

# RODA and CRiB: A Service-Oriented Digital Repository

---

José Carlos Ramalho  
[jcr@di.uminho.pt](mailto:jcr@di.uminho.pt)

Francisco Barbedo  
[frbarbedo@iantt.pt](mailto:frbarbedo@iantt.pt)

Miguel Ferreira  
[mferreira@dsi.uminho.pt](mailto:mferreira@dsi.uminho.pt)

Cecília Henriques  
[chenriques@iantt.pt](mailto:chenriques@iantt.pt)

Rui Castro  
[Rcastro@iantt.pt](mailto:Rcastro@iantt.pt)

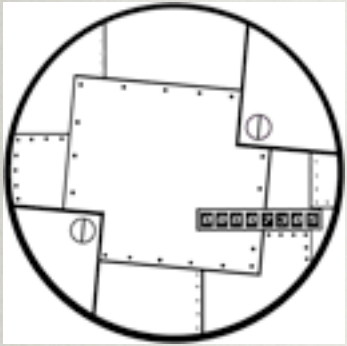
Luis Corujo  
[lcorujo@iantt.pt](mailto:lcorujo@iantt.pt)

Luis Faria  
[lfaria@iantt.pt](mailto:lfaria@iantt.pt)



# CONTEXT

---



## RODA (2006-2009)

- Metadata management (EAD based)
- Digital object management (...)
- Digital preservation protocols and policies
- National project with public funds (Portuguese National Archives)



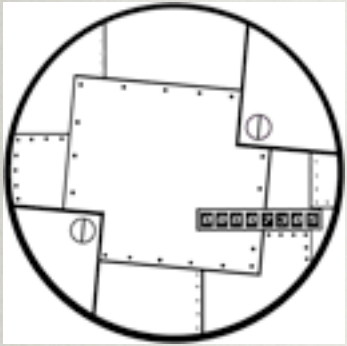
## CRiB: Preservation Services for Digital Repositories (2005-2008)

- Distributed Migration Service
- Migration Adviser Service
- Phd thesis at U.Minho (Miguel Ferreira)



# CONTEXT

---



## RODA (2006-2009)

- Metadata management (EAD based)
- Digital object management (...)
- Digital preservation protocols and policies
- National project with public funds (Portuguese National Archives)



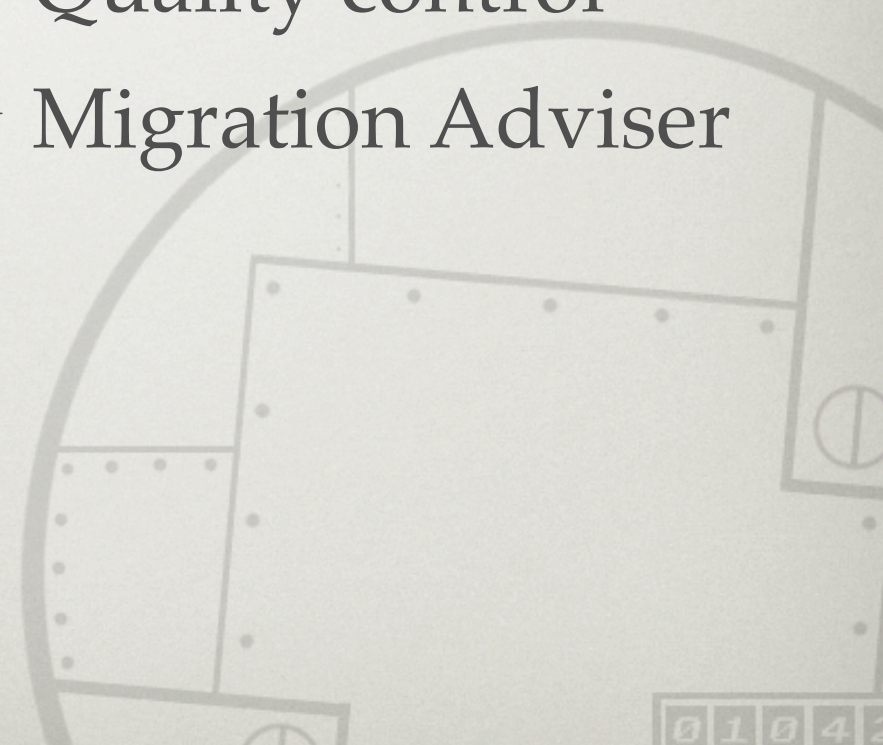
## CRiB: Preservation Services for Digital Repositories (2005-2008)

- Distributed Migration Service
- Migration Adviser Service
- Phd thesis at U.Minho (Miguel Ferreira)



# PREMISSES

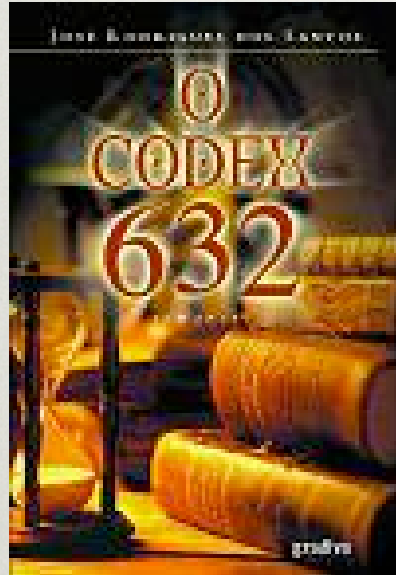
---

- RODA:
    - \* OAIS as reference model
    - \* EAD as descriptive metadata
    - \* Open Source
    - \* SOA
    - \* Preservation Policy: Migration
    - \* Authenticity
  - CRiB:
    - \* Distributed Migration
    - \* SOA
    - \* Quality control
    - \* Migration Adviser
- 



# AUTHENTICITY

---



“O Codex 632” by José Rodrigues dos Santos

Subject: Who really was Cristophoros Colombus?

Was he italian? Spanish? Or a portuguese belonging to a jewish family?

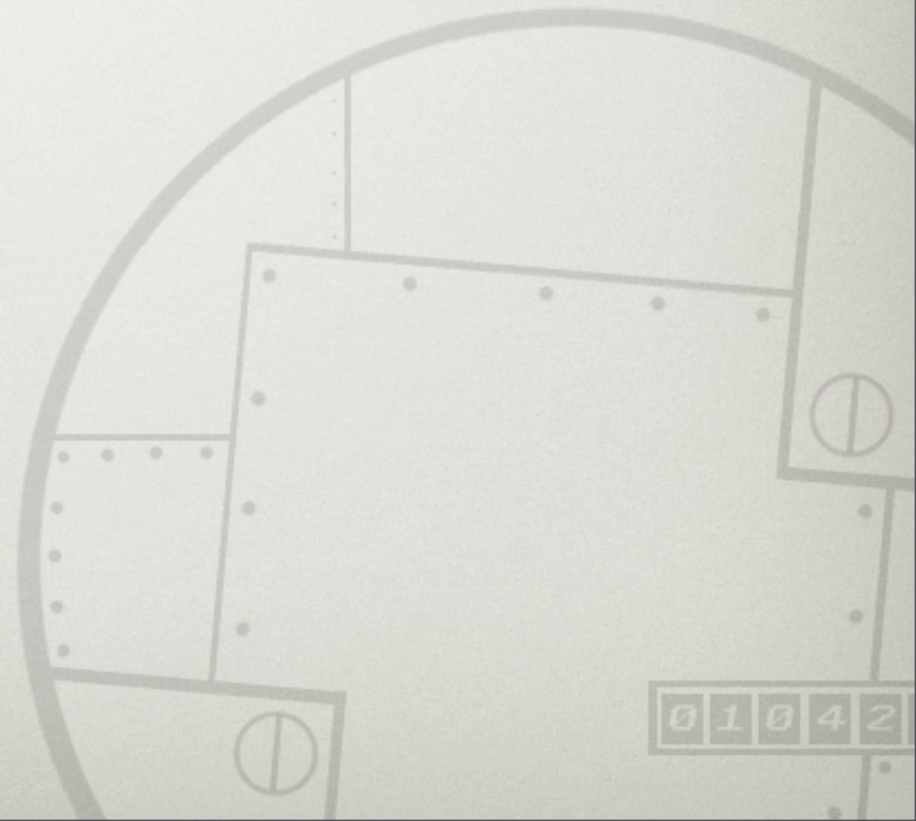


# AUTHENTICITY

---

We must trust our sources: in ancient History there are no direct speech or evidence.

EX: the bible





# AUTHENTICITY

---

We must trust our sources: in ancient History there are no direct speech or evidence.

EX: the bible

How do we become trustful?



# AUTHENTICITY

---

We must trust our sources: in ancient History there are no direct speech or evidence.

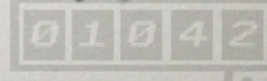
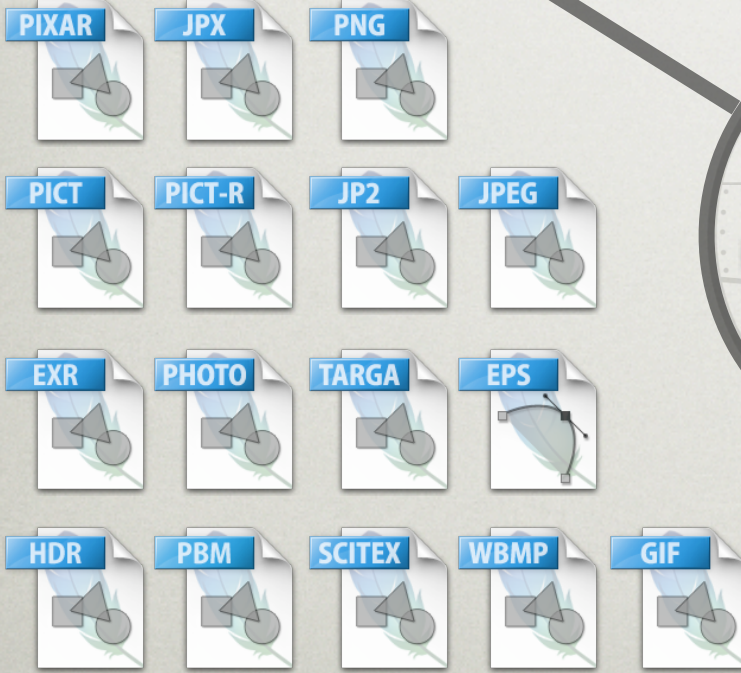
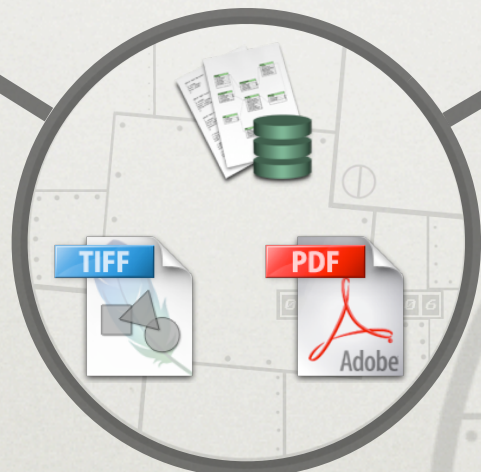
EX: the bible

How do we become trustful?

- Reputation
- Documenting every action taken upon DOs

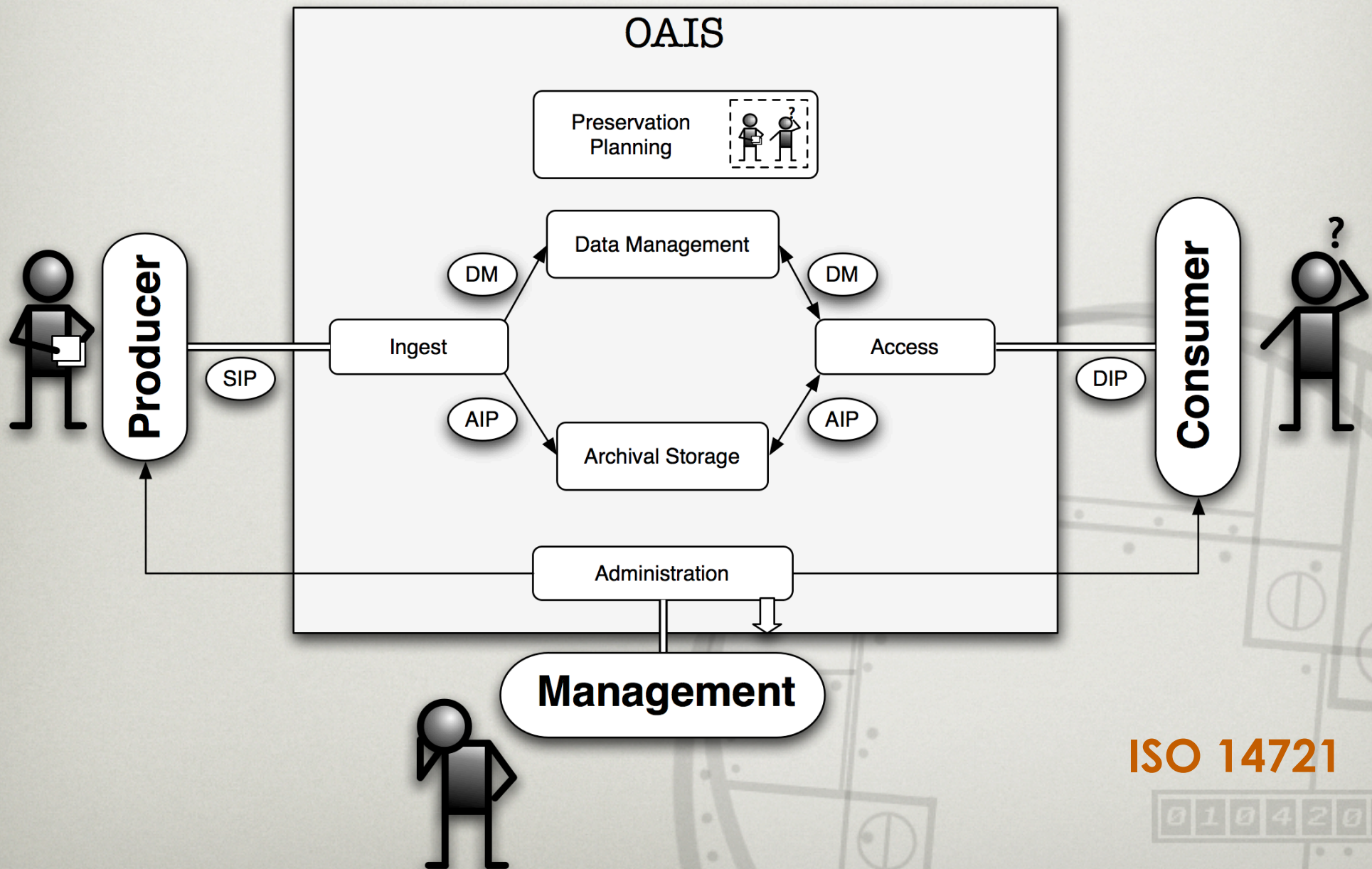


# DO CLASSES: NORMALIZATION





# OPEN ARCHIVAL INFORMATION SYSTEM



ISO 14721

01042005



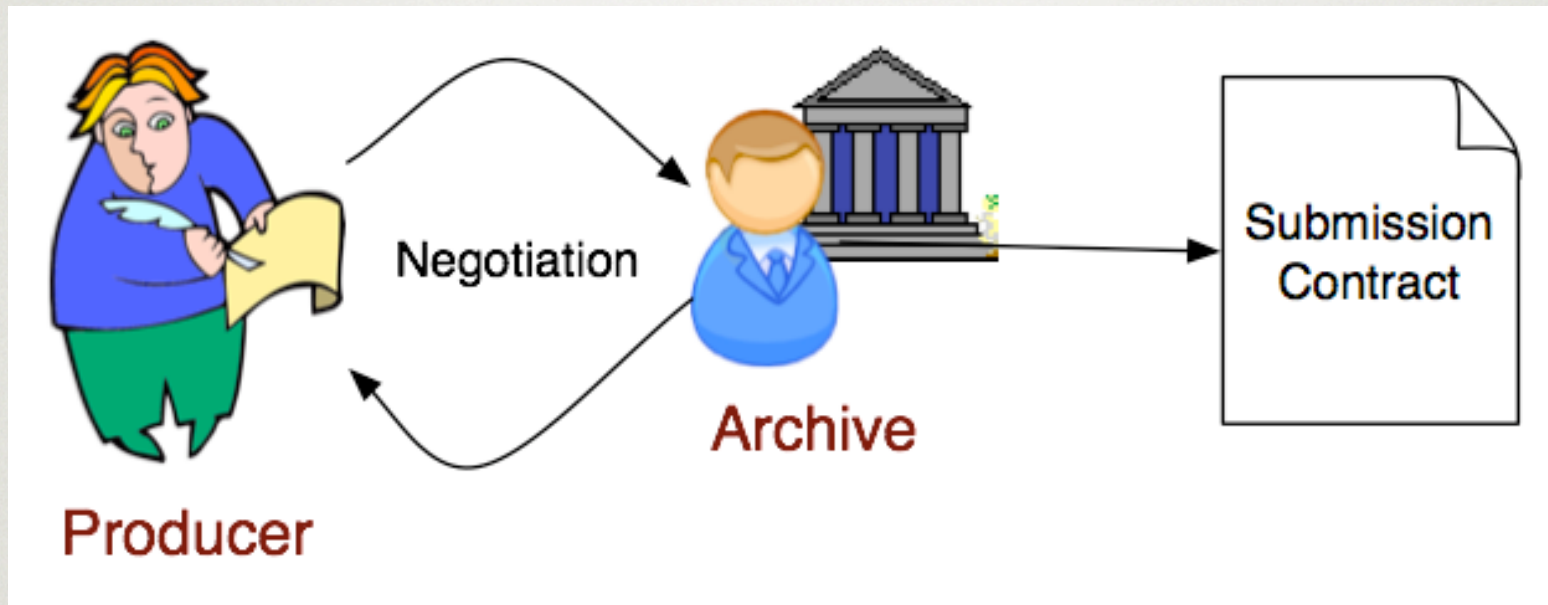
# OAIS (INFORMATION PACKAGES)

---

- Submission Information Package (SIP)
  - \* **Digital Object**
  - \* **Metadata created by producer**
    - ▶ **too open...**
- Archival Information Package (AIP)
  - \* **Digital Object to be stored**
  - \* **Metadata:** enough to ensure DO's preservation and access
    - ▶ **model defined by PREMIS**
- Dissemination Information Package (DIP)
  - DO transformed into the **format** that will be **delivered** to the **consumer**
  - **Metadata**

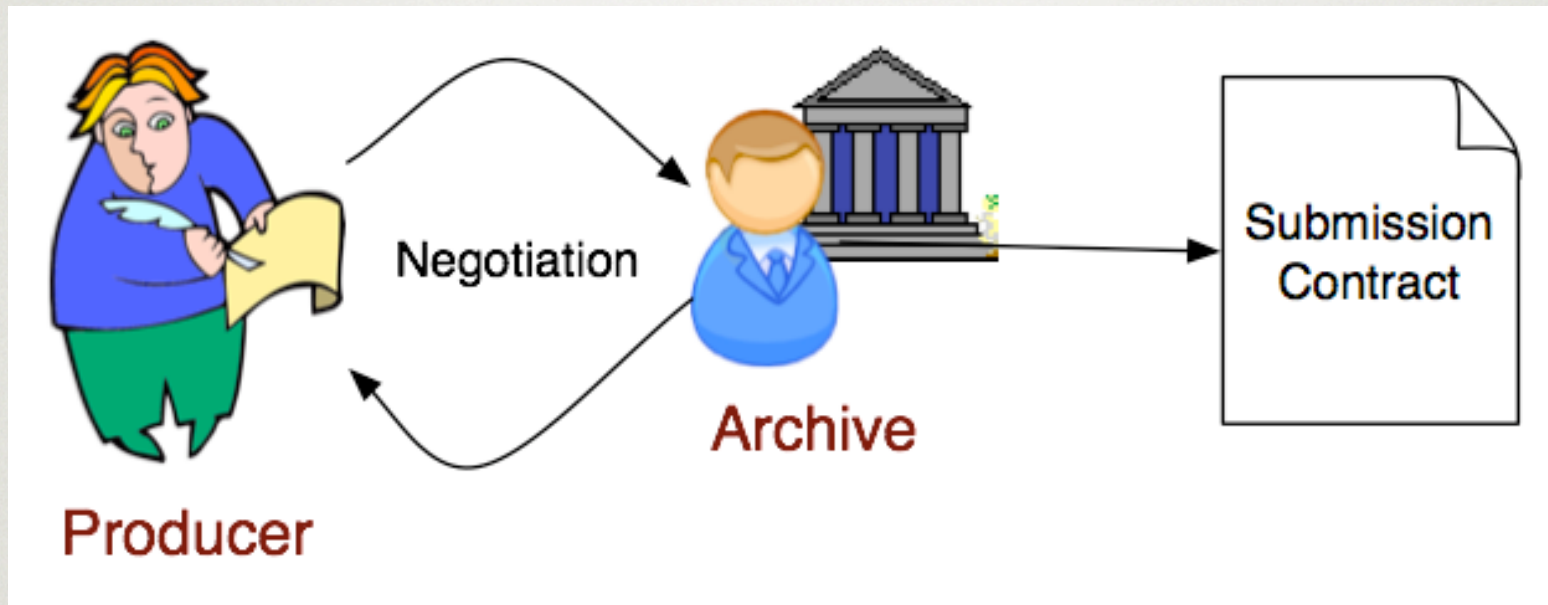


# INGESTION





# INGESTION



## Submission Contract

- SIP specification
- Ingestion workflow specification

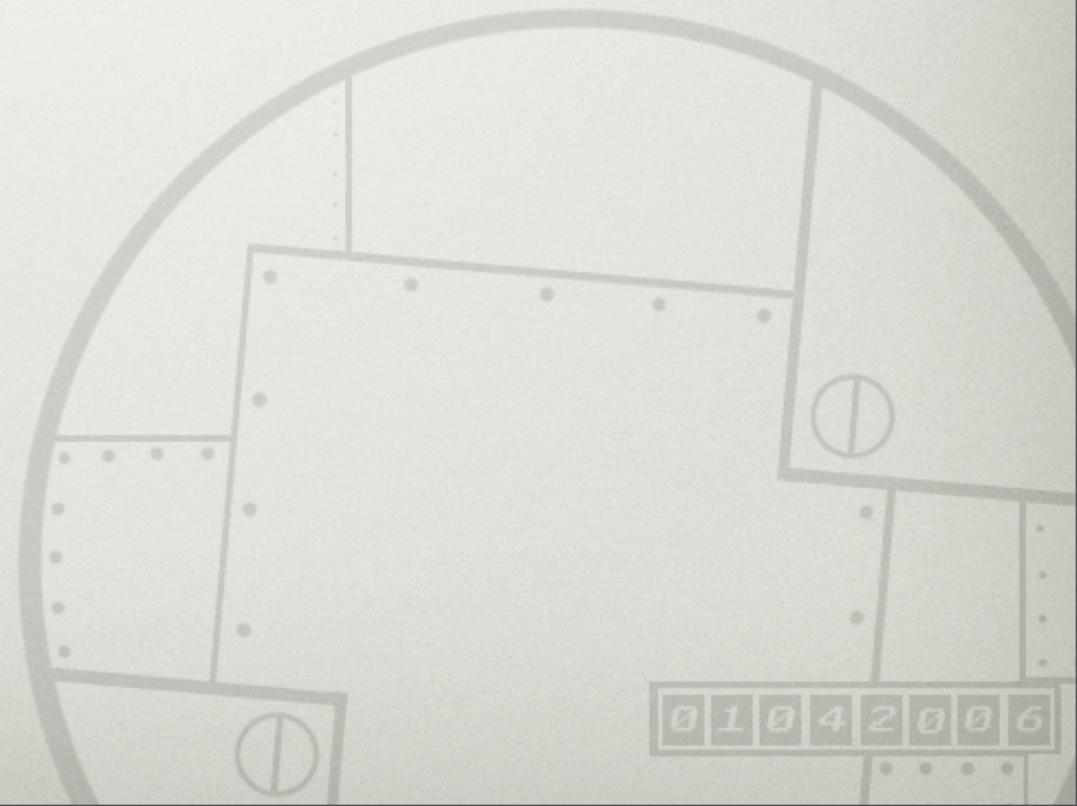


# SIP STRUCTURE (EXAMPLE)

---



one still image



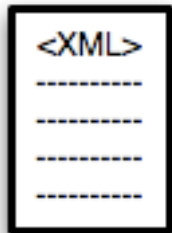


# SIP STRUCTURE (EXAMPLE)

---



one still image



creation

properties:

- date
- hardware
- ...

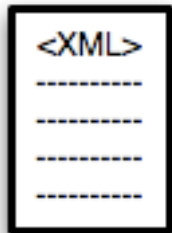


# SIP STRUCTURE (EXAMPLE)

---



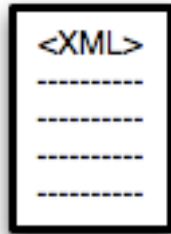
one still image



creation

properties:

- date
- hardware
- ...



Technical Metadata:

- color
- dimensions
- ...

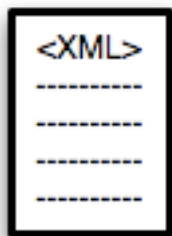


# SIP STRUCTURE (EXAMPLE)

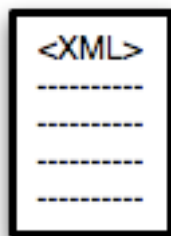
---



one still image

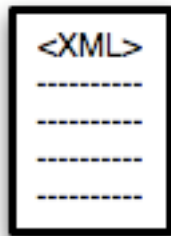


creation  
properties:  
- date  
- hardware  
- ...



Technical Metadata:

- color
- dimensions
- ...

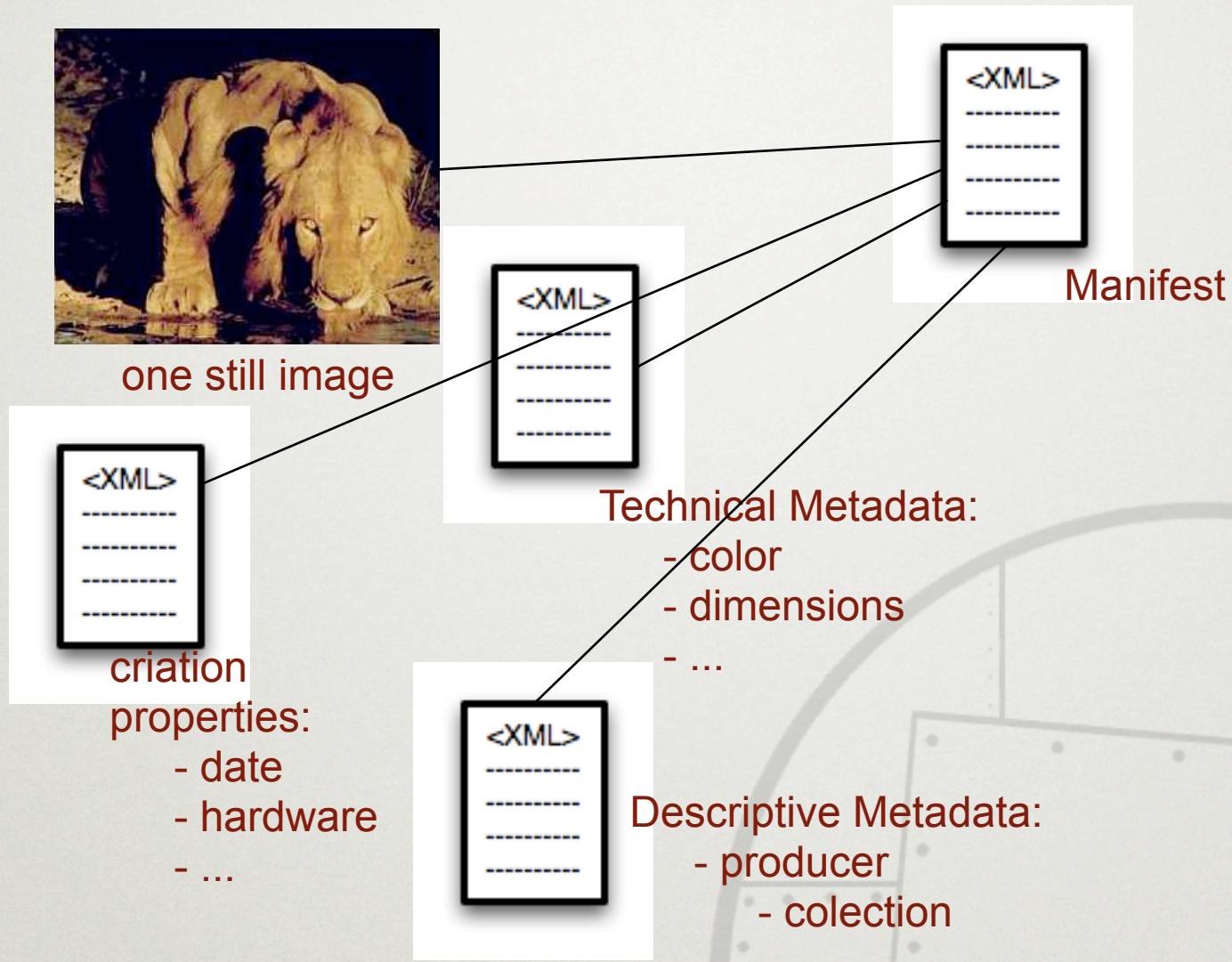


Descriptive Metadata:

- producer
- collection
- ...

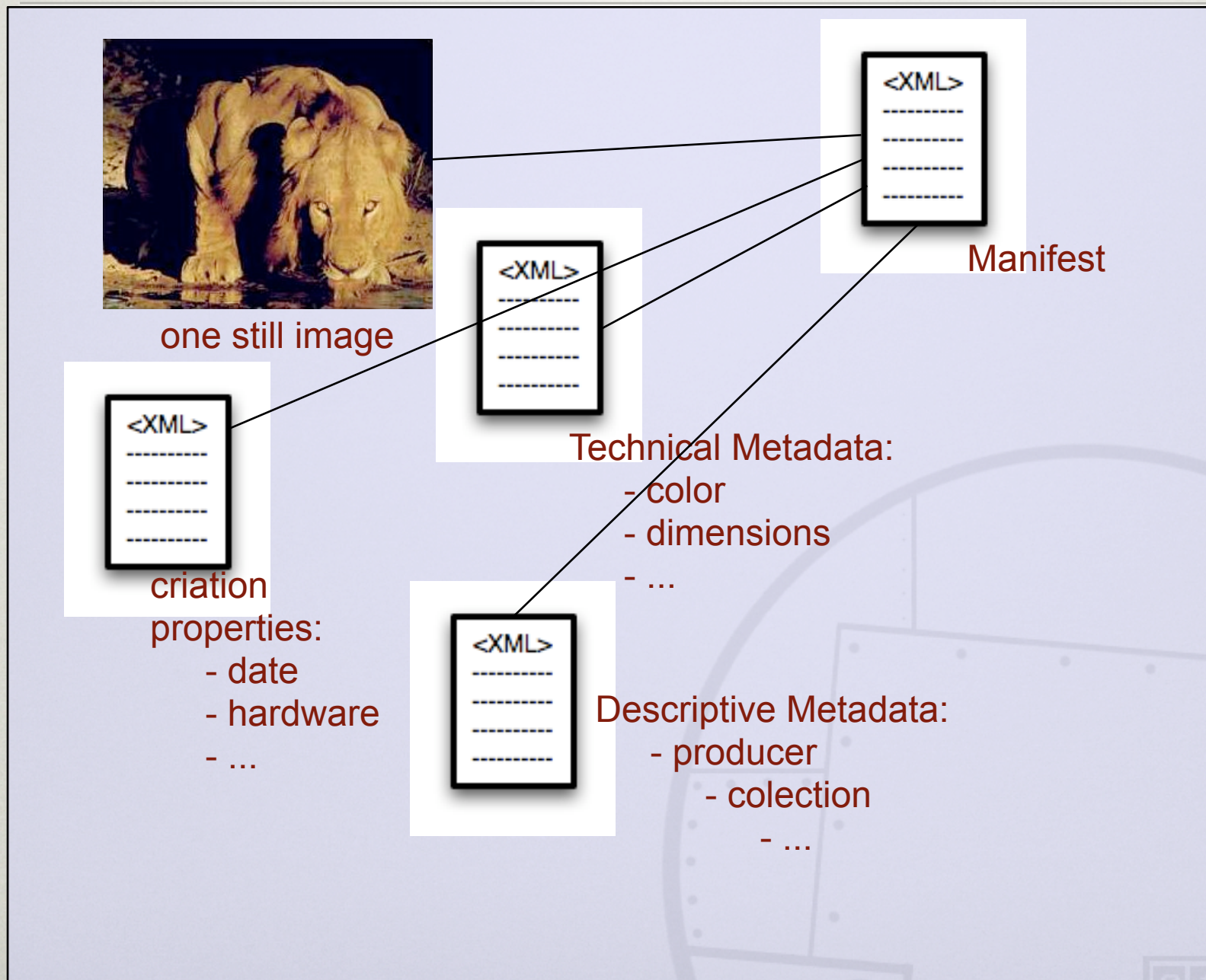


# SIP STRUCTURE (EXAMPLE)





# SIP STRUCTURE (EXAMPLE)

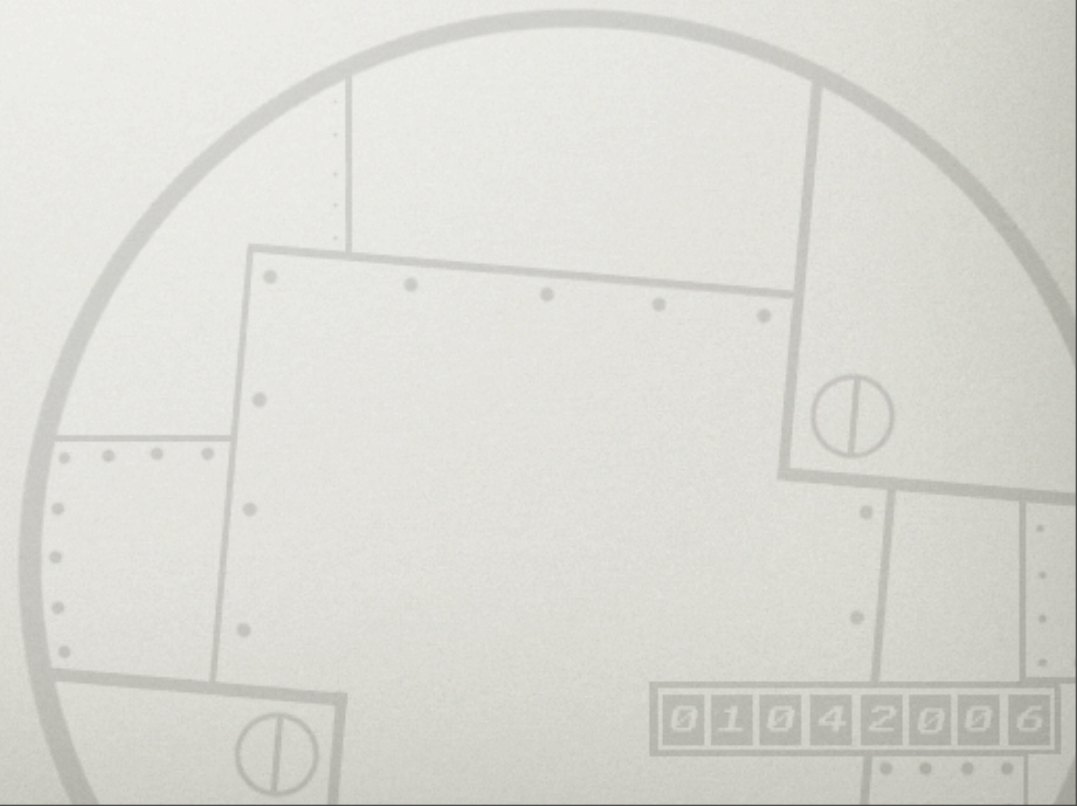


**Compressed File**



# SIP STRUCTURE (+COMPLEX)

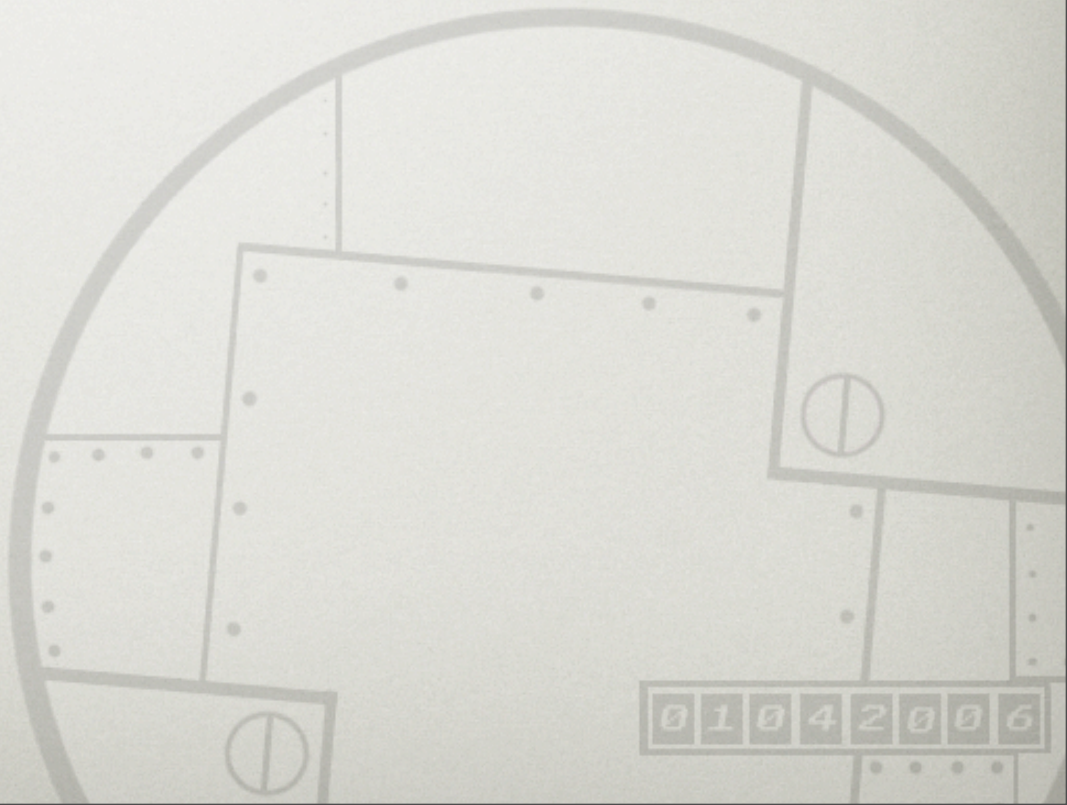
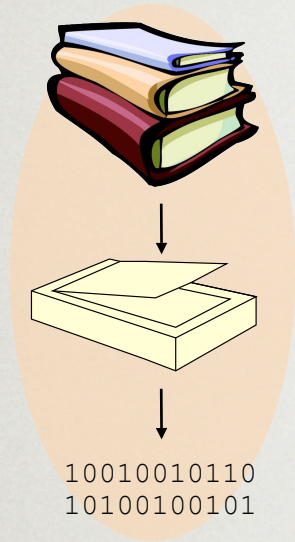
---





# SIP STRUCTURE (+COMPLEX)

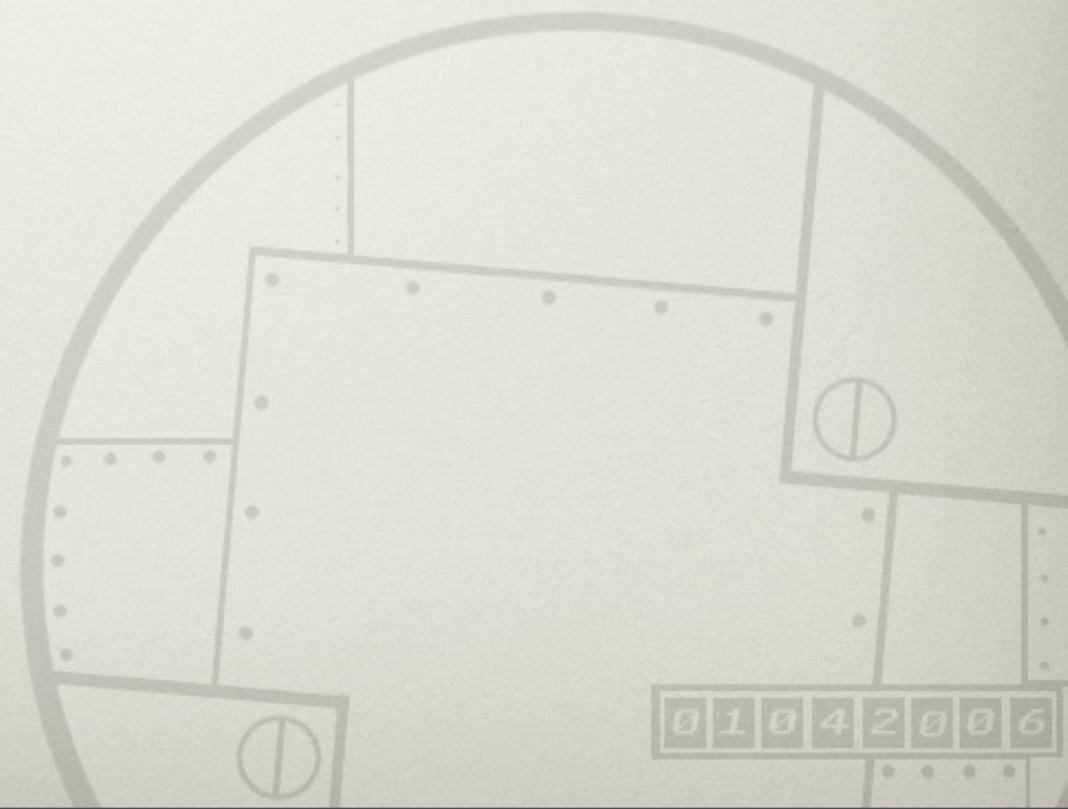
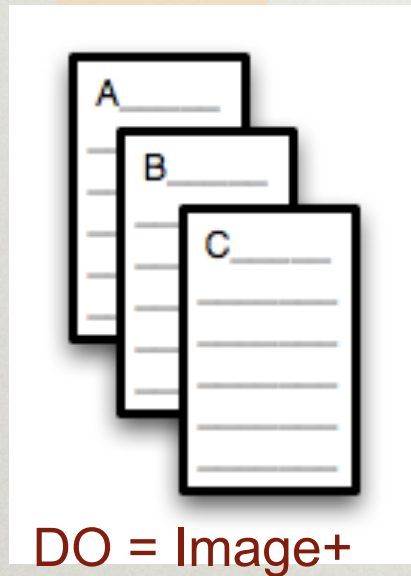
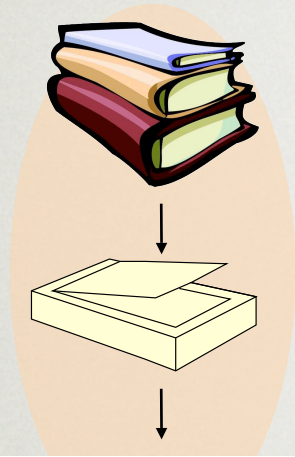
---





# SIP STRUCTURE (+COMPLEX)

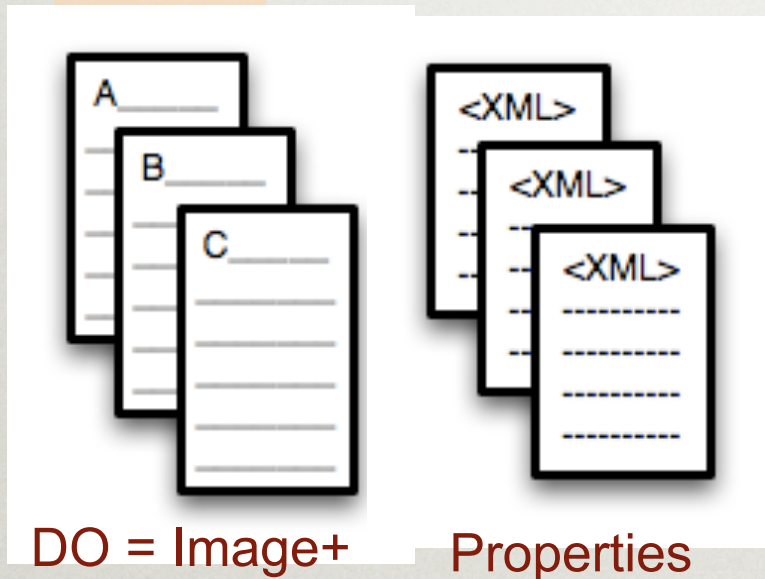
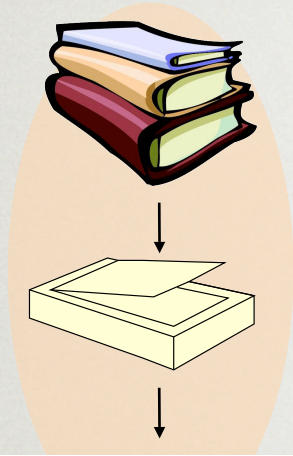
---





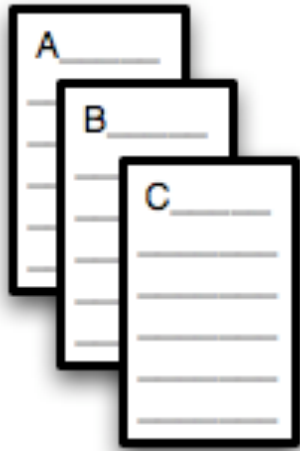
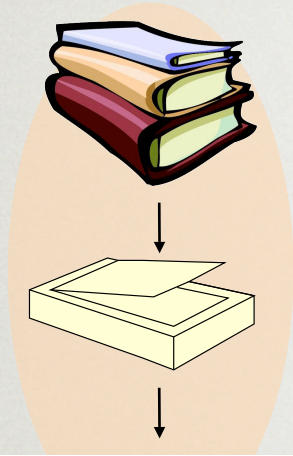
# SIP STRUCTURE (+COMPLEX)

---





# SIP STRUCTURE (+COMPLEX)



DO = Image+



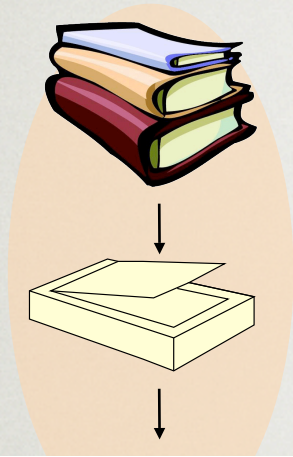
Properties



Technical Metadata



# SIP STRUCTURE (+COMPLEX)



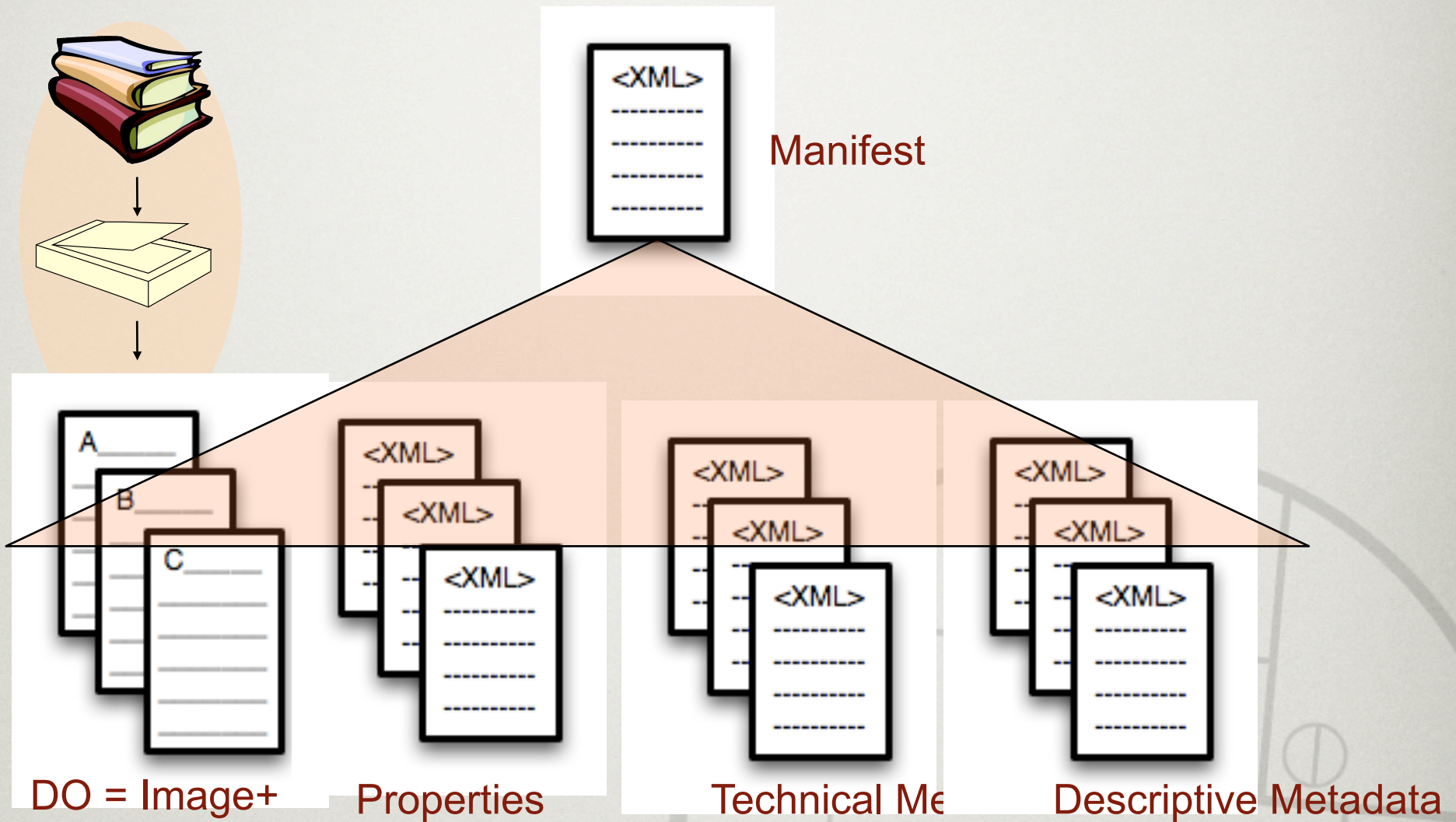
DO = Image+ Properties

Technical Me

Descriptive Metadata

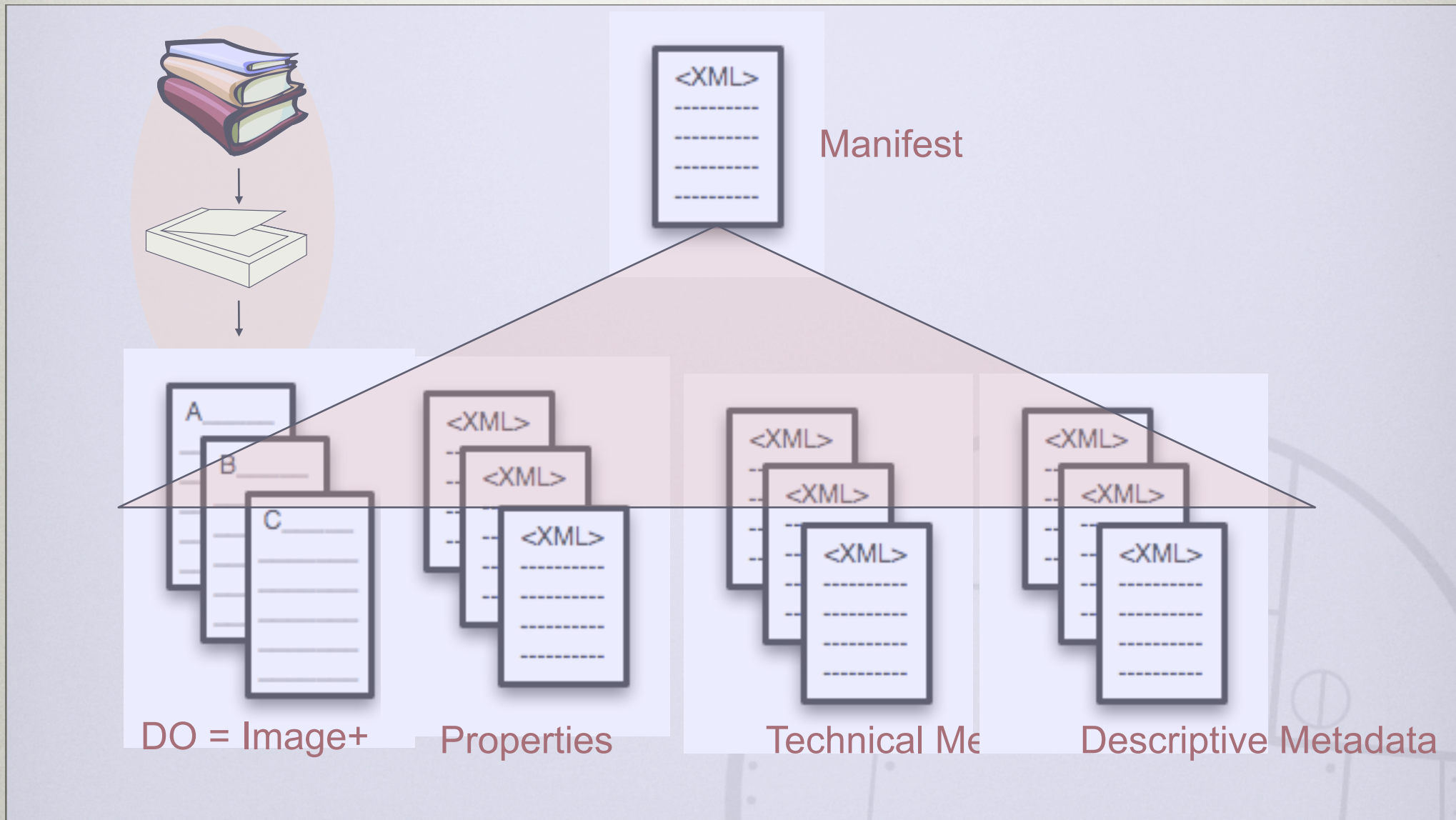


# SIP STRUCTURE (+COMPLEX)





# SIP STRUCTURE (+COMPLEX)



**Compressed File**



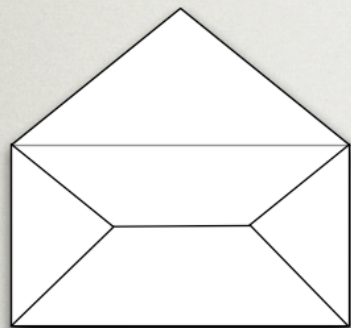
# INGESTION WORKFLOW

---

**SIP**



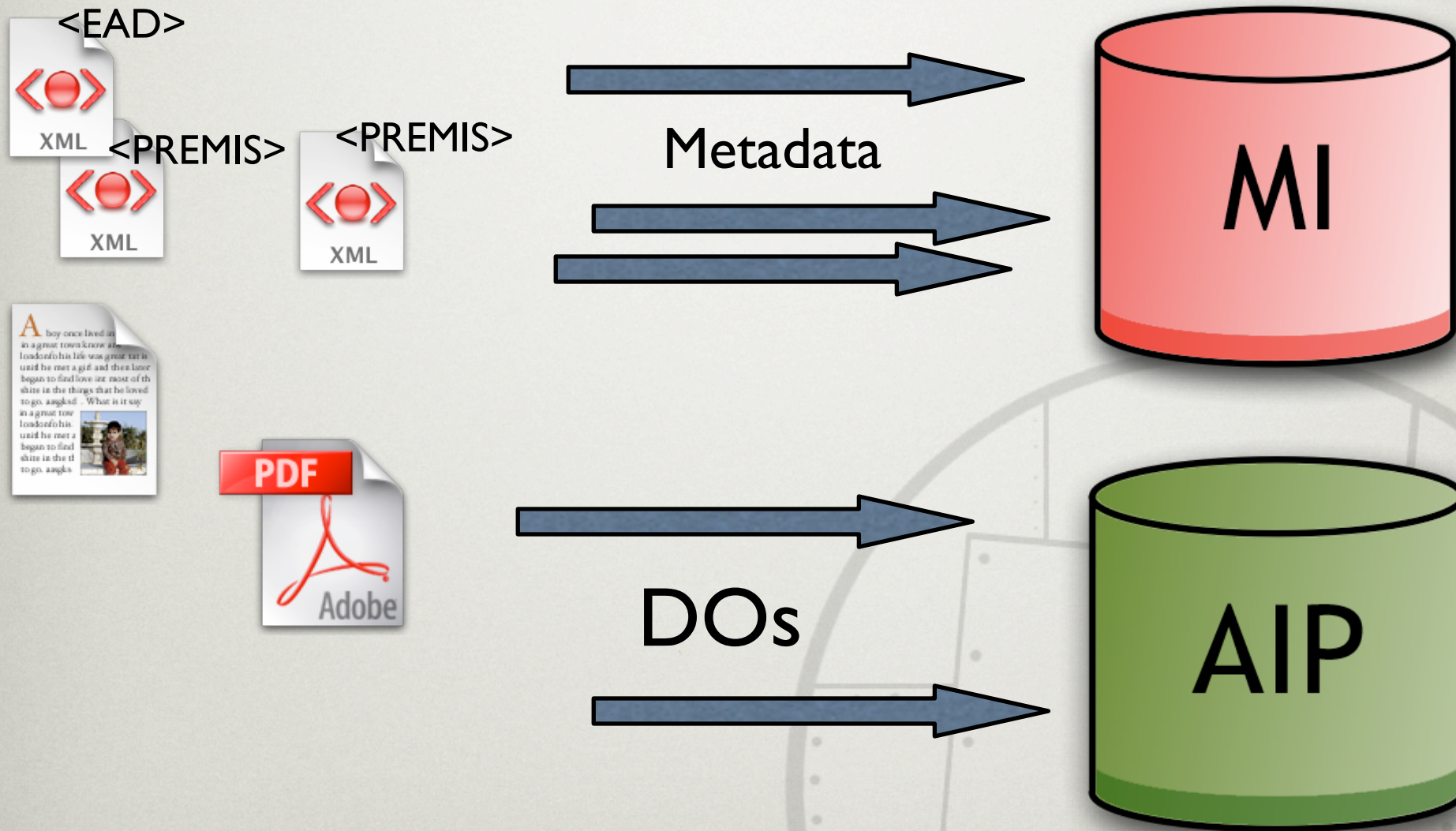
**AIP**



- integrity check
- virus check
- generation of preservation metadata (PREMIS)
- conversion to a normalized format
- generation of technical metadata
- generation of preservation metadata (PREMIS)

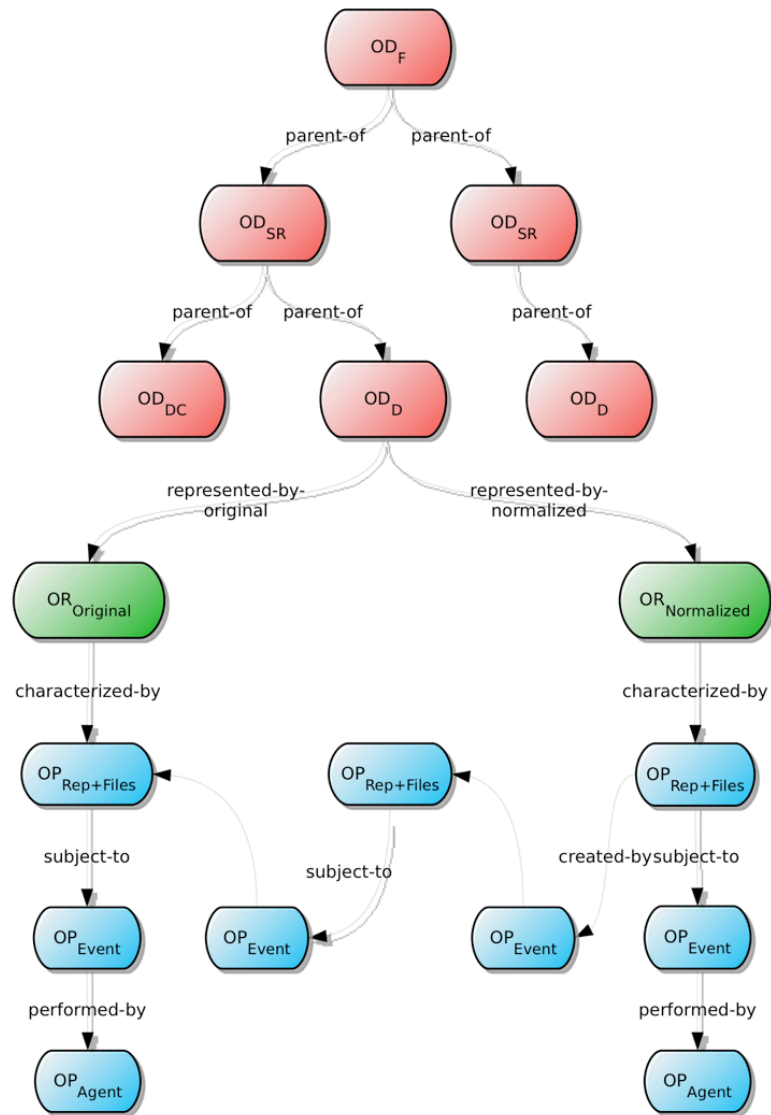


# AIP STORAGE





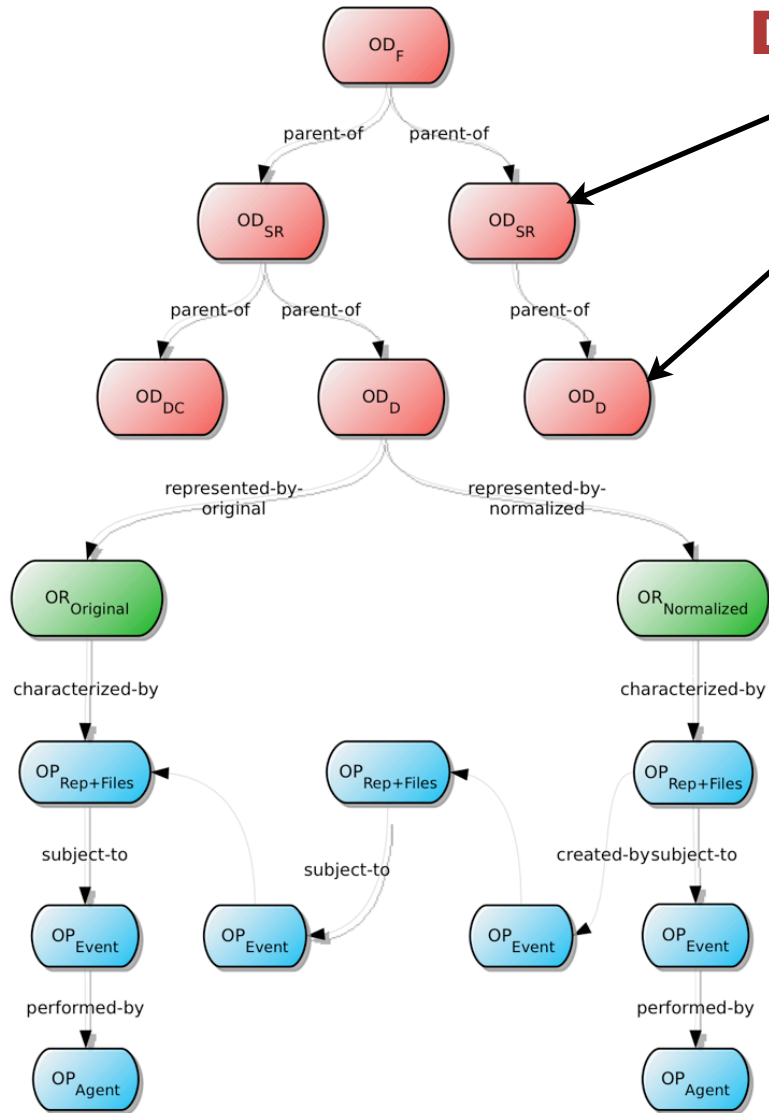
# RODA DATA MODEL





# RODA DATA MODEL

## Description Objects

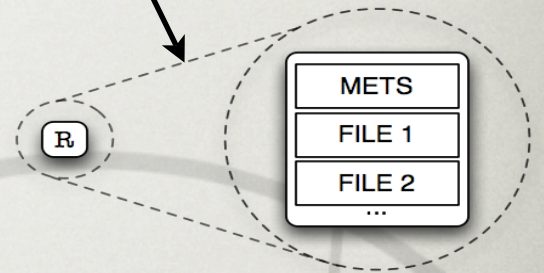
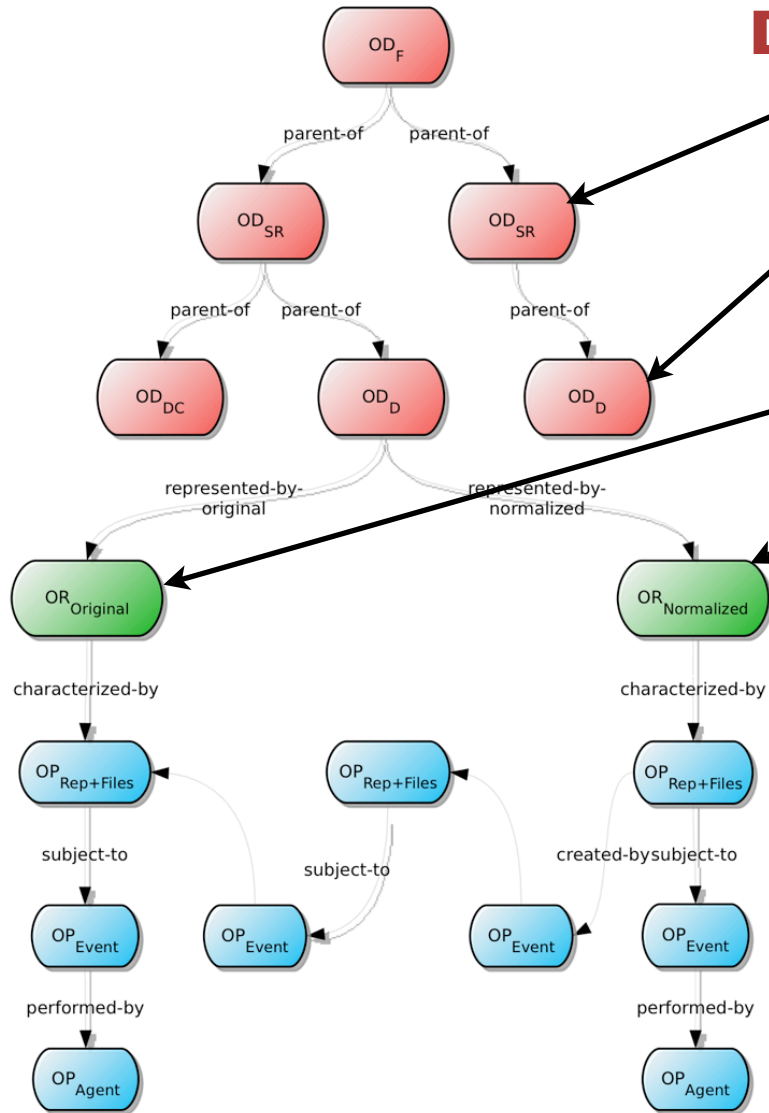




# RODA DATA MODEL

**Description Objects**

**Representation Objects**



01042006

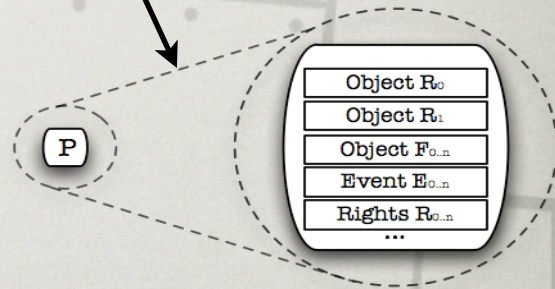
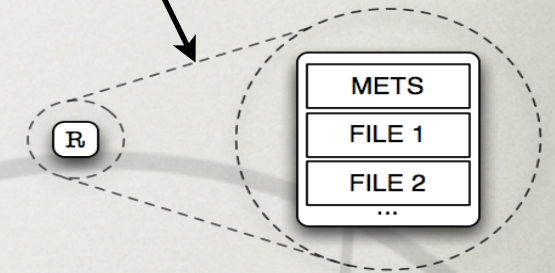
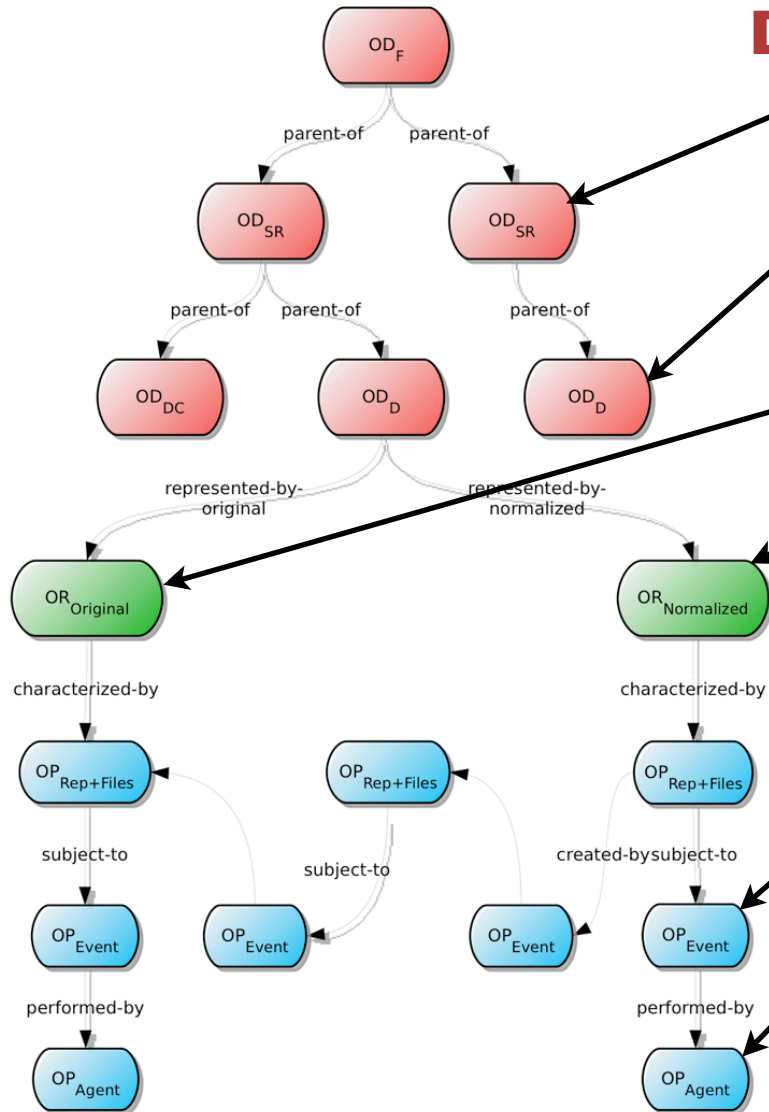


# RODA DATA MODEL

## Description Objects

## Representation Objects

## Preservation Objects

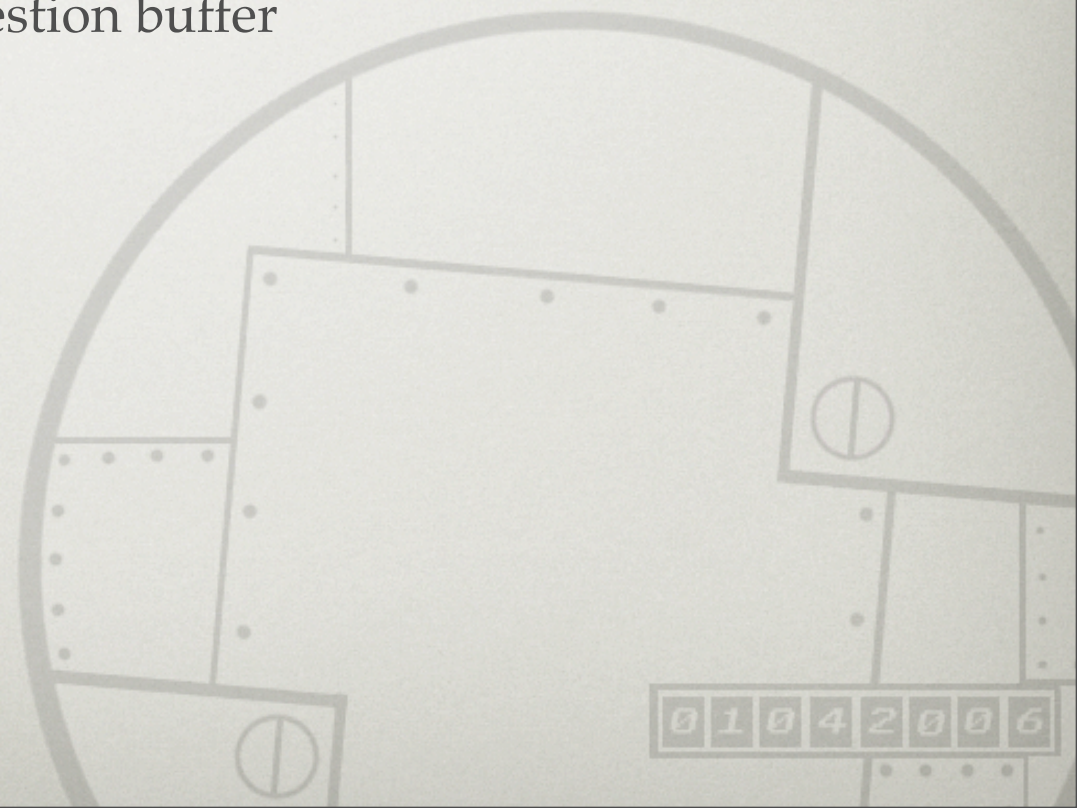




# IMPLEMENTATION REQUISITES

---

- Graphical Interface for Ingestion process
- Producer registry
- SIP production tool
- Ingestion feedback
- Partial Ingestion
- “Quarantine” zone: cache, ingestion buffer
- SIP validation
- Error reporting
- Persistent identifiers
- PREMIS event generation
- DIP digital signature
- ...





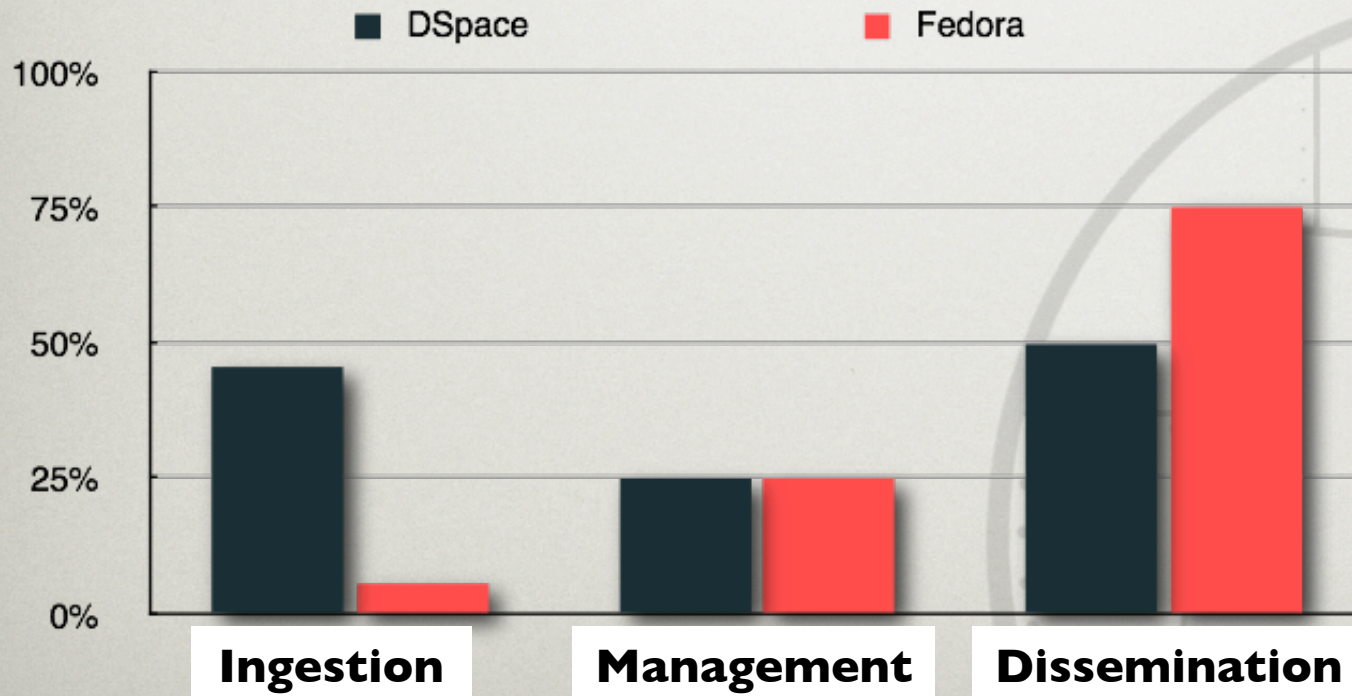
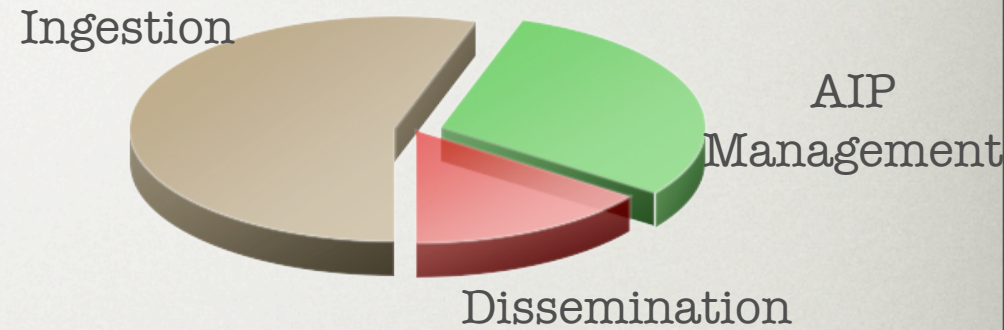
# DEVELOPMENT FRAMEWORK

---





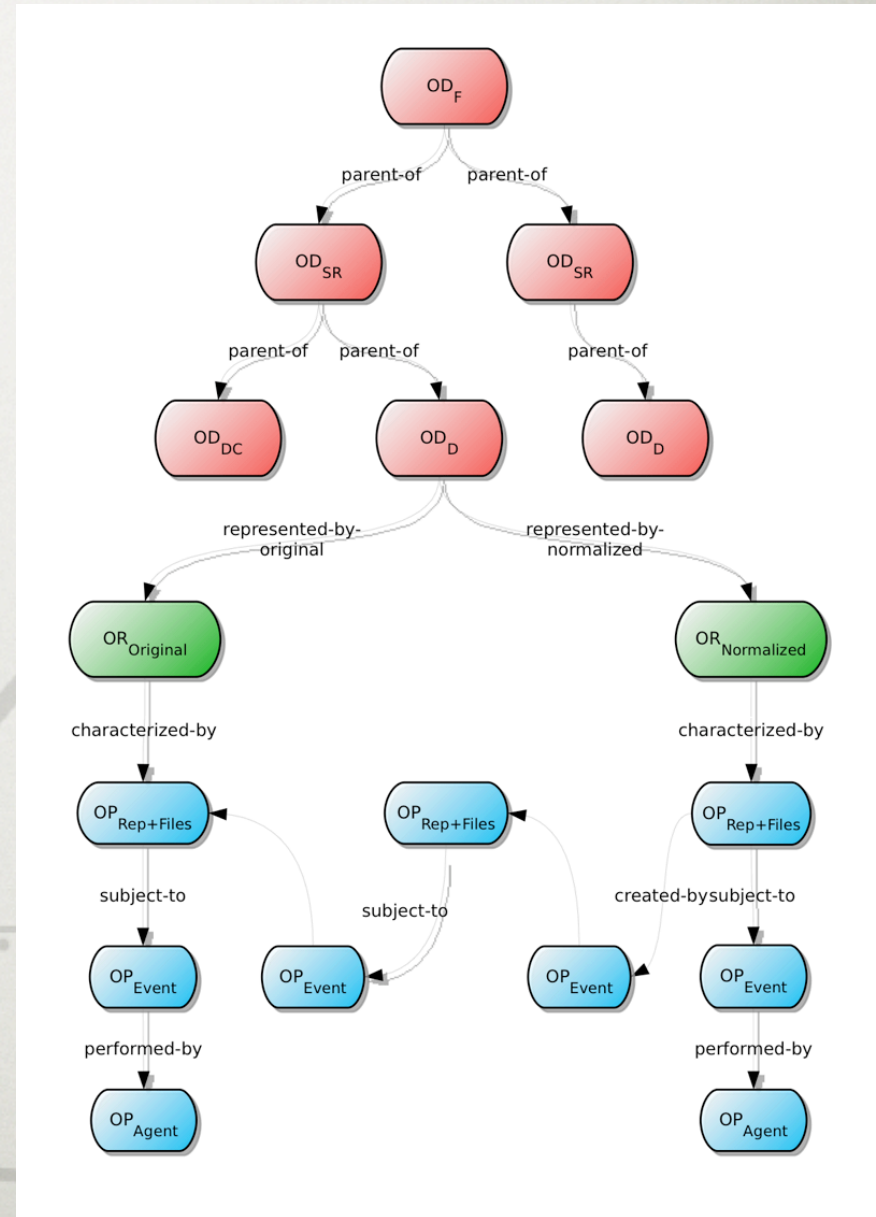
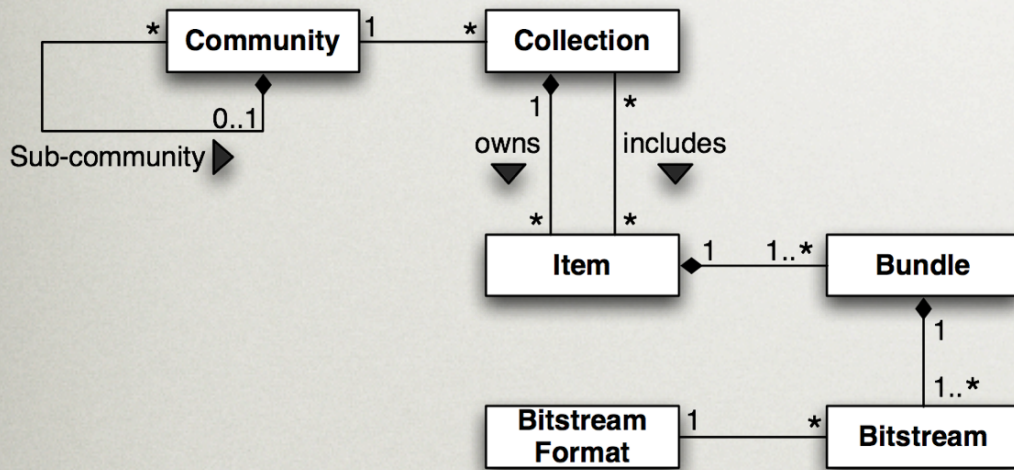
# REQUISITES BASED COMPARAISON





# MATCHING DATA MODELS

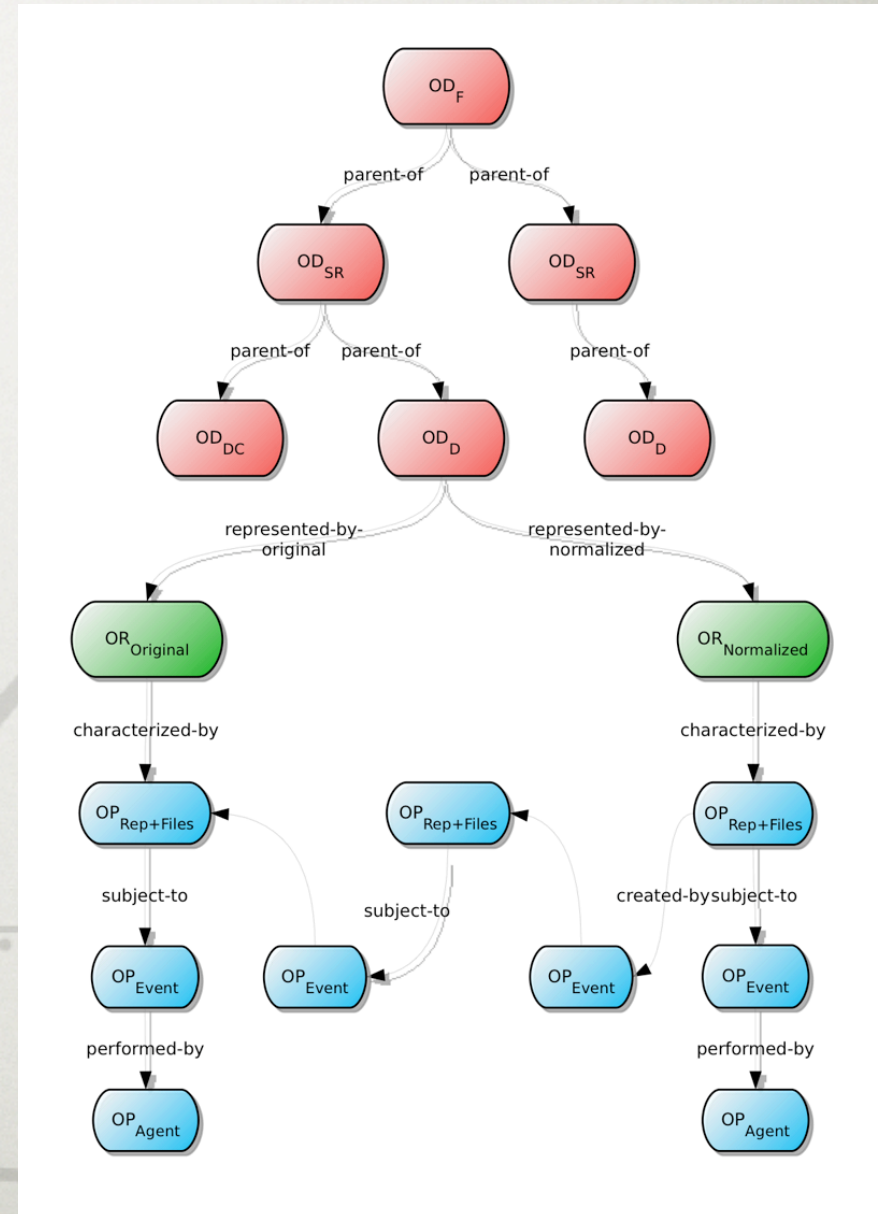
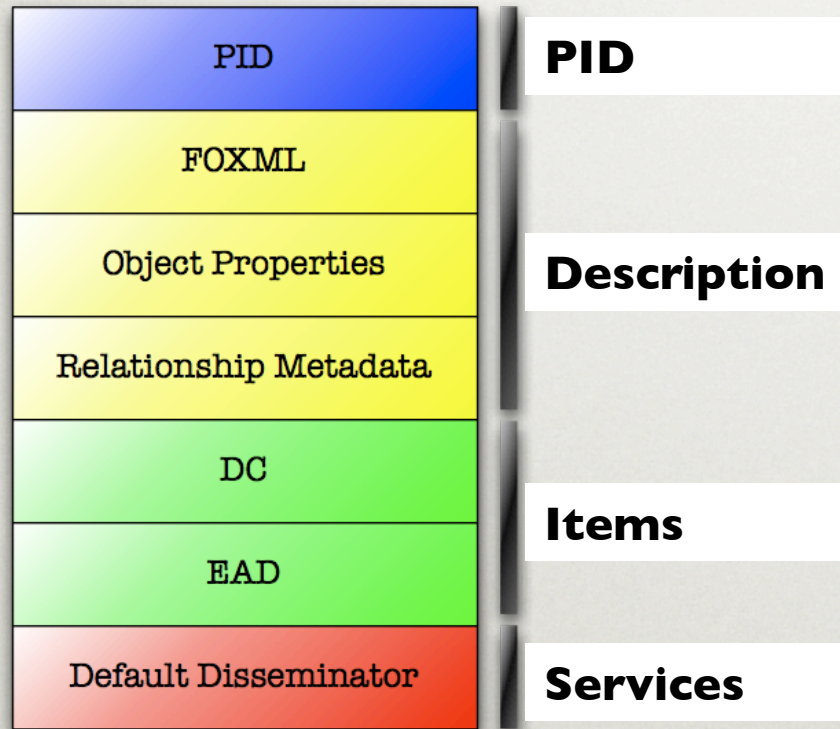
DSpace





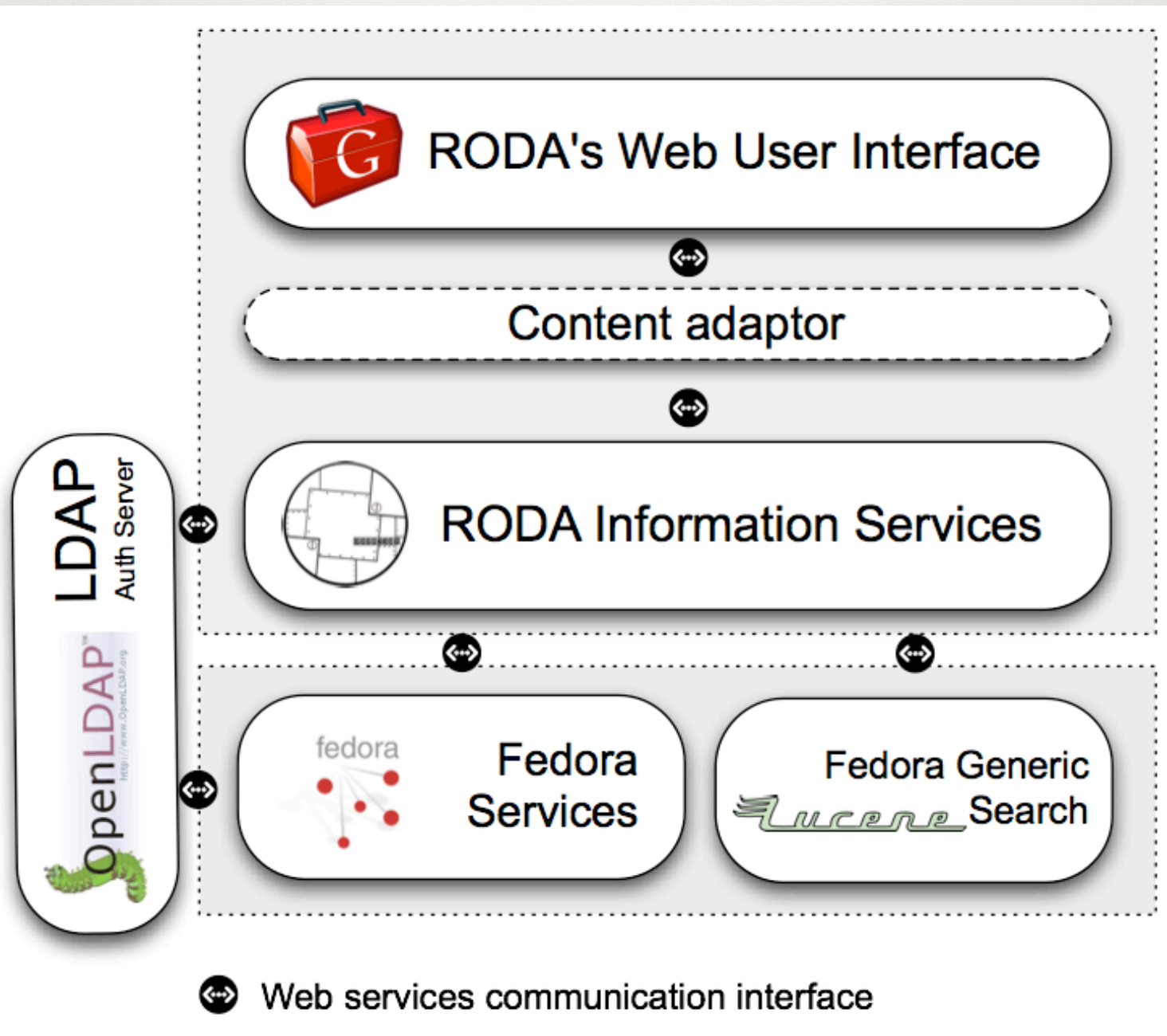
# MATCHING DATA MODELS

## Fedora



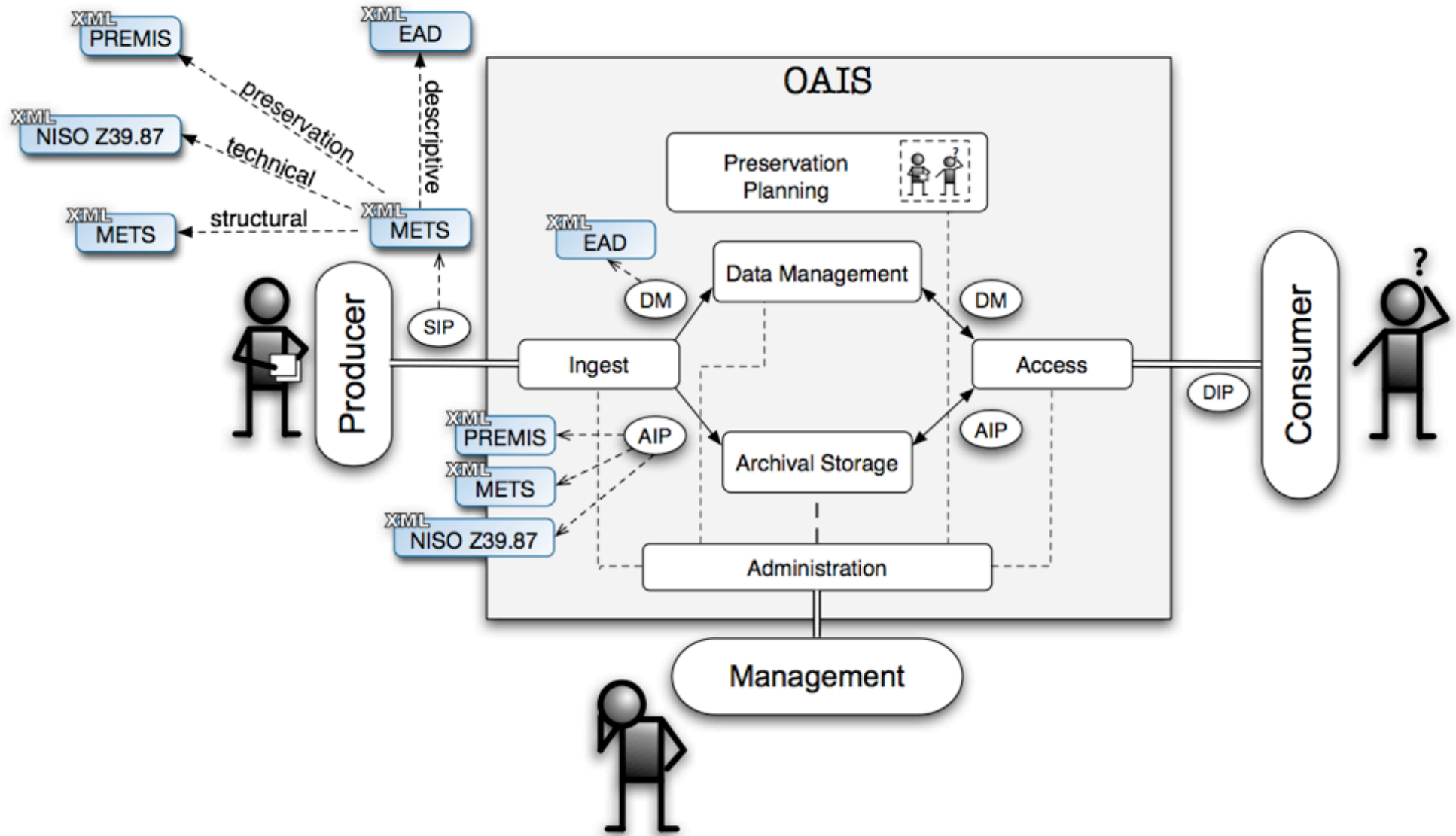


# ARCHITECTURE





# RODA SCHEMAS





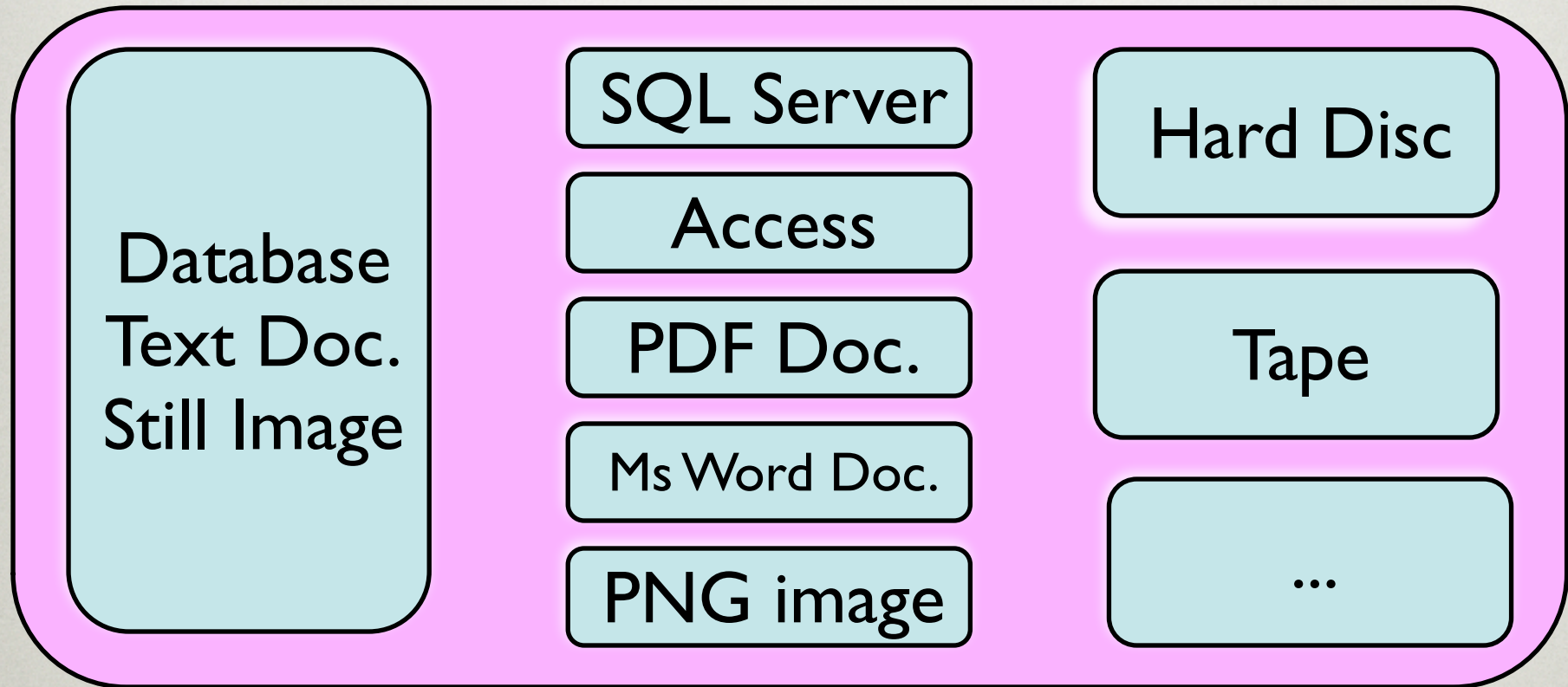
# DO Anatomy

---

Conceptual  
level

Logical  
level

Physical  
level



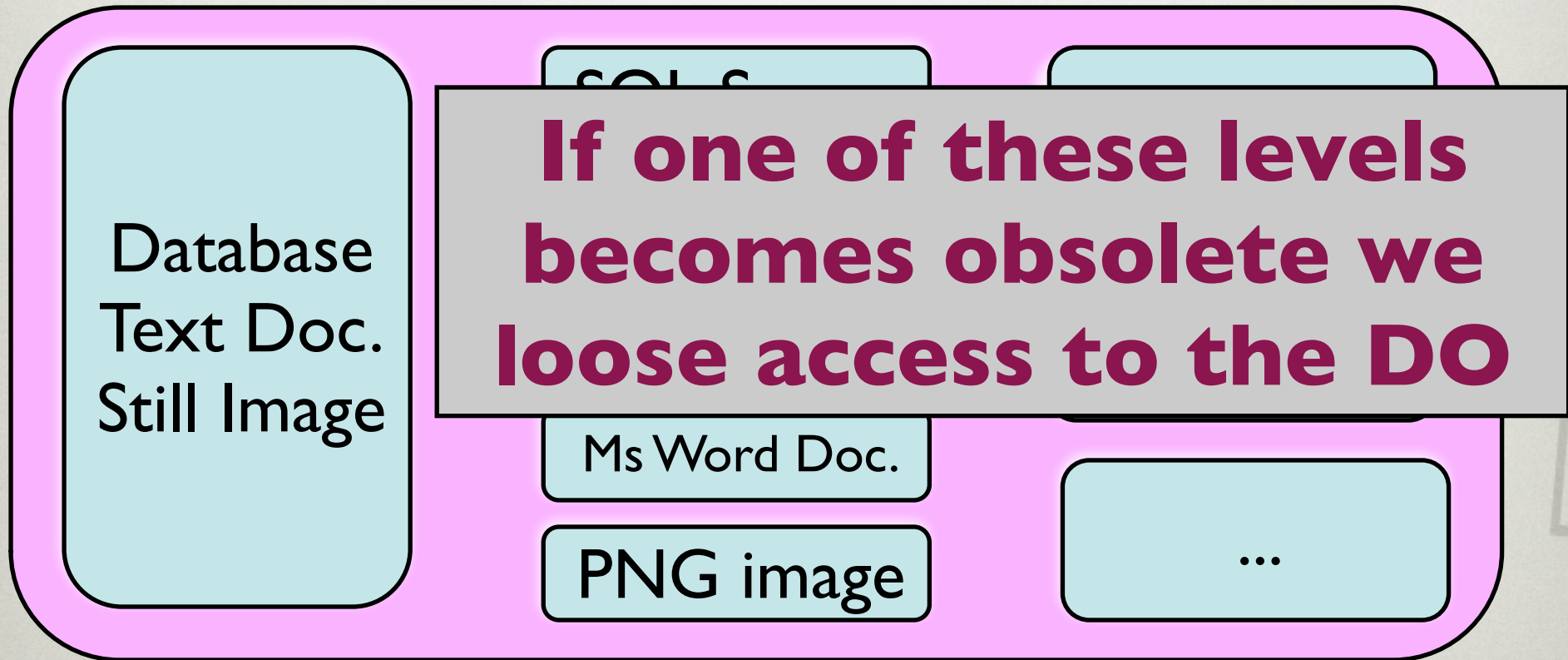


# DO Anatomy

Conceptual  
level

Logical  
level

Physical  
level





# DO Preservation Strategies

---

- Focusing the **physical/logical object**
  - Centered in preserving information in her **logical format** or / and **physical support**
  - Uses original technology associated to these objects to ensure the access to them
  - **Technology preservation**
- Focusing the **conceptual object**
  - Centered in **preserving the object core properties in a way that is independent from hardware and software**
  - **Conceptual object preservation**



# Conceptual object preservation

---

**Migration:** periodic DO transfer from one hw/sw configuration into an updated one (centered in preserving significant properties other than preserving the original bit stream).

## Advantages

- DO are disseminated in formats known to users
- No need to preserve the original hw/sw platform
- Most used strategy and the only that has worked so far

## Disadvantages

- Possible loss of information during conversion
- Continued maintenance is needed
- **In the longterm perspective costs are high**



# Conceptual object preservation

---

**Migration:** periodic DO transfer from one hw/sw configuration into an updated one (centered in preserving significant properties other than preserving the original bit stream).

Advantages

- DO are
- No need to preserve the original hw/sw platform
- Most used strategy and the only that has worked so far

**What are the significant properties?**

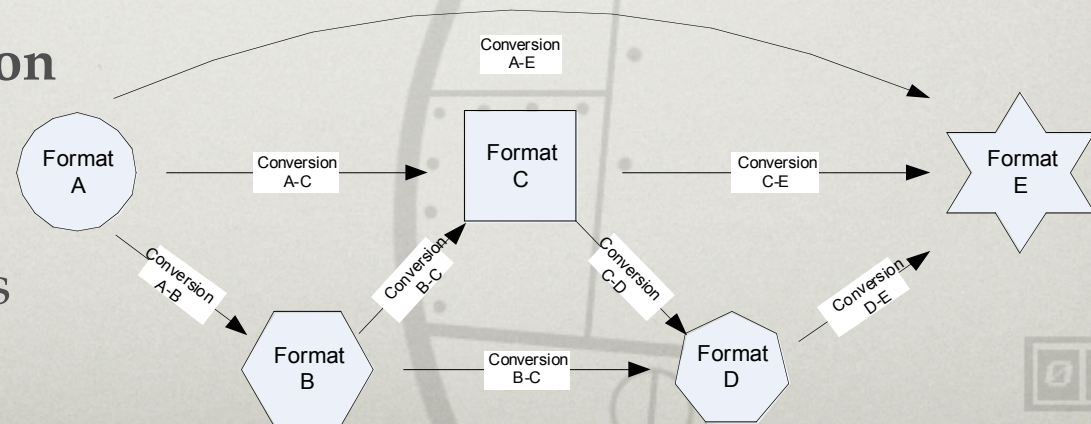
Disadvantages

- Possible loss of information during conversion
- Continued maintenance is needed
- **In the longterm perspective costs are high**



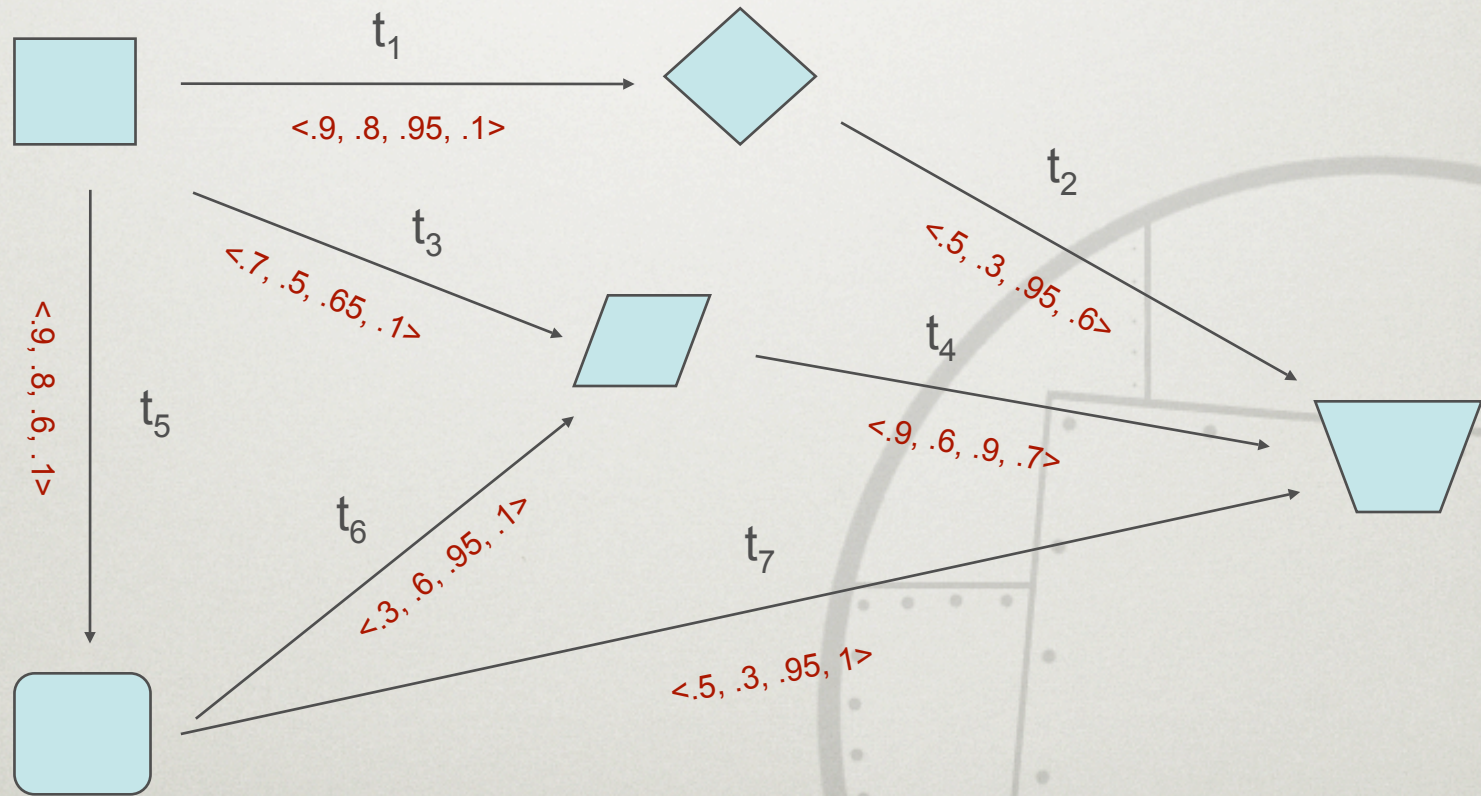
# DISTRIBUTED MIGRATION

- Remote conversion services
  - o known APIs
  - o descriptive metadata for localization and invocation (UDDI)
- Advantages
  - o Platform independency
  - o Redundancy/multiple migration paths
  - o Compatible with other migration strategies
    - Normalization, migration on request
  - o Generalized cost reduction
- Disadvantages
  - o Bandwidth requirements
  - o Slow





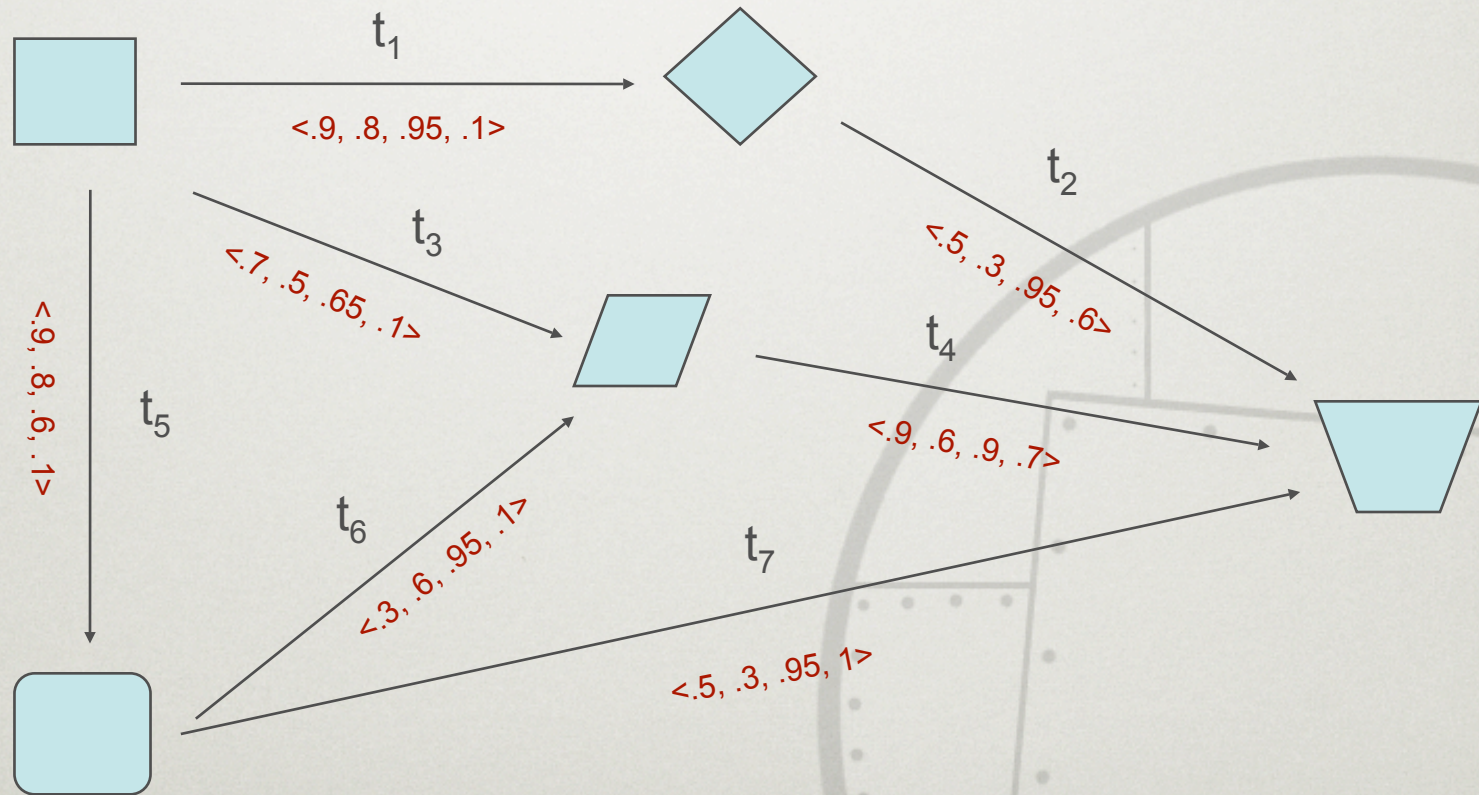
# Distributed Preservation Services





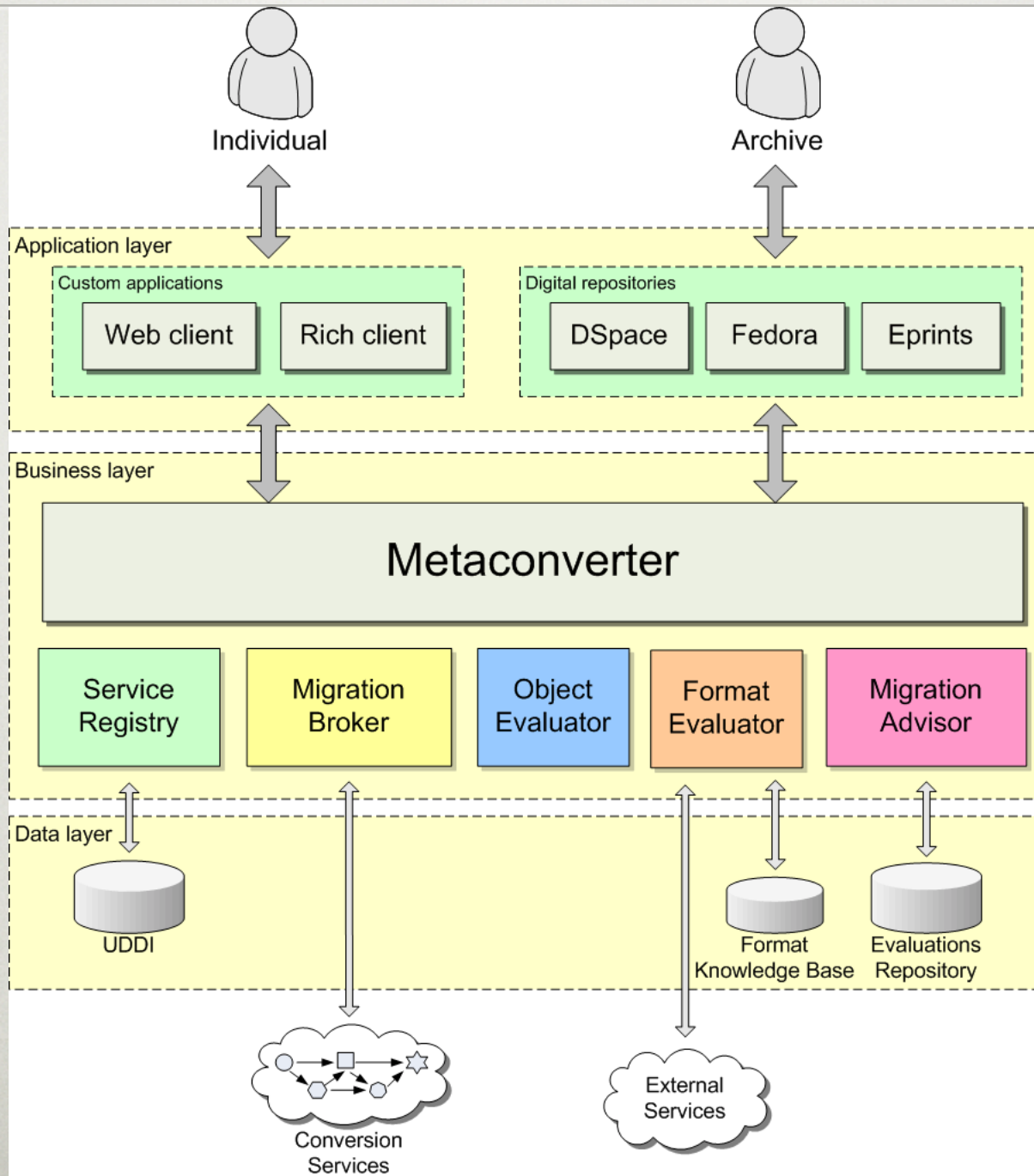
# Distributed Preservation Services

**CRiB project:** <http://crib.dsi.uminho.pt>





# CRIB: ARCHITECTURE





# MIGRATION BROKER

---

- Carries out format conversions
  - o Invokes the necessary conversion services
- Measures the performance of the conversion process
  - o Availability
  - o Stability
  - o Throughput
  - o Scalability
  - o Cost
  - o Size ratio
  - o File count ratio





# OBJECT EVALUATOR

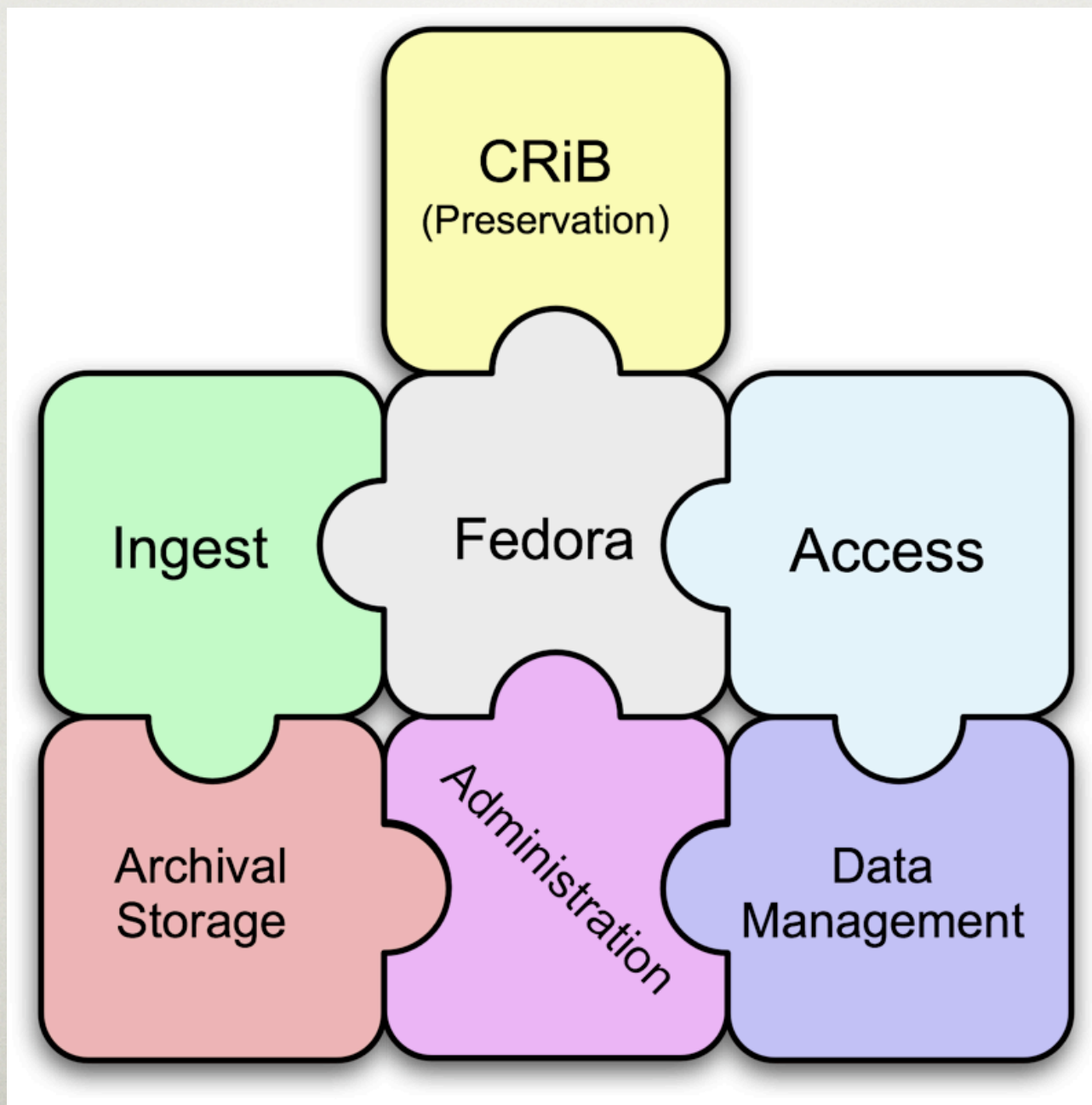
---

- Determines the amount of data loss involved in migration
- Detects similarity between significant properties of digital objects
  - Object class dependent;
  - Different significant properties for bitmap images, text documents, relational databases, etc.
- Produces evaluation reports in PREMIS format
  - Datetime of intervention
  - Description of involved agents
  - Type of event (i.e. Migration)
  - Outcome of intervention





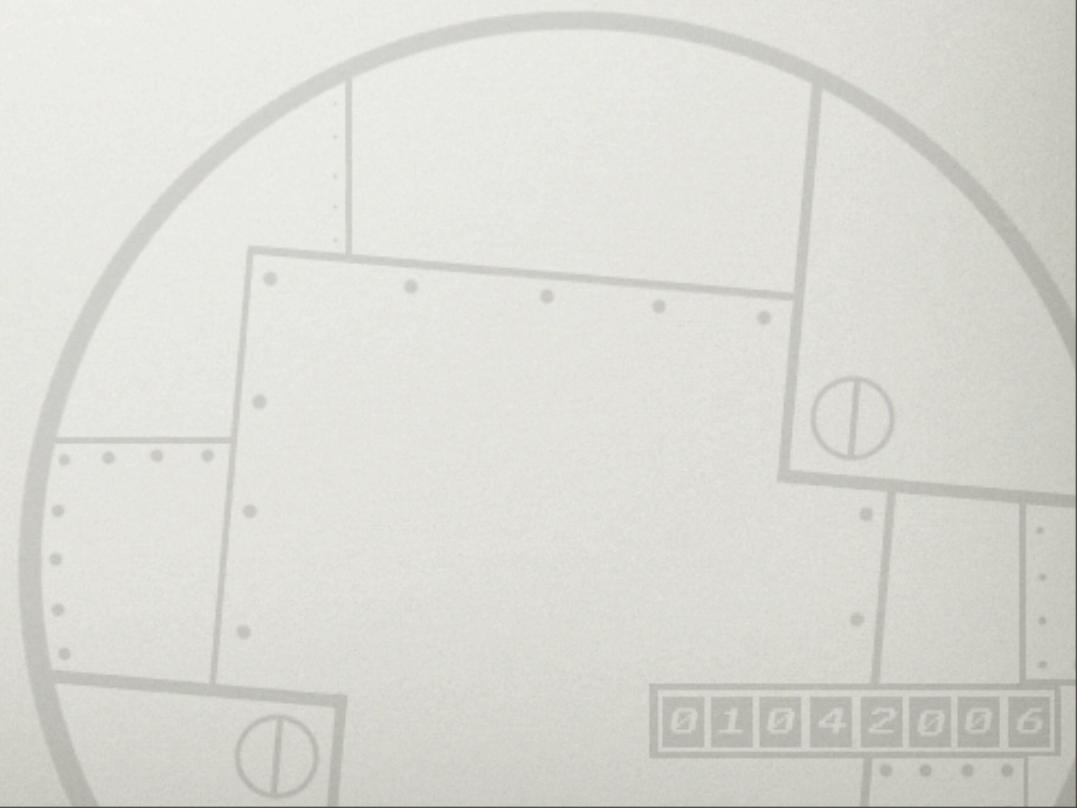
# RODA+CRIB





# TEXT DOCUMENTS AND STILL IMAGES

---





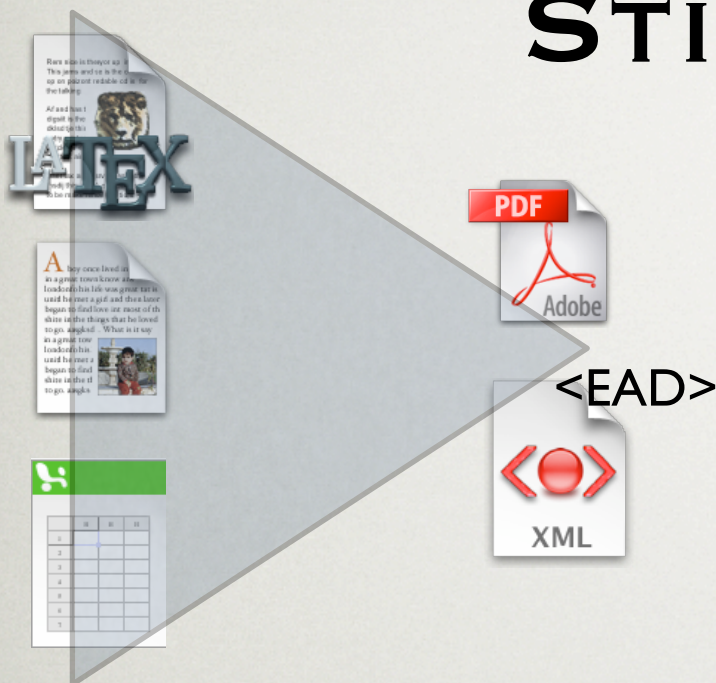
# TEXT DOCUMENTS AND STILL IMAGES

---

- EAD elements capture most of the significant properties: provenance, producer history, context, ...
- Content is kept in a normalized format: PDF and uncompressed TIFF.



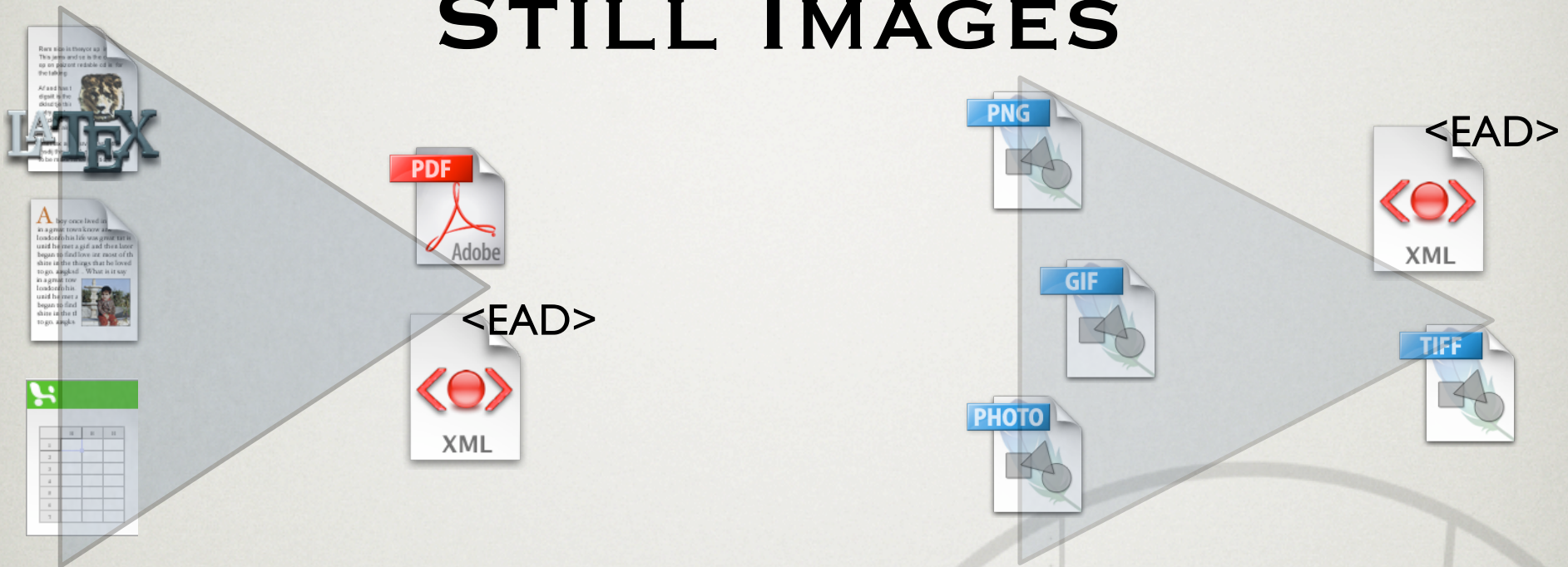
# TEXT DOCUMENTS AND STILL IMAGES



- EAD elements capture most of the significant properties: provenance, producer history, context, ...
- Content is kept in a normalized format: PDF and uncompressed TIFF.



# TEXT DOCUMENTS AND STILL IMAGES



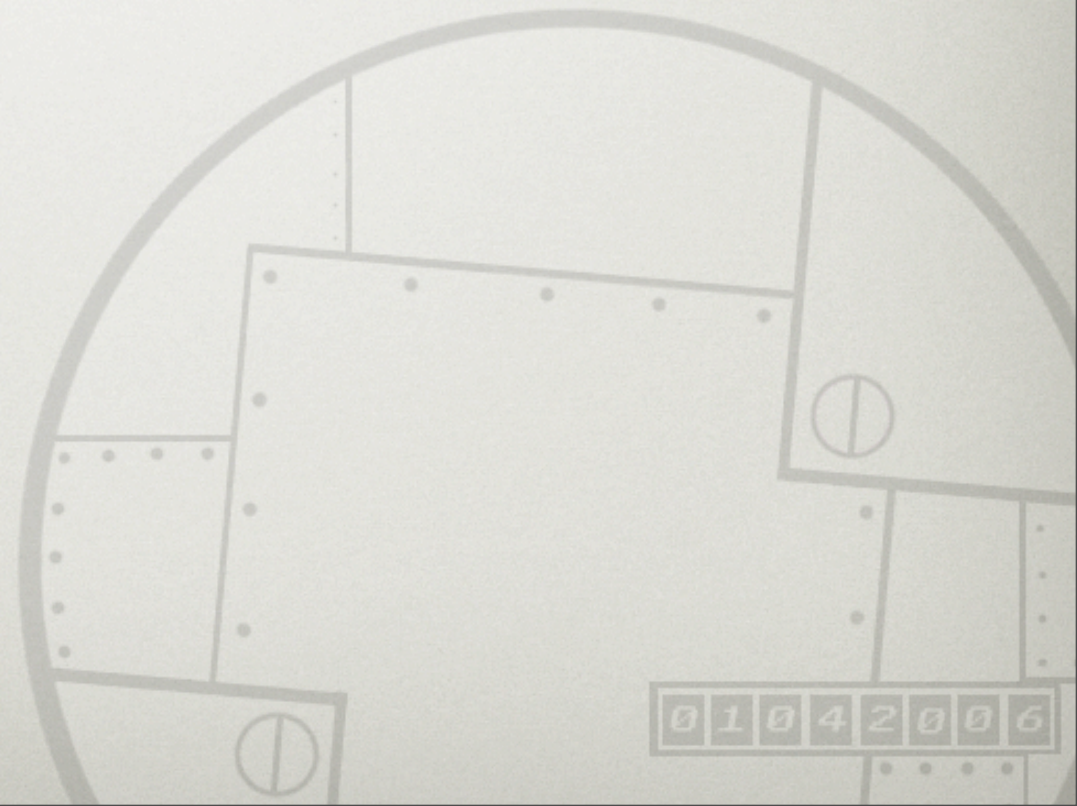
- EAD elements capture most of the significant properties: provenance, producer history, context, ...
- Content is kept in a normalized format: PDF and uncompressed TIFF.



# DATABASES

---

- Data?
- Structure?
- Views?
- Reports?
- Stored Procedures?
- ...





# DATABASES

---

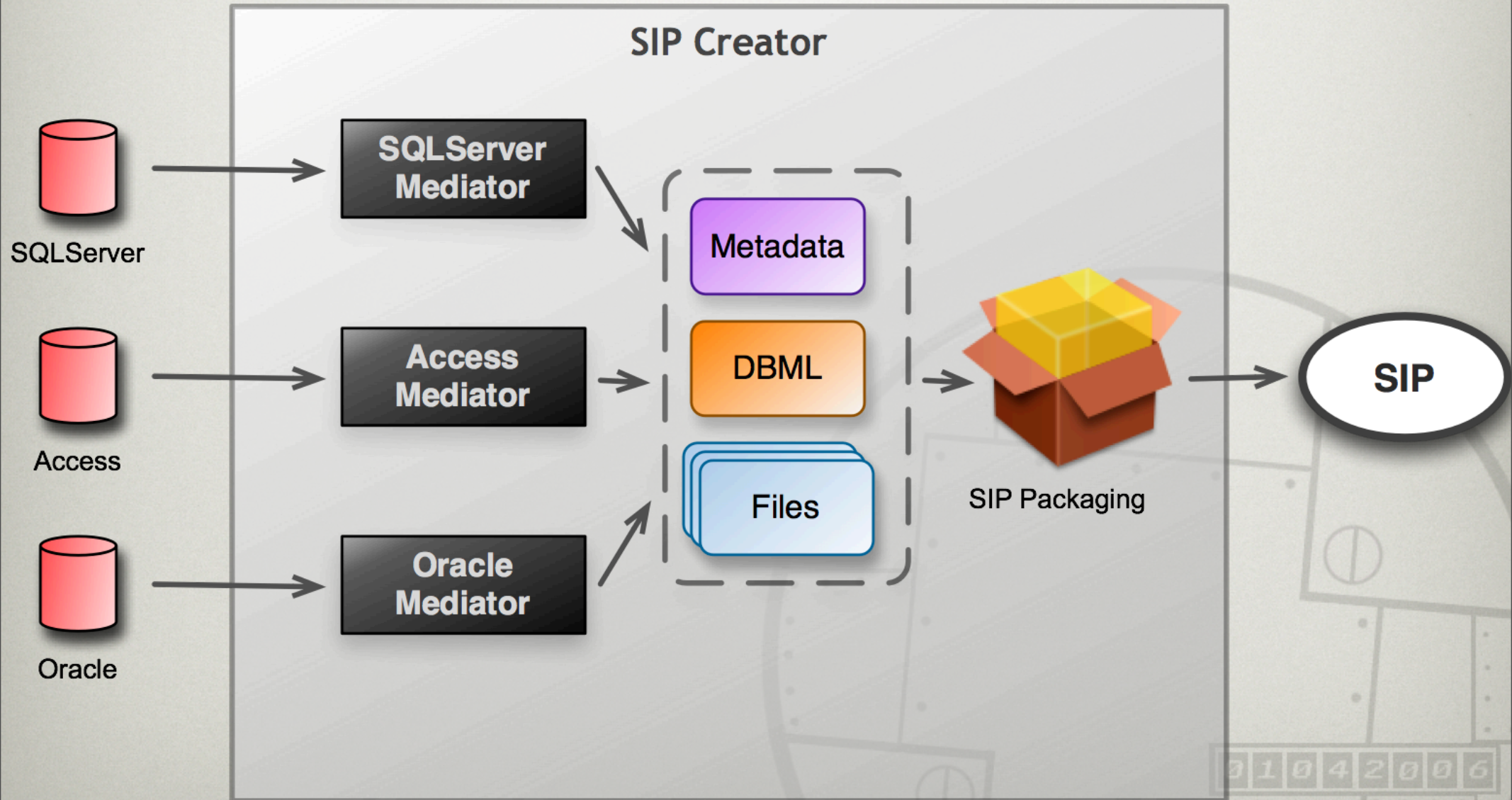
- Data?
- Structure?
- Views?
- Reports?
- Stored Procedures?
- ...

## First prototype:

- Data
- Structure



# SIP BUILDER





# BROWSER

Arquivo Editar Ver Histórico Delicious Marcadores Ferramentas Ajuda

http://roda.di.uminho.pt:8080/roda-wui/#dissemination.browse.17

RODA@DI GWT Javadoc

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

**RODA**

Bem-vindo visitante

Nome de utilizador

Palavra-passe

Registar · Recuperar registo

SOBRE O RODA

CATÁLOGO

catálogo / navegar fundos / 17

## NAVEGAR FUNDOS

6 fundos

- DEMO
  - 0001
    - 00001
    - 2
    - 2
    - 00002**
    - 000005
    - 00006
    - <indefinida>
    - <indefinida>
    - PESSOA
    - xpto
  - 0002
- DGARQ
- FB
- KEEP

Descrição | Visualização | Preservação

### Identificação

Referência	00002
Referência completa	PT/DGARQ/DEMO/0001/00002
Título	Demo de imagens
Nível de descrição	documento simples
Código do país	PT
Código do repositório	DGARQ
Produtor	Ifaria

### Conteúdo e estrutura

Âmbito e conteúdo	demons
-------------------	--------

FECHAR

Concluído



# DISSEMINATORS

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

# RODA

Bem-vindo visitante

[Registar](#) · [Recuperar registo](#)

[SOBRE O RODA](#)

[FASES DO PROJECTO](#) · [RESTAURO DIGITAL](#) · [PROJECTOS ASSOCIADOS](#) · [CONTACTOS](#) · [AJUDA](#)

**CATÁLOGO**

[NAVEGAR FUNDOS](#) · [PESQUISA BÁSICA](#) · [PESQUISA AVANÇADA](#)

*catálogo / navegar fundos / 17*

## NAVEGAR FUNDOS

6 fundos

DEMO

SR 0001

OS 00001

OS 2

OS 2

OS 00002

OS 000005

OS 00006

OS <indefinida>

OS <indefinida>

OS PESSOA

OS xpto

SR 0002

F DGARQ

F FB

F KEEP

**Descrição** | Visualização | Preservação

Disseminações de 00002 - "Demo de imagens"



Descarregar representação  
**(original)**



Abrir no pré-visualizador de  
fotos



Abrir no pré-visualizador de  
livros

FECHAR ✕



# METADATA EDITOR

Ficheiro Editar Ver Histórico Deliciosos Marcadores Ferramentas Ajuda

← → ↻ × 🏠 📄 📁 TAG <http://roda.di.uminho.pt:8080/roda-wui/#dissemination.browse> Google

RODA@DI GWT Javadoc

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÉNTICOS

# RODA

Bem-vindo admin

Terminar sessão

Preferências

SOBRE O RODA

CATÁLOGO

ADMINISTRAÇÃO

INCORPORAÇÃO

catálogo / navegar fundos / 17

## NAVEGAR FUNDOS

6 fundos

- DEMO
  - 0001
    - 00001
    - 2
    - 2
    - 00002**
    - 000005
    - 00006
    - <indefinida>
    - <indefinida>
    - PESSOA
    - xpto
  - 0002
- DGARQ
- FB
- KEEP

Descrição Visualização Preservação

### Identificação

Esconder campos opcionais

Referência	00002
Referência completa	PT/DGARQ/DEMO/0001/00002
Título	Demo de imagens
Nível de descrição	documento simples
Data inicial	<input type="text"/> Janeiro
Data final	<input type="text"/> Janeiro
Código do país	PT
Código do repositório	DGARQ
Produtor	Ifaria
Detalhes específicos	

GRAVAR SUBNÍVEL DUPLICAR MOVER ELIMINAR FECHAR

1 0 4 2 0 0 6

Concluído



# PRESERVATION METADATA VIEWER

The screenshot displays the RODA (Repositório de Objectos Digitais Autênticos) interface. At the top, there is a navigation bar with links for 'SOBRE O RODA', 'FASES DO PROJECTO', 'RESTAURO DIGITAL', 'PROJECTOS ASSOCIADOS', 'CONTACTOS', and 'AJUDA'. Below this is a 'CATÁLOGO' section with links for 'NAVEGAR FUNDOS', 'PESQUISA BÁSICA', and 'PESQUISA AVANÇADA'. The main content area is titled 'NAVEGAR FUNDOS' and shows a list of 6 funds on the left. The selected fund is '00002', which is expanded to show a list of objects including '00001', '2', '2', '00002', '00005', '00006', '<indefinida>', '<indefinida>', 'PESSOA', and 'xpto'. The main view is a calendar for January 2008, showing 'Integrity Check' events for each day from Jan 12 to Jan 18. A modal dialog box titled 'Integrity Check' is open, displaying the message: 'Object integrity was successfully checked' followed by two timestamps: 'Tue, 15 Jan 2008 01:00:00 GMT' and 'Tue, 15 Jan 2008 02:00:01 GMT'. The dialog has a 'FECHAR' button at the bottom right.

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

# RODA

Bem-vindo visitante

Nome de utilizador

Palavra-passe

Registar · Recuperar registo

SOBRE O RODA FASES DO PROJECTO · RESTAURO DIGITAL · PROJECTOS ASSOCIADOS · CONTACTOS · AJUDA

CATÁLOGO NAVEGAR FUNDOS · PESQUISA BÁSICA · PESQUISA AVANÇADA

catálogo / navegar fundos / 17

## NAVEGAR FUNDOS

6 fundos

- DEMO
  - SR 0001
    - OS 00001
    - OS 2
    - OS 2
    - OS 00002
    - OS 00005
    - OS 00006
    - OS <indefinida>
    - OS <indefinida>
    - OS PESSOA
    - OS xpto
  - SR 0002
- DGARQ
- FB
- KEEP

Descrição Visualização P

Original

Integrity Check

Integrity Check

Integrity Check

Integrity Check

Integrity Check

Integrity Check

Integrity Check

Jan 12 Jan 13 Jan 14 Jan 15 Jan 16 Jan 17 Jan 18

2008 Feb

FECHAR



# ADVANCED SEARCH

Ficheiro Editar Ver Histórico Deliciosos Marcadores Ferramentas Ajuda

← → ↻ × 🏠 📄 TAG 🌐 http://roda.di.uminho.pt:8080/roda-wui/#dissemination.search. ⭐ 🔍 Google

RODA@DI GWT javadoc

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

# RODA

Bem-vindo visitante

Nome de utilizador

Palavra-passe

Registrar · Recuperar registo

SOBRE O RODA

CATÁLOGO    NAVEGAR FUNDOS    PESQUISA BÁSICA    PESQUISA AVANÇADA

catálogo / pesquisa / pesquisa avançada

## PESQUISA AVANÇADA

Selecione os níveis de descrição que pretende incluir na pesquisa

todos os níveis     F fundo     C classe     SR serie     DC documento composto

escolher níveis     SF sub-fundo     SC sub-classe     SSR sub-serie     DS documento simples

Selecione o intervalo de tempo em que pretende pesquisar

sem restrições temporais

especificar intervalo    desde  Janeiro  até 2005  Dezembro  31

Preencha as palavras que pretende encontrar

todos os campos   

escolher campos     todas as palavras     frase exacta     pelo menos uma das palavras     sem as palavras

Todas as palavras "testes" terão de aparecer em um dos campos Título dos resultados.

PESQUISAR

1 resultados obtidos

LC	Testes
----	--------

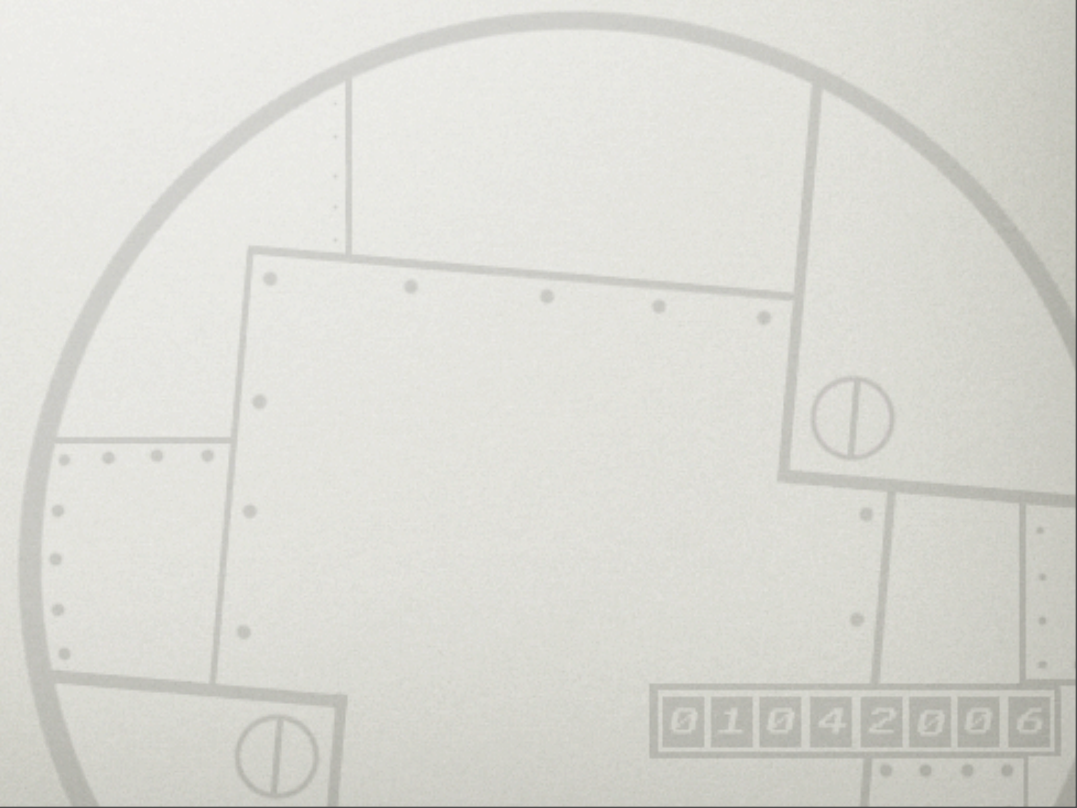
Concluído



# FINAL THOUGHTS

---

*“Data Preservation is a people problem”*  
*Michael Lesk*





# FINAL THOUGHTS

---

*“Data Preservation is a people problem”*  
*Michael Lesk*

- People need to be trained to save data in a proper way.
- What to preserve? Data, Structure, Semantics...
- Preservation is for future users but only today users vote on budget
- We need to make data collecting people have preservation concerns
- Preservation is fault tolerance. All systems are imperfect



# RODA HOMEPAGE



**RODA**

Repositório de Objectos Digitais Autênticos

Pesquisar

## Navegação

- Notícias
- Sobre o RODA
- Disseminação
- Ferramentas e normas
- Projectos associados
- Eventos
- Fóruns
- Votações
- Equipa & contactos

## Society of Archivists 2007 Conference

Submetido por [lfaria](#) em Quarta, 18/10/2006 - 10:57 :: [Arquivística](#)

Start: 28/08/2007 - 08:00

End: 31/08/2007 - 17:00

Timezone: Etc/GMT

The Society of Archivists is hosting a major international conference in 2007 - Differing Directions: Challenging Communities

The Conference will be held at Queen's University Belfast, Northern Ireland on 28 - 31 August 2007. The Society of Archivists (UK and Ireland) covers archivists, records and information management, and archive conservators - and the programme will include three parallel sessions relating to all these professional areas.

For more information see the [Call for Papers](#)

[calendar](#)

## Participação do RODA no 4<sup>a</sup> Congresso Nacional de Administração Pública

Submetido por [rcaastro](#) em Terça, 17/10/2006 - 16:38 :: [Arquivística](#)

Start: 02/11/2006 - 07:00

Timezone: Etc/GMT+1

A Administração Pública portuguesa está a viver dias de mudança profunda. Mudança

## Idiomas

[Português](#)

[English](#)

## Próximos eventos

[Digital rights and asset management: access and digital preservation forum](#) (evento)  
(10 horas)

[mais](#)

## Novos tópicos no fórum

[Questionário sobre PREMIS e EAD no DSpace](#)

[Requisitos do RODA](#)

[Criação de uma identidade para o grupo](#)

[Coleções fotográficas: como estruturá-las?](#)

[Abertura do fórum...](#)

[mais](#)

Nome de utilizador: \*

Palavra passe: \*

[Iniciar sessão](#)

[Criar uma nova conta](#)  
[Pedir uma nova palavra passe](#)

<http://roda.iantt.pt/>

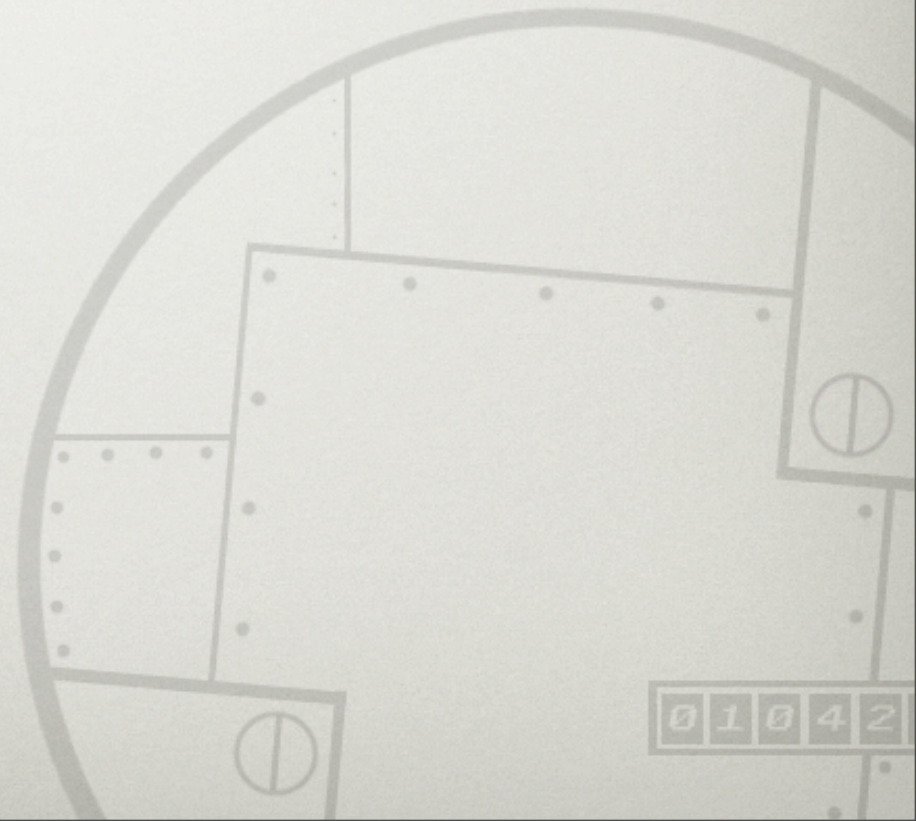
0 1 0 4 2 0 0 6



# BUSINESS MODEL

---

- CRiB is free: [mferreira@keep.pt](mailto:mferreira@keep.pt)
- RODA is *free*: property of Portuguese National Archives
- Support, maintenance and internationalization:  
KEEP Solutions Lda, [www.keep.pt](http://www.keep.pt)





The background features a large, faint circular globe with a grid of latitude and longitude lines. Overlaid on the globe is a rectangular frame with a grid of dots, resembling a technical drawing or a map projection. The text is centered over this graphic.

LET'S PRESERVE  
TOMORROW'S HISTORY...

QUESTIONS?