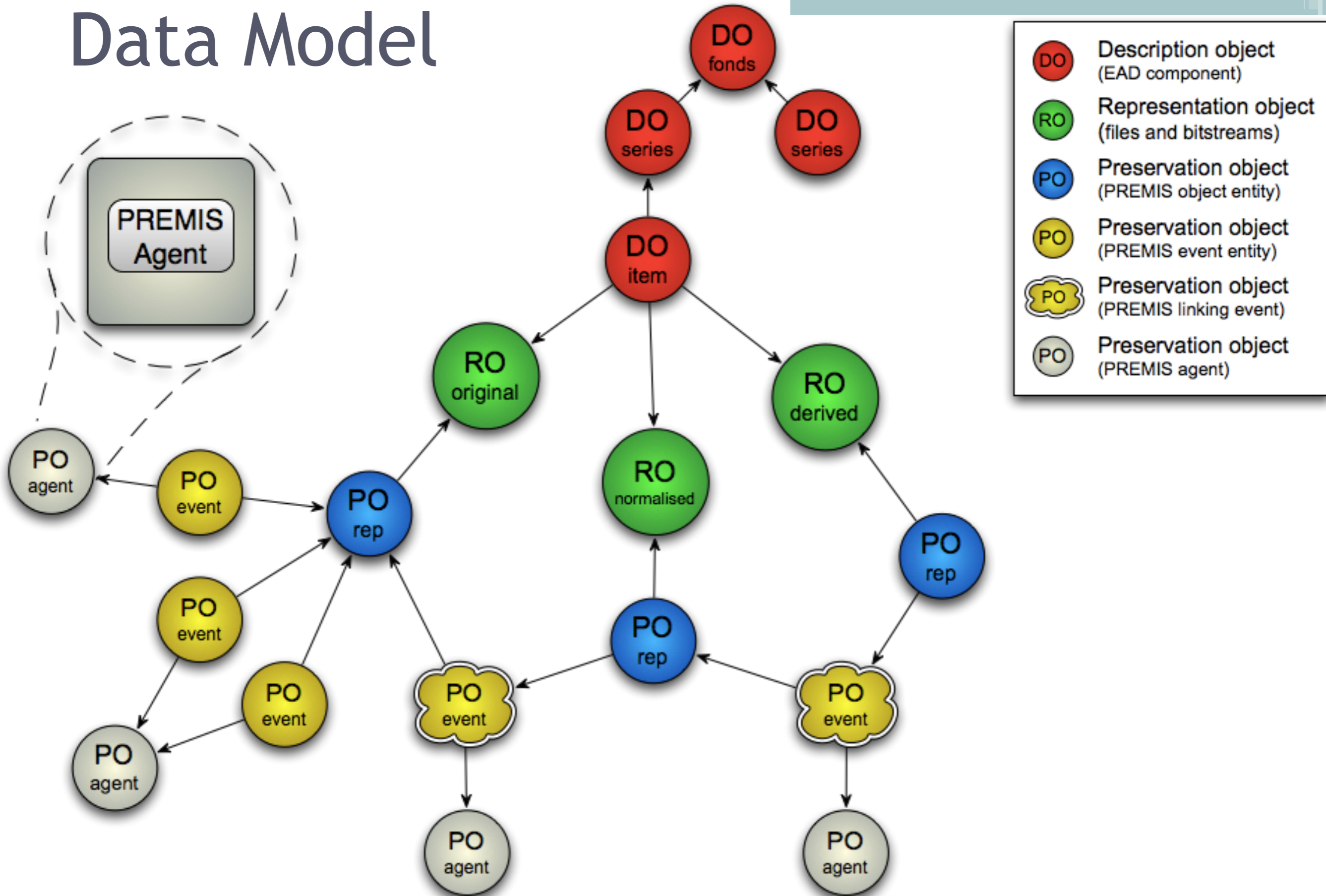
Data Model



Data Model

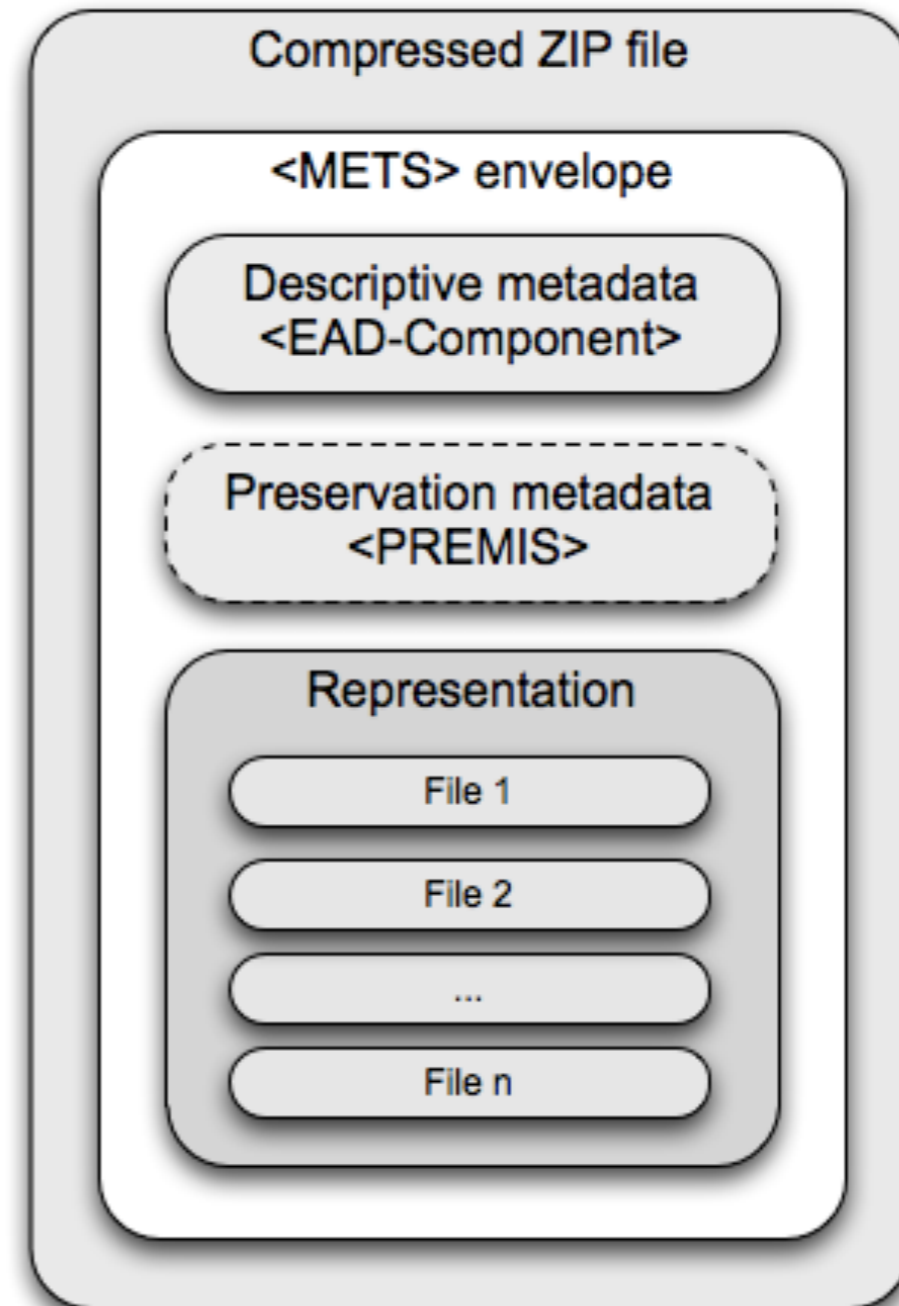


Data Model



Estrutura do SIP

- Identificação do OD e seu formato
- Localização no plano de classificação
- Metadados descritivos
- Metadados de Preservação
- Metadados Técnicos
- Representações
- Identificação do ficheiro principal/raíz



ABOUT RODA

SERVICES · POLICY AND TECHNICAL DOCUMENTS · R&D · CONTACTS · ABOUT THE REGISTER · HELP

CATALOG

BROWSE · SEARCH · ADVANCED SEARCH · HELP

INGEST

PRE-INGEST · SUBMIT · STATUS · HELP

ADMINISTRATION

USERS · SCHEDULER · STATISTICS · LOG · HELP

catalog / browse / 50218

BROWSE

4 fonds

Reference

- AACC
- DGARQ
- TEST
 - 692
 - LFARIA
 - sf1
 - c1
 - sr1
 - 101
 - 815
 - 1030
 - 50192
 - 50218
- fbarbedo

Identification

Reference	50218
Complete reference	PT/DGARQ/TEST/sf1/c1/sr1/50218
Handle	http://hdl.handle.net/10384/50218
Title	National Geographic Photos
Level	item
Initial date	2008-01-01
Final date	2009-01-01
Country code	PT
Repository code	DGARQ
Origination	National Geographic
Acquisition number	NG
Acquisition date	2009-05-11

Content and structure

Scope and content	tests for or09
-------------------	----------------

Notes

Note	Pacote de submissão validado semanticamente por Ifaria
------	--

EDIT

SUBLEVEL

MOVE

EAD

CLOSE

Identification

Reference	50218
Complete reference	PT/DGARQ/TEST/sf1/c1/sr1/50218
Handle	http://hdl.handle.net/10384/50218
Title	National Geographic Photos
Level	item
Initial date	2008-01-01
Final date	2009-01-01
Country code	PT
Repository code	DGARQ
Origination	National Geographic
Acquisition number	NG
Acquisition date	2009-05-11

Content and structure

Scope and content	tests for or09
-------------------	----------------

Notes

Note	Pacote de submissão validado semanticamente por Ifaria
------	--

EDIT

SUBLEVEL

MOVE

EAD

CLOSE

Identification


Reference	50218
Complete reference	PT/DGARQ/TEST/sf1/c1/sr1/50218
Handle	http://hdl.handle.net/10384/50218
Title	National Geographic Photos
Level	item
Initial date	2008-01-01
Final date	2009-01-01
Country code	PT
Repository code	DGARQ
Origination	National Geographic
Acquisition number	NG
Acquisition date	2009-05-11

Content and structure


Scope and content	tests for or09
-------------------	----------------


Notes


Note	Pacote de submissão validado semanticamente por Ifaria
------	--

EDIT 

SUBLEVEL 

MOVE 

EAD 

CLOSE 

Description

Visualization

Preservation

Permissions

Disseminassions of 50218 - "National Geographic Photos"



JPEG
(original)



TIFF
(normalized)

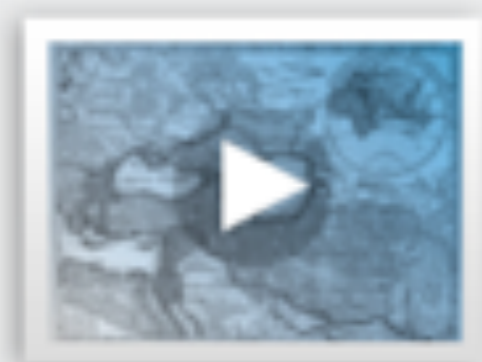


Photo pre-viewer



Book pre-viewer

CLOSE



Description

Visualization

Preservation

Permissions

Disseminassions of 50218 - "National Geographic Photos"



JPEG
(original)



TIFF
(normalized)



Photo pre-viewer



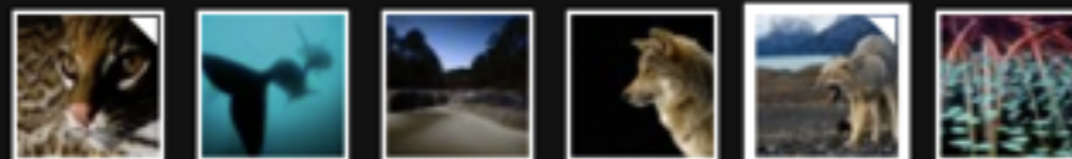
Book pre-viewer

CLOSE





National Geographic Photos



(5/13) gray-fox-dusk-1201638-xl.jpg.tiff

SIMPLEVIEWER

ABOUT RODA

SERVICES · POLICY AND TECHNICAL DOCUMENTS · R&D · CONTACTS · ABOUT THE REGISTER · HELP

CATALOG

BROWSE · SEARCH · ADVANCED SEARCH · HELP

INGEST

PRE-INGEST · SUBMIT · STATUS · HELP

ADMINISTRATION

USERS · SCHEDULER · STATISTICS · LOG · HELP

catalog / browse / 50218

BROWSE

4 fonds

Reference

- AACC
- DGARQ
- TEST
 - 692
 - LFARIA
 - sf1
 - c1
 - sr1
 - 101
 - 815
 - 1030
 - 50192
 - 50218
- fbarbedo

Description
 Visualization
 Preservation
 Permissions

roda:50219 (original)
 roda:50222 (normalized)

ingestion
 normalization

Day

Apr	May 8	May 9	May 10	May 11	May 12	May 13	Jun

PREMIS

CLOSE

ingestion

Description: The objects inside the SIP were added to the preservation repository

Result: success

Ingest details:

Description Visualization Preservation Permissions

● roda:50219 (original) ● roda:50222 (normalized)



ingestion

Description: The objects inside the SIP were added to the preservation repository

Result: success

Ingest details:

○ SIP for ingested.com success.com

Day

PREMIS

CLOSE

- sr1
- 101
- 815
- 1030
- 50192
- 50218

Fork us on GitHub



RODA

The world's most advanced **open-source** digital repository



An open-source **digital repository** designed for preservation

RODA is a complete digital repository that delivers functionality for all the main units of the OAIS reference model. RODA is capable of ingesting, managing and providing access to the various types of digital content produced by large corporations or public bodies. RODA is based on open-source technologies and is supported by existing standards such as the OAIS, METS, EAD and PREMIS.



Conforms to open standards

RODA follows open standards using EAD for description metadata, PREMIS for preservation metadata, METS for structural metadata, several standards for technical metadata (e.g. NISO Z39.87 for digital still images).



Vendor independent

RODA is 100% built on top of open-source technologies. The entire infrastructure required to support RODA is vendor independent. This means that you may use the hardware and Linux distributions that best fit your institutional needs.



Scalable

The service-oriented nature of RODA's architecture allows the system to be highly



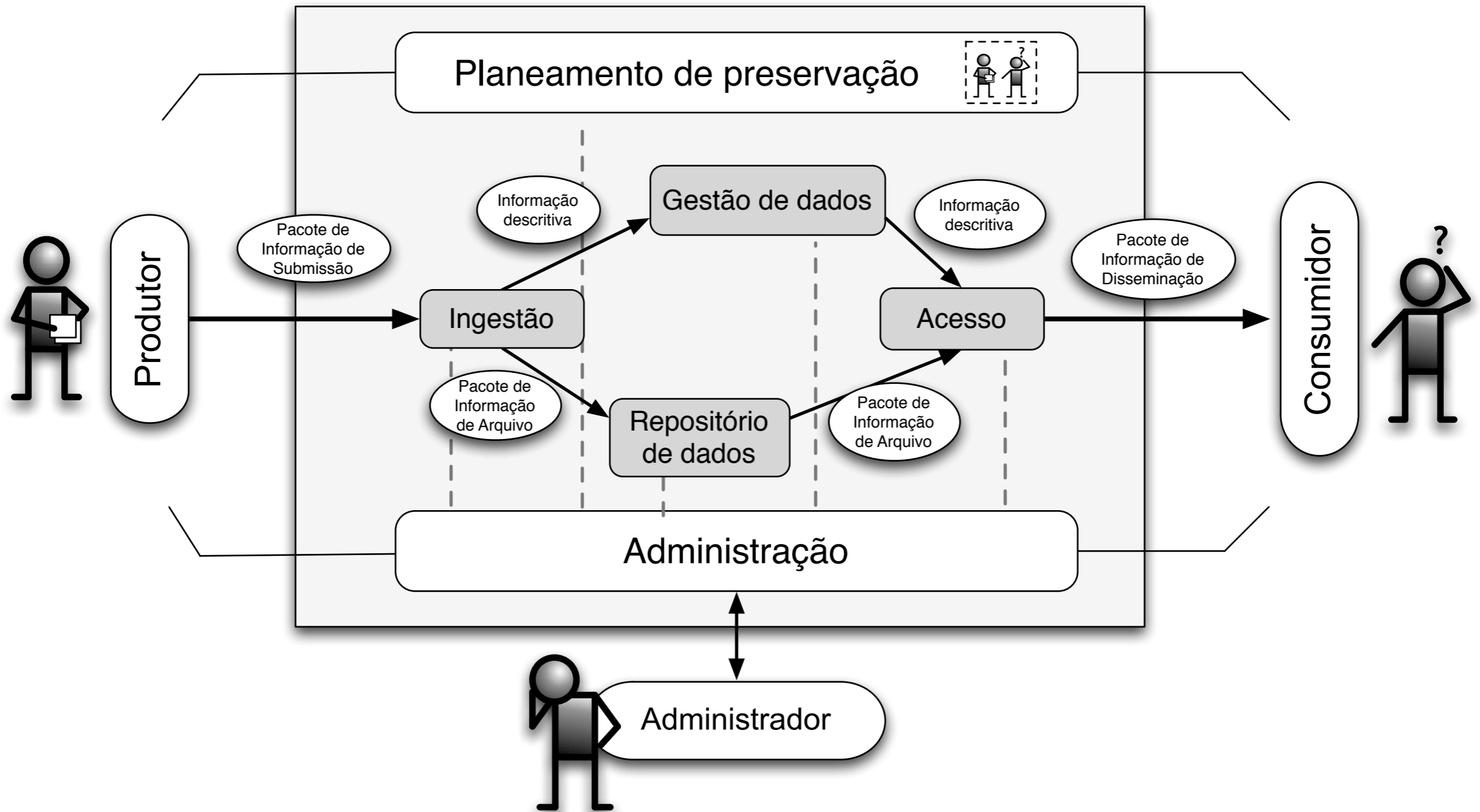
Embedded preservation actions

Preservation actions and management within RODA is handled by a task scheduler. The task

Presente e futuro:

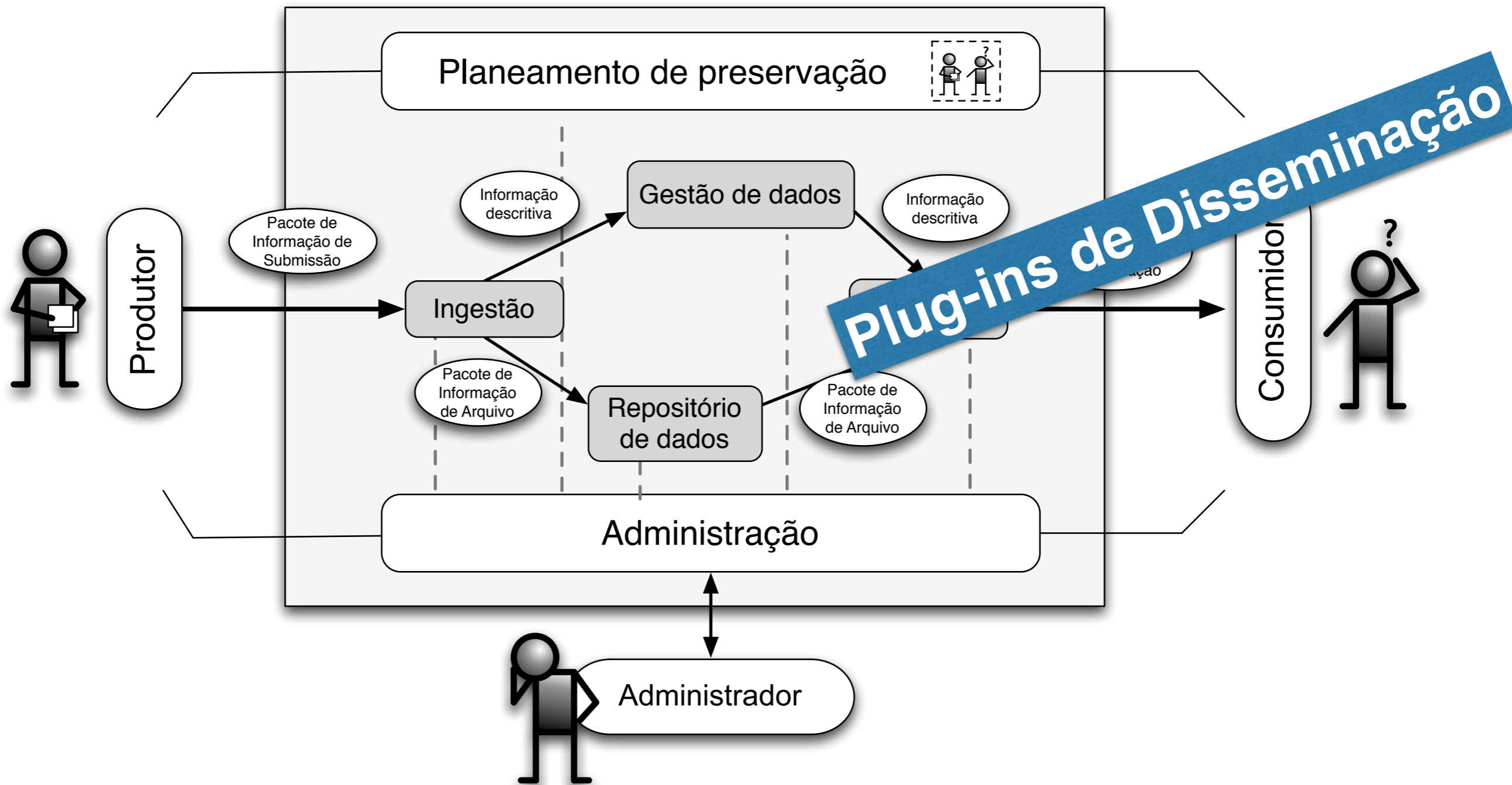
- Desenvolvimento continuado: projetos europeus
- Disponível em várias línguas: português, inglês e checo
- Em análise em várias instituições ao redor do globo
- É e continuará a ser Open Source
- Vários demonstradores em linha
- Problemas detetados:
 - Uniformização de SIPs e DIPs

Open Archival Information System - OAIS ISO 14721:2003



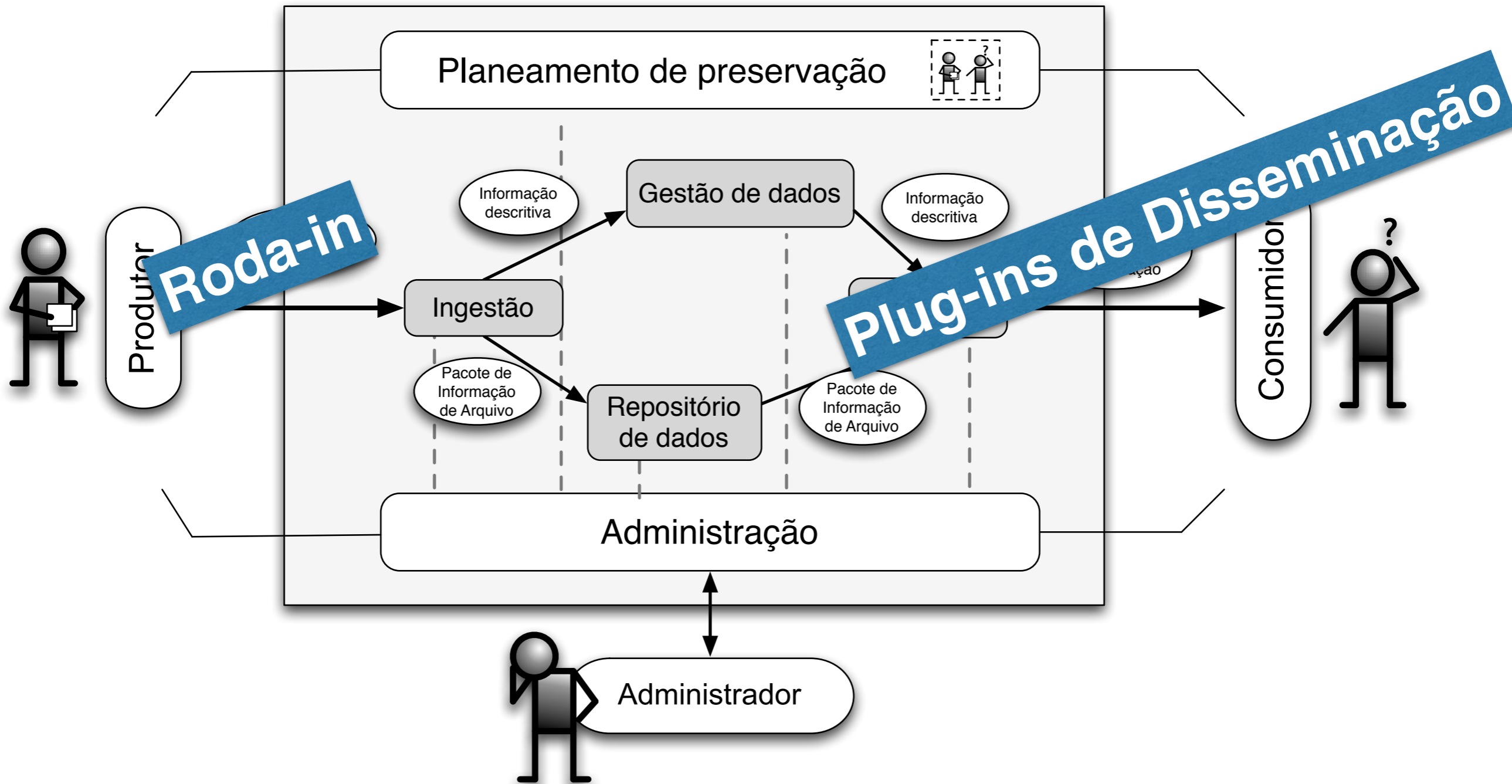
Nós:

Open Archival Information System - OAIS ISO 14721:2003



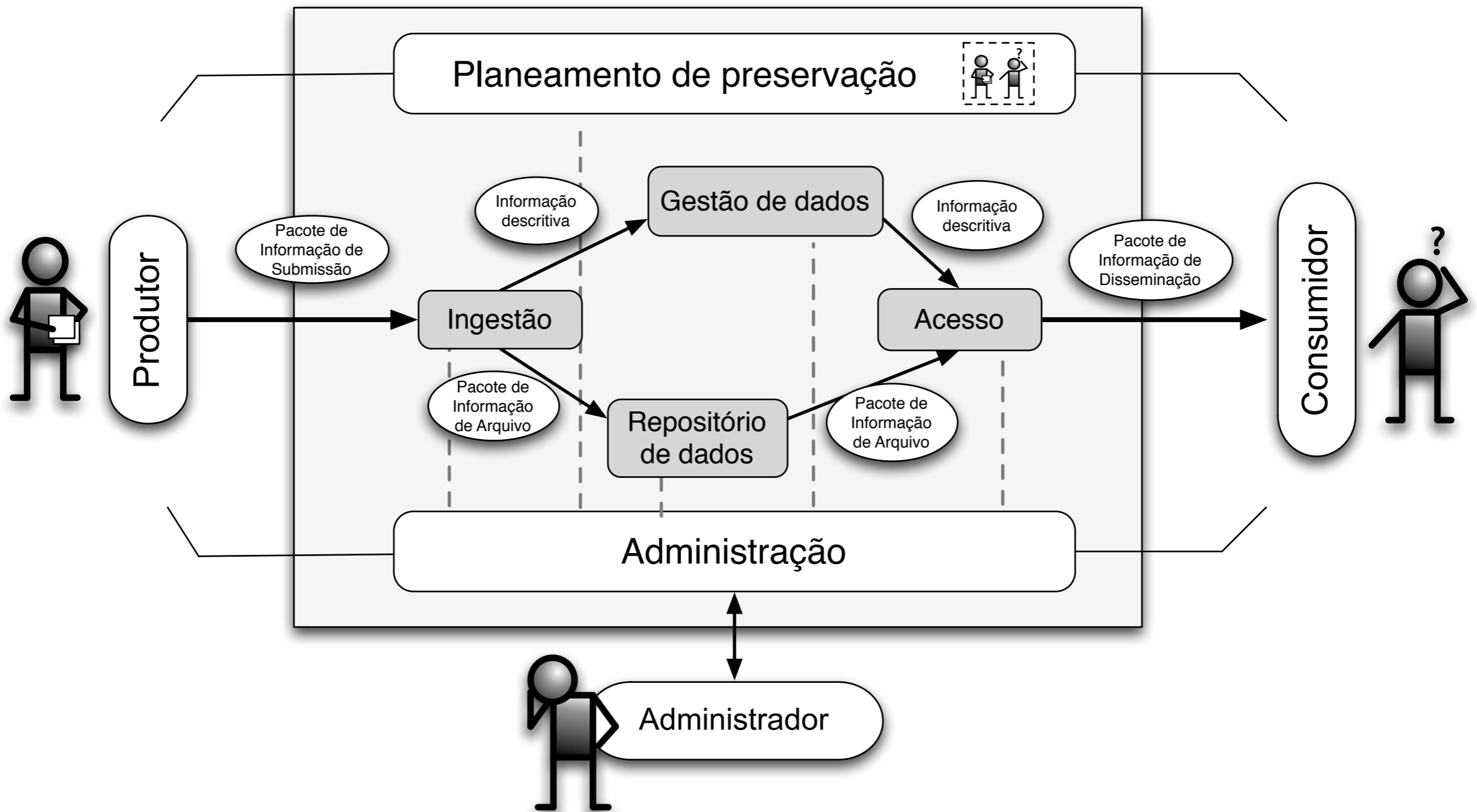
Nós:

Open Archival Information System - OAIS ISO 14721:2003



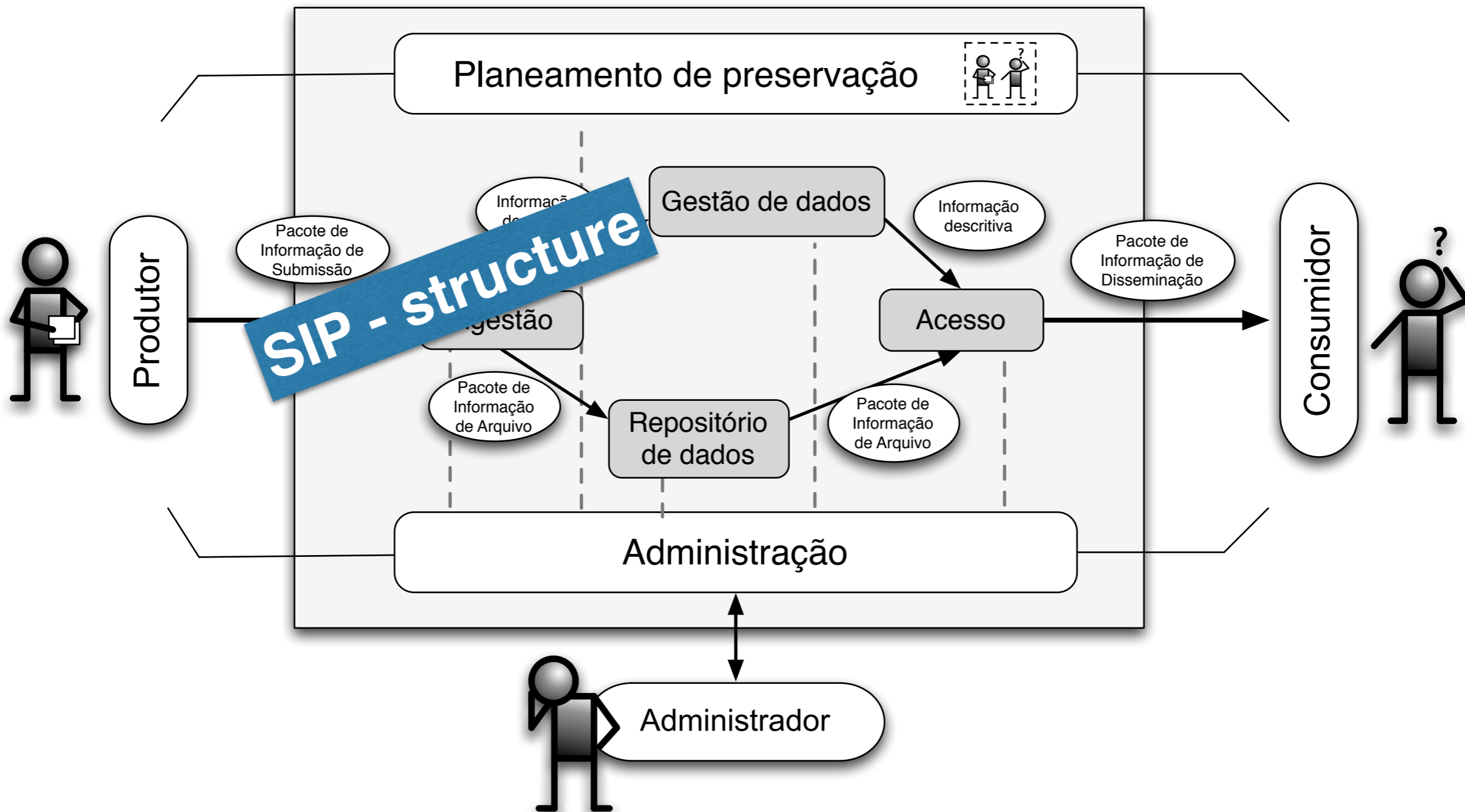
Os outros:

Open Archival Information System - OAIS ISO 14721:2003



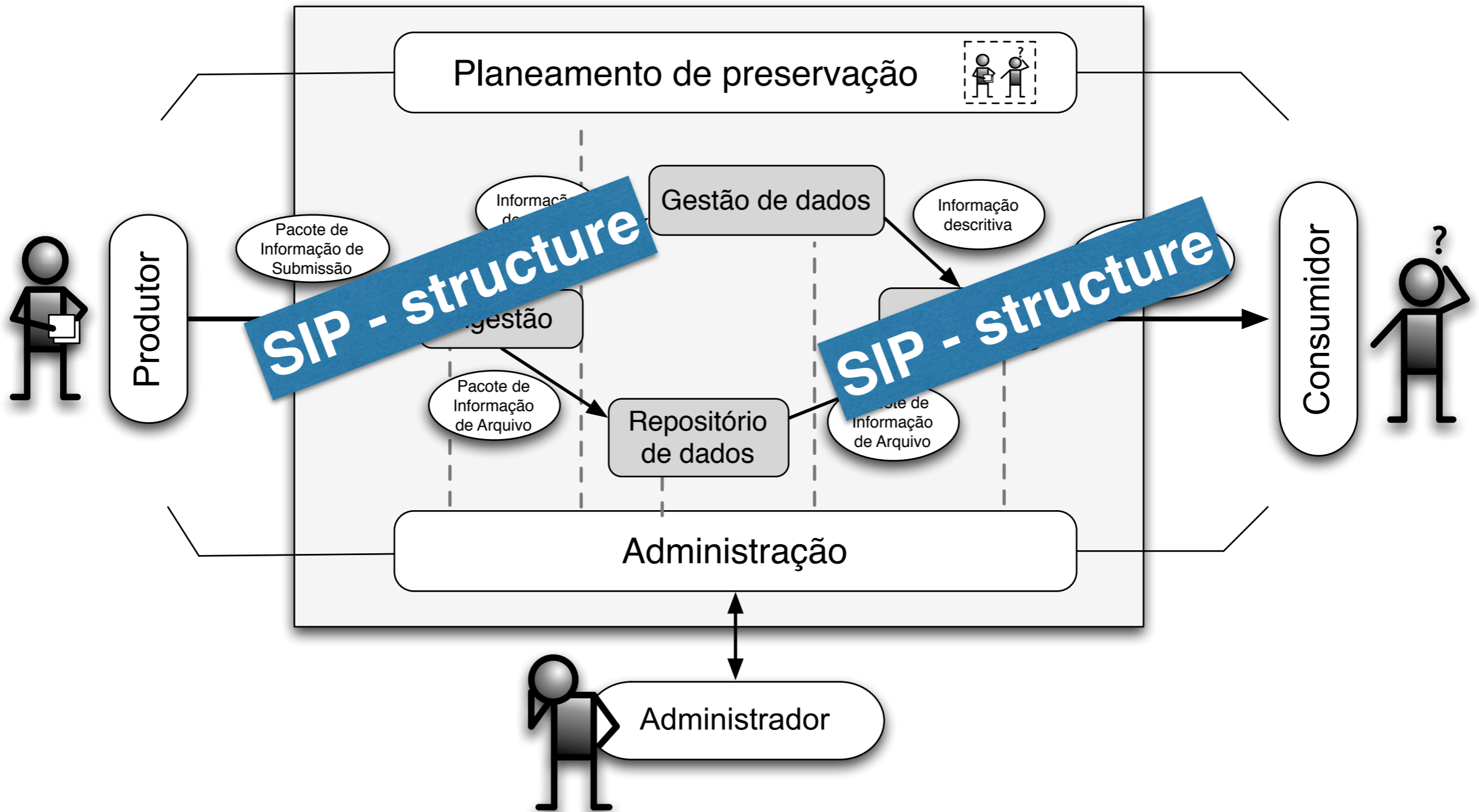
Os outros:

Open Archival Information System - OAIS ISO 14721:2003



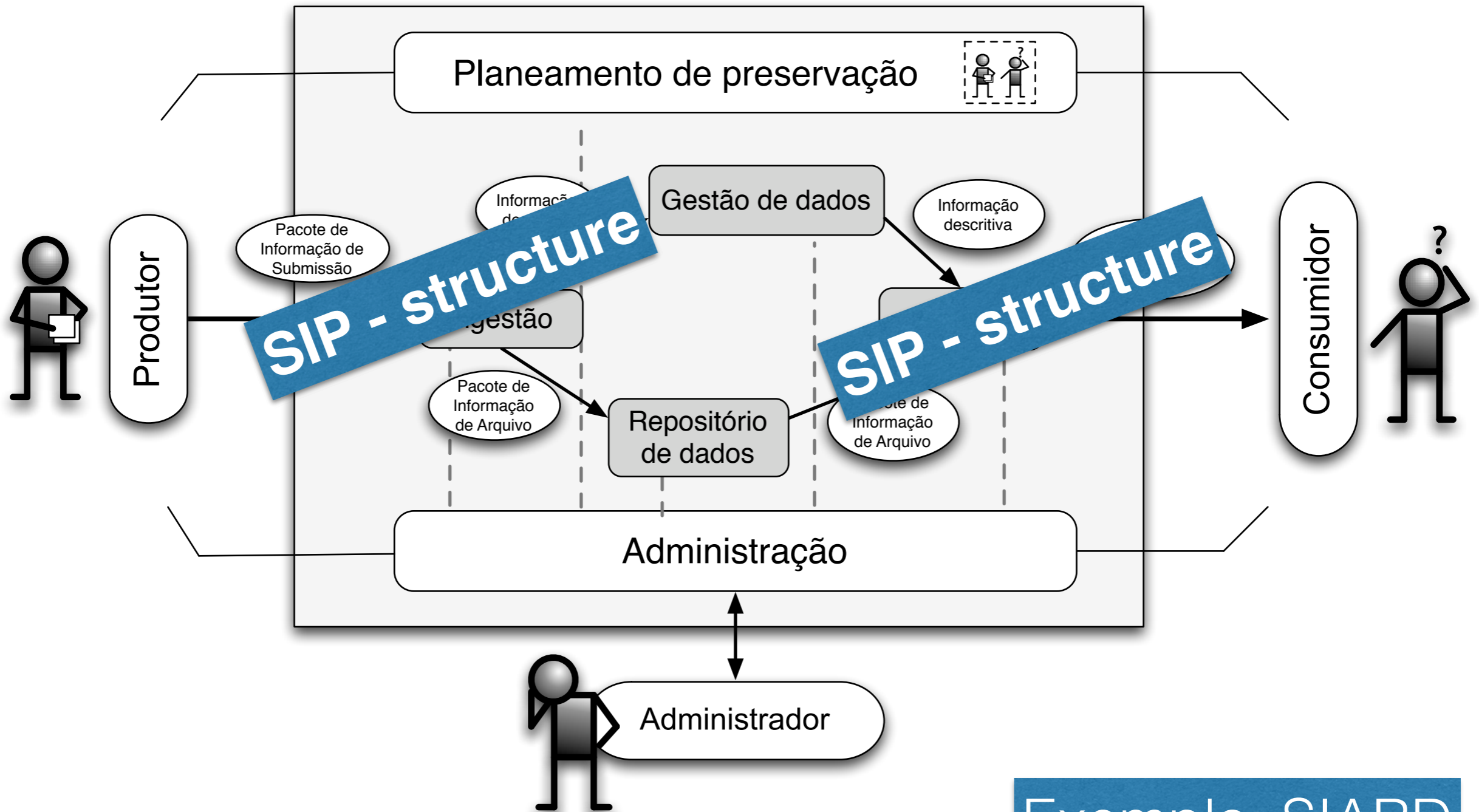
Os outros:

Open Archival Information System - OAIS ISO 14721:2003



Os outros:

Open Archival Information System - OAIS ISO 14721:2003



Exemplo: SIARD

Ingest with RODA

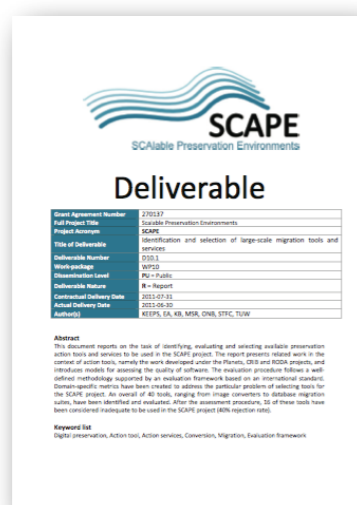
The present and the future of repository ingest

<http://bit.ly/scape-tools>

<http://www.scape-project.eu>

<http://redmine.keep.pt>

<http://roda.keep.pt>



Questões?



José Carlos Ramalho

Consultor / Investigador

jcr@keep.pt / jcr@di.uminho.pt

KEEPSOLUTIONS

University of Minho SPIN-OFF

ARQUIVOS | BIBLIOTECAS | MUSEUS

www.keep.pt