



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

2007-10-17

# Repositórios Digitais

---

José Carlos Ramalho

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



Universidade  
do Minho

# Conteúdo

- O que são?
- Como são constituídos?
- Como são concebidos e implementados?
- Que agentes e actores?
- Perspectivas de trabalho



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

3

# Por aí...



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

3

# Por aí...

## Bibliotecas Digitais



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

3

# Por aí...

## Bibliotecas Digitais

## Arquivos Digitais



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

3

# Por aí...

## **Bibliotecas Digitais**

## **Arquivos Digitais**

## **Repositórios Digitais**



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

3

# Por aí...

**Bibliotecas Digitais**

**Biblioteca Electrónica**

**Arquivos Digitais**

**Repositórios Digitais**



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

3

# Por aí...

**Bibliotecas Digitais**

**Biblioteca Electrónica**

**Arquivos Digitais**

**Repositórios Digitais**

**Biblioteca Virtual**



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

Por aí...

# **Bibliotecas Digitais**

**Biblioteca Electrónica**

# **Arquivos Digitais**

**Repositórios Digitais**

**Biblioteca Virtual**



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

Por aí...

# **Bibliotecas Digitais**

# **Arquivos Digitais**



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

# Por aí...



# O que são?

## Bibliotecas Digitais

Uma biblioteca digital é uma biblioteca na qual as colecções estão armazenadas em formato digital (por oposição ao papel, microfilme, ou outro suporte físico) e acessíveis por computador.

O conteúdo digital pode estar armazenado localmente ou pode ser acedido remotamente através da Internet.



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

O que são?

# Arquivos Digitais (diferem)



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

# O que são?

## Arquivos Digitais (diferem)

● *Fontes primárias*



# O que são?

## Arquivos Digitais (diferem)

### ● *Fontes primárias*

Contêm essencialmente *fontes primárias* de informação (cartas, processos judiciais, registos paroquiais, ..., produzidos directamente por um indivíduo ou organização) em lugar das fontes secundárias normalmente encontradas numa biblioteca (livros, etc);



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

# O que são?

## Arquivos Digitais (diferem)

- *Fontes primárias*



# O que são?

## Arquivos Digitais (diferem)

- *Fontes primárias*
- *Conteúdos organizados em grupos*



# O que são?

## Arquivos Digitais (diferem)

- *Fontes primárias*
- *Conteúdos organizados em grupos*  
Têm os seus **conteúdos organizados em grupos**. Enquanto numa biblioteca os itens são catalogados individualmente, num arquivo são agrupados por proveniência (indivíduo ou instituição criador) e ordem original (ordem mantida pelo criador);



# O que são?

## Arquivos Digitais (diferem)

- *Fontes primárias*
- *Conteúdos organizados em grupos*



# O que são?

## Arquivos Digitais (diferem)

- *Fontes primárias*
- *Conteúdos organizados em grupos*
- *Conteúdos únicos*



# O que são?

## Arquivos Digitais (diferem)

- *Fontes primárias*
- *Conteúdos organizados em grupos*
- *Conteúdos únicos*

Têm ***conteúdos únicos***. Enquanto um livro pode ser encontrado em várias bibliotecas, os registos de um arquivo são únicos e não podem ser encontrados ou consultados noutro local que não seja o seu arquivo.



# O que são?

## Arquivos Digitais (diferem)

- *Fontes primárias*
- *Conteúdos organizados em grupos*
- *Conteúdos únicos*



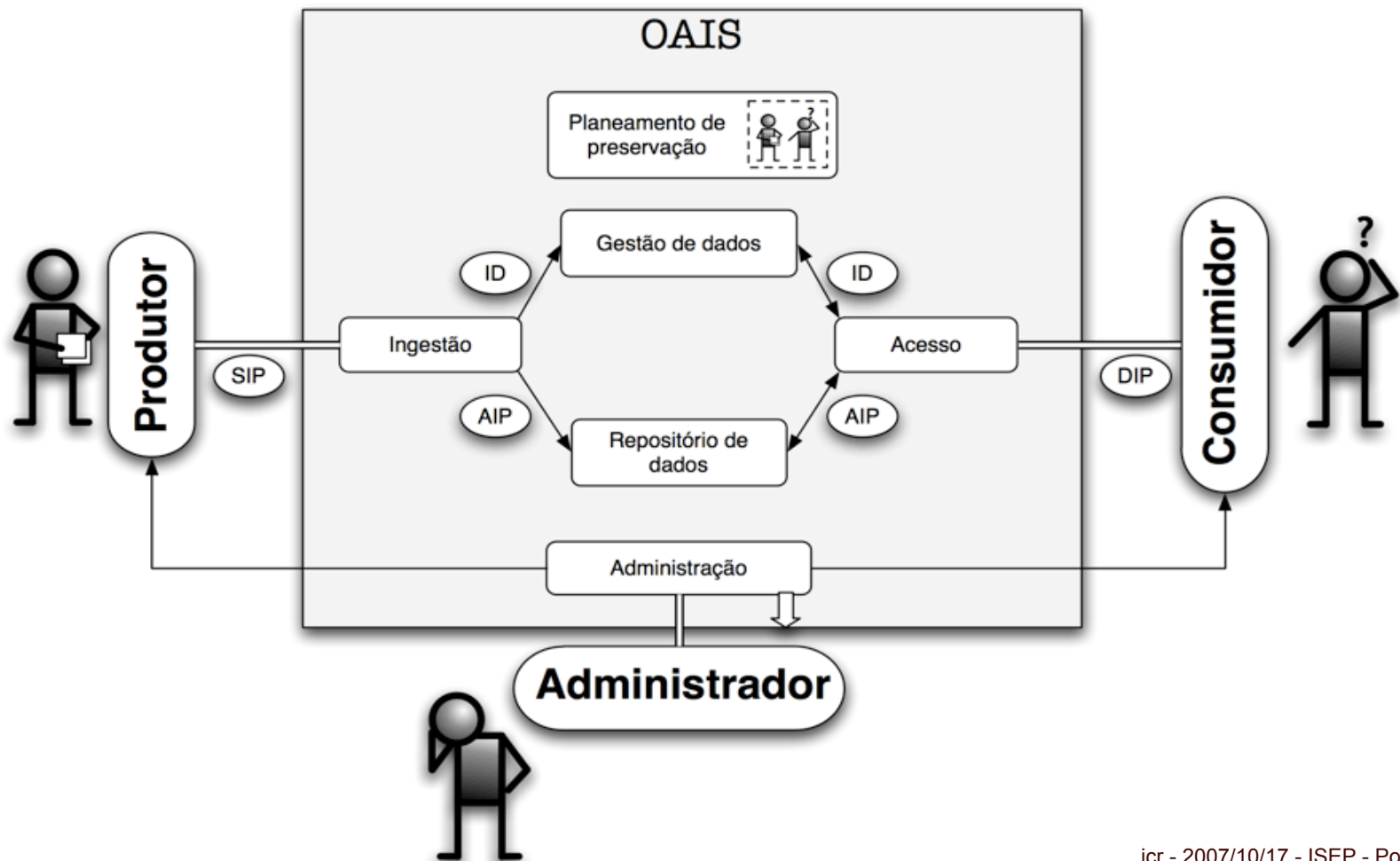
# Vantagens

- Ausência de fronteiras geográficas;
- Disponibilidade 24h/24h, 365 dias ao ano;
- Acessos Múltiplos;
- Acesso Estruturado ao conteúdo;
- Recuperação de Informação;
- Conservação e Preservação;
- Espaço Físico;
- Networking;
- Custo.



Universidade  
do Minho

# OAIS (Open Archival Information system)





# OAIS (resumo)

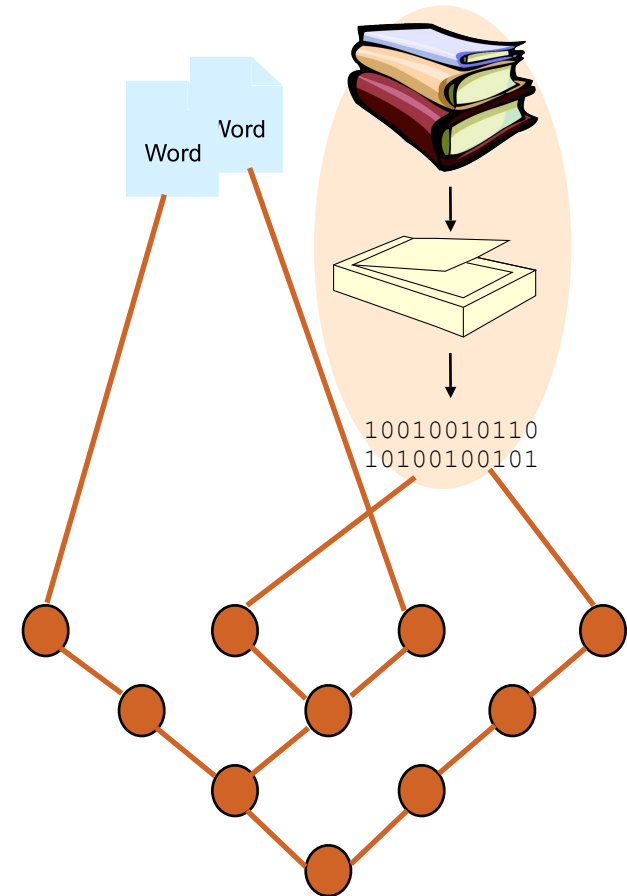
- 3 actores: produtor, administrador, consumidor
- 3 mega-processos: ingestão, administração, disseminação/distribuição
- 3 pacotes de informação: SIP, AIP e DIP



Universidade  
do Minho

# Objecto Digital (OD)

domínio físico

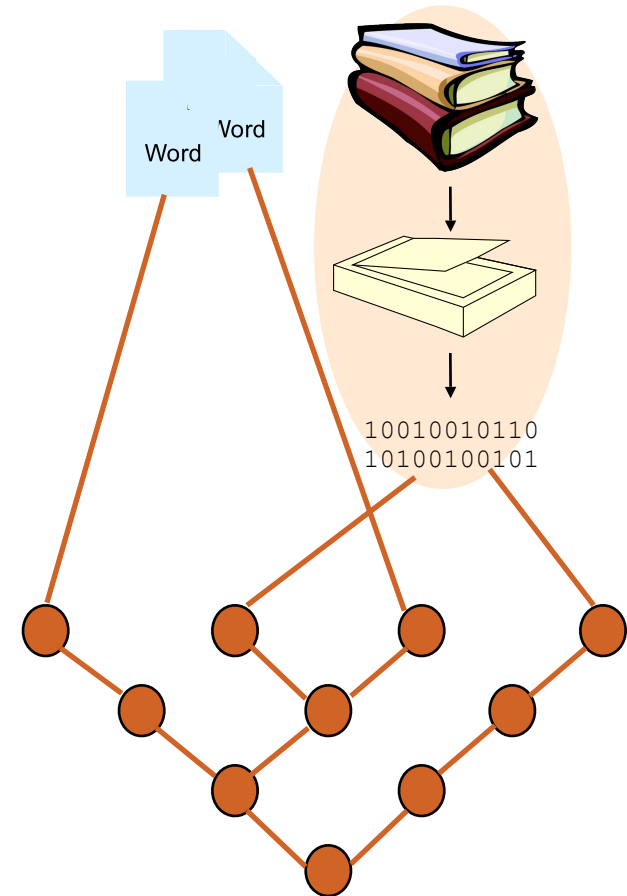




Universidade  
do Minho

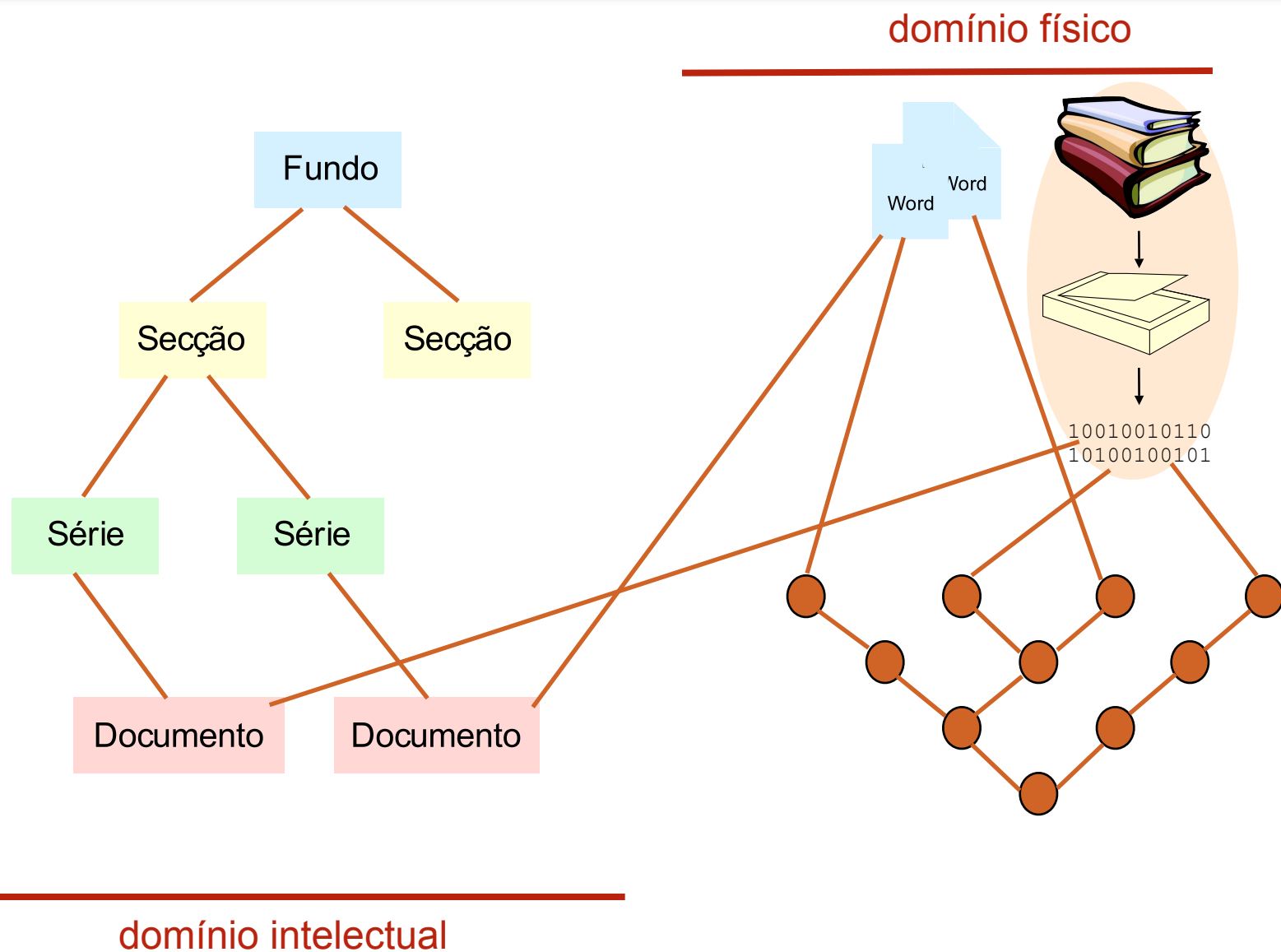
# Objecto Digital (OD)

domínio físico





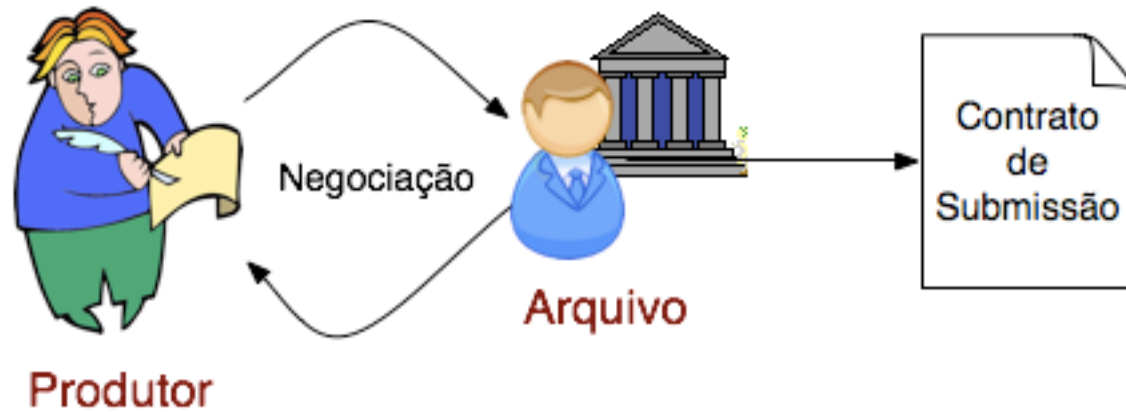
# Objecto Digital (OD)





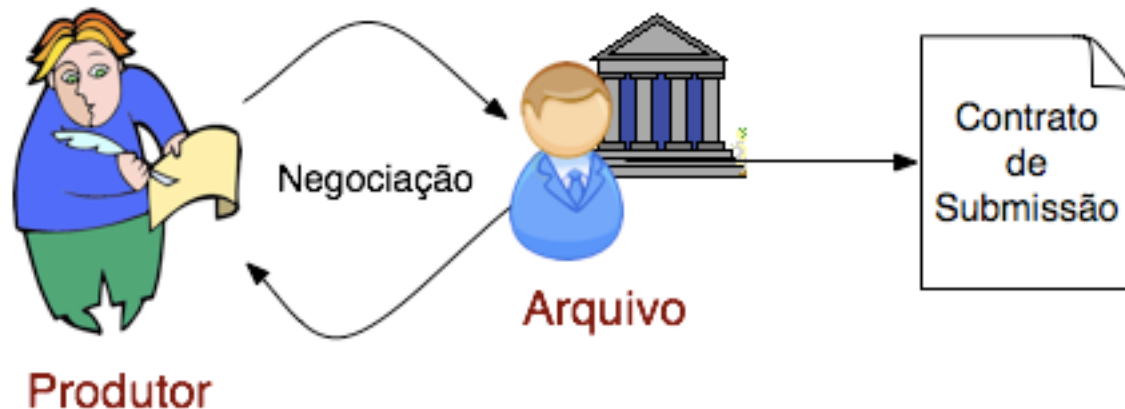
Universidade  
do Minho

# Ingestão





# Ingestão



## Contrato de Submissão

- Definição da estrutura do SIP
- Especificação das várias etapas do processo de ingestão



Universidade  
do Minho

# Estrutura dum SIP (exemplo)



Uma imagem

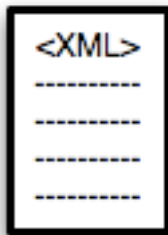


Universidade  
do Minho

# Estrutura dum SIP (exemplo)



Uma imagem



propriedades  
da criação:

- data
- hardware
- ...

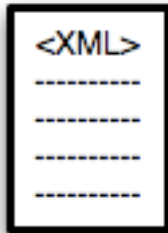


Universidade  
do Minho

# Estrutura dum SIP (exemplo)

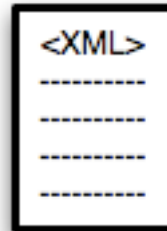


Uma imagem



propriedades  
da criação:

- data
- hardware
- ...



Metainformação Técnica:

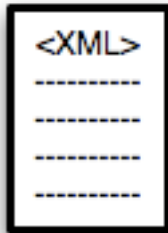
- cor
- dimensões
- ...



# Estrutura dum SIP (exemplo)

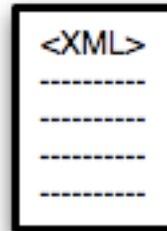


Uma imagem



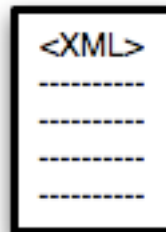
propriedades  
da criação:

- data
- hardware
- ...



Metainformação Técnica:

- cor
- dimensões
- ...

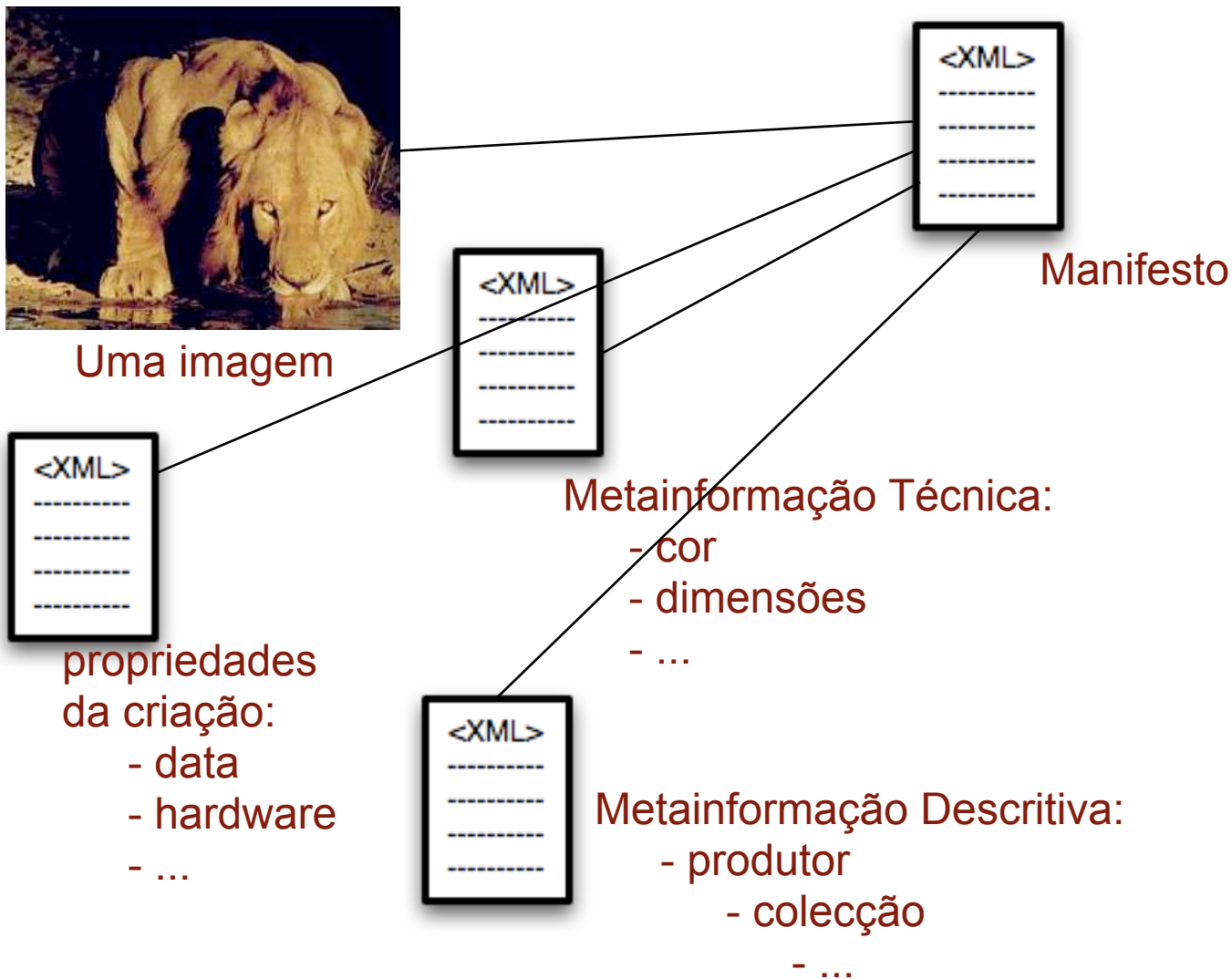


Metainformação Descritiva:

- produtor
- colecção
- ...

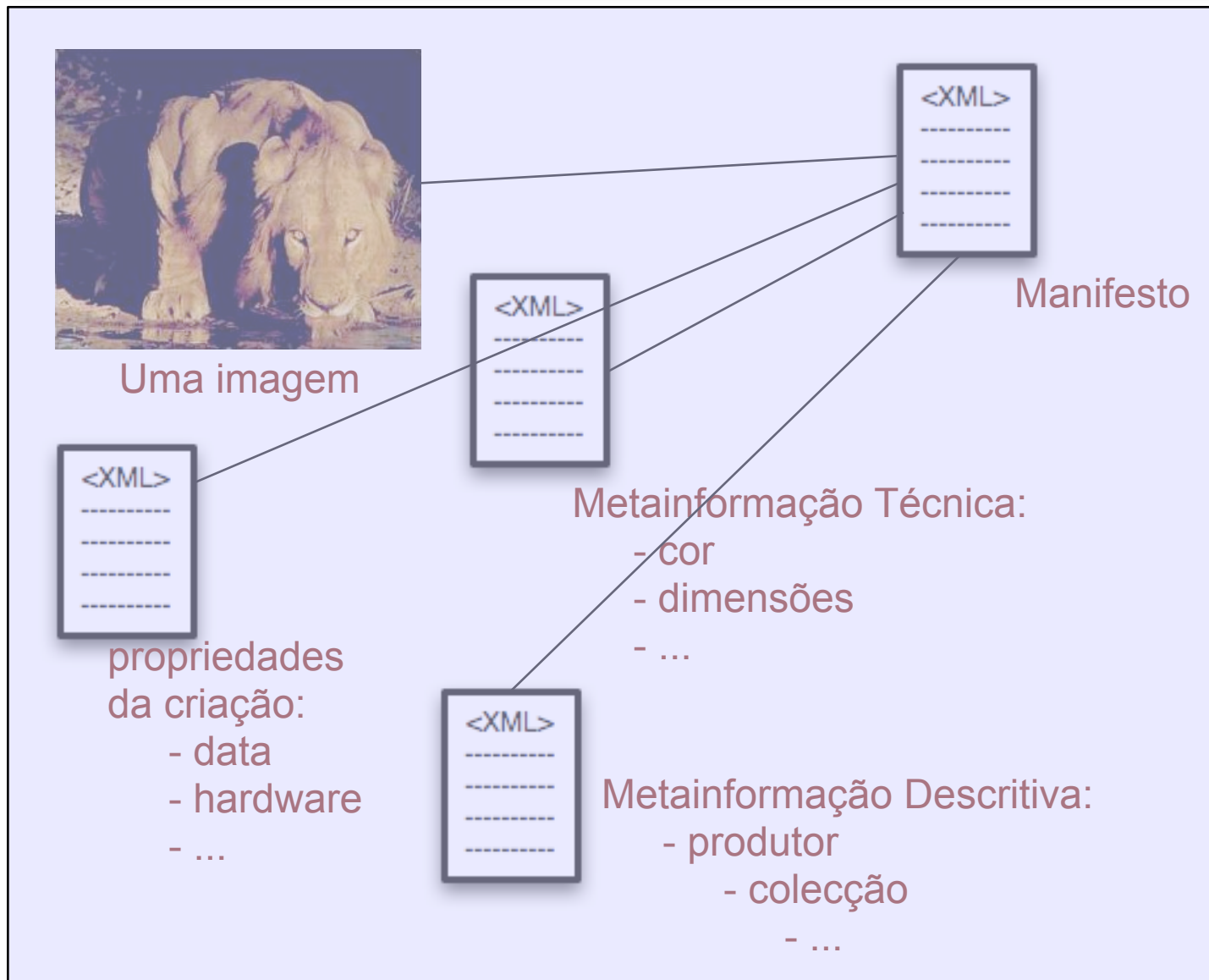


# Estrutura dum SIP (exemplo)





# Estrutura dum SIP (exemplo)





Universidade  
do Minho

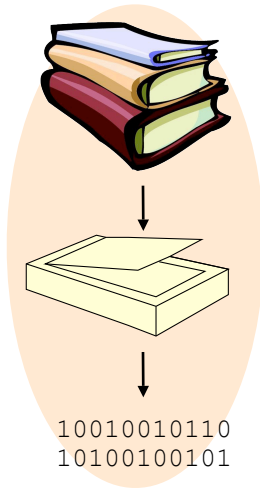
Pesquisa de Informação e Inteligência Artificial

# Estrutura dum SIP (+complexo)



Universidade  
do Minho

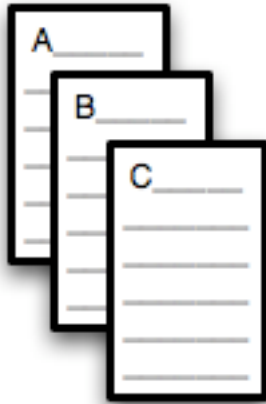
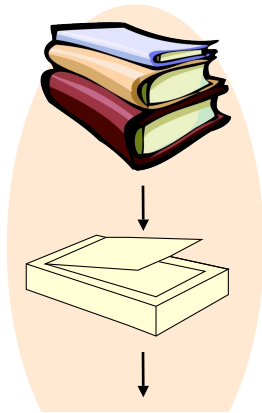
# Estrutura dum SIP (+complexo)





Universidade  
do Minho

# Estrutura dum SIP (+complexo)

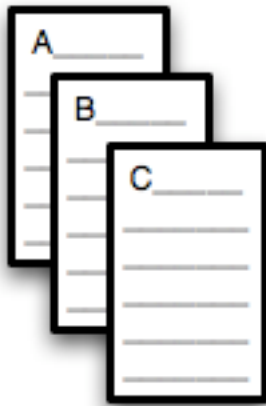
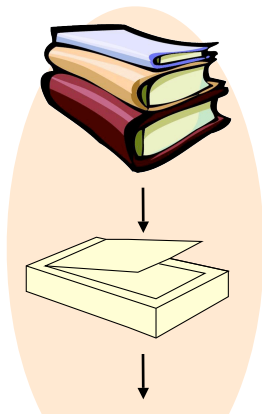


OD = Imagem+



Universidade  
do Minho

# Estrutura dum SIP (+complexo)



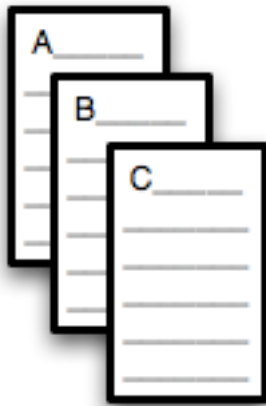
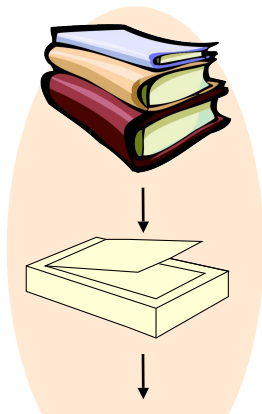
OD = Imagem  $\tau$  Propriedades





Universidade  
do Minho

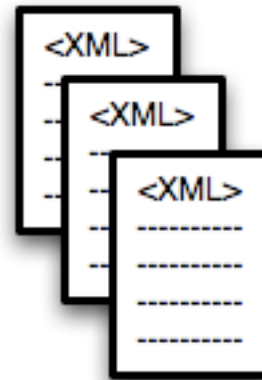
# Estrutura dum SIP (+complexo)



OD = Imagem



Propriedades

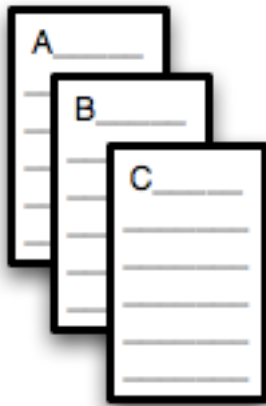
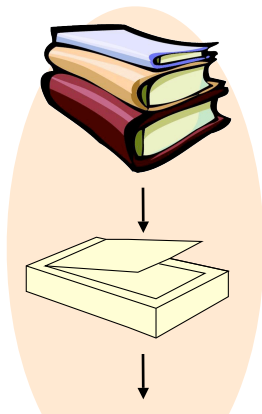


MI Técnica



Universidade  
do Minho

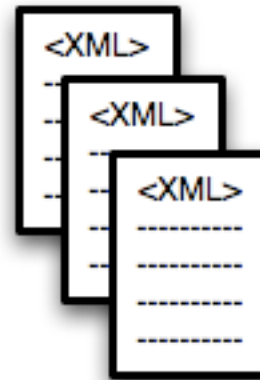
# Estrutura dum SIP (+complexo)



OD = Imagem



Propriedades



MI Técnica

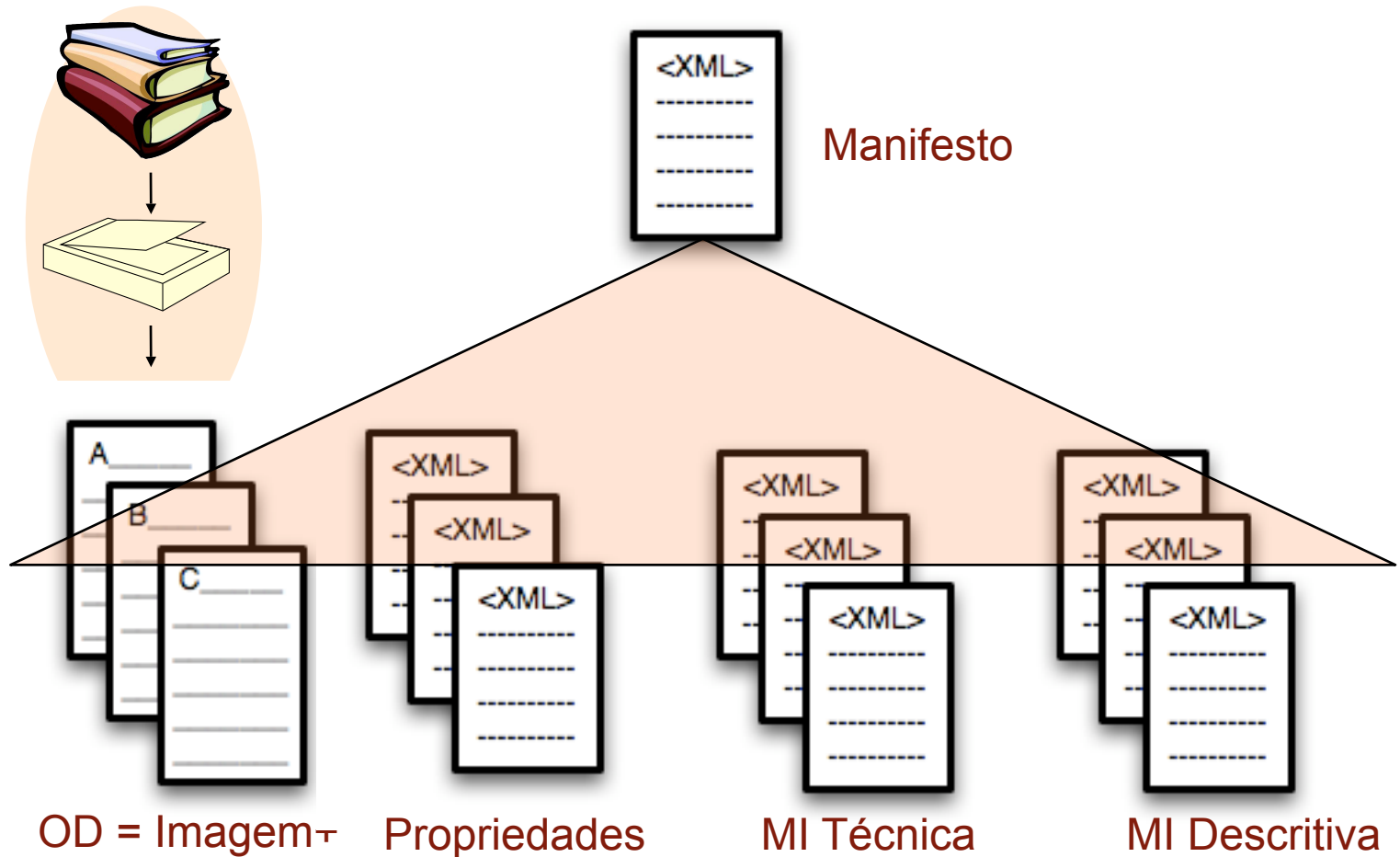


MI Descritiva



Universidade  
do Minho

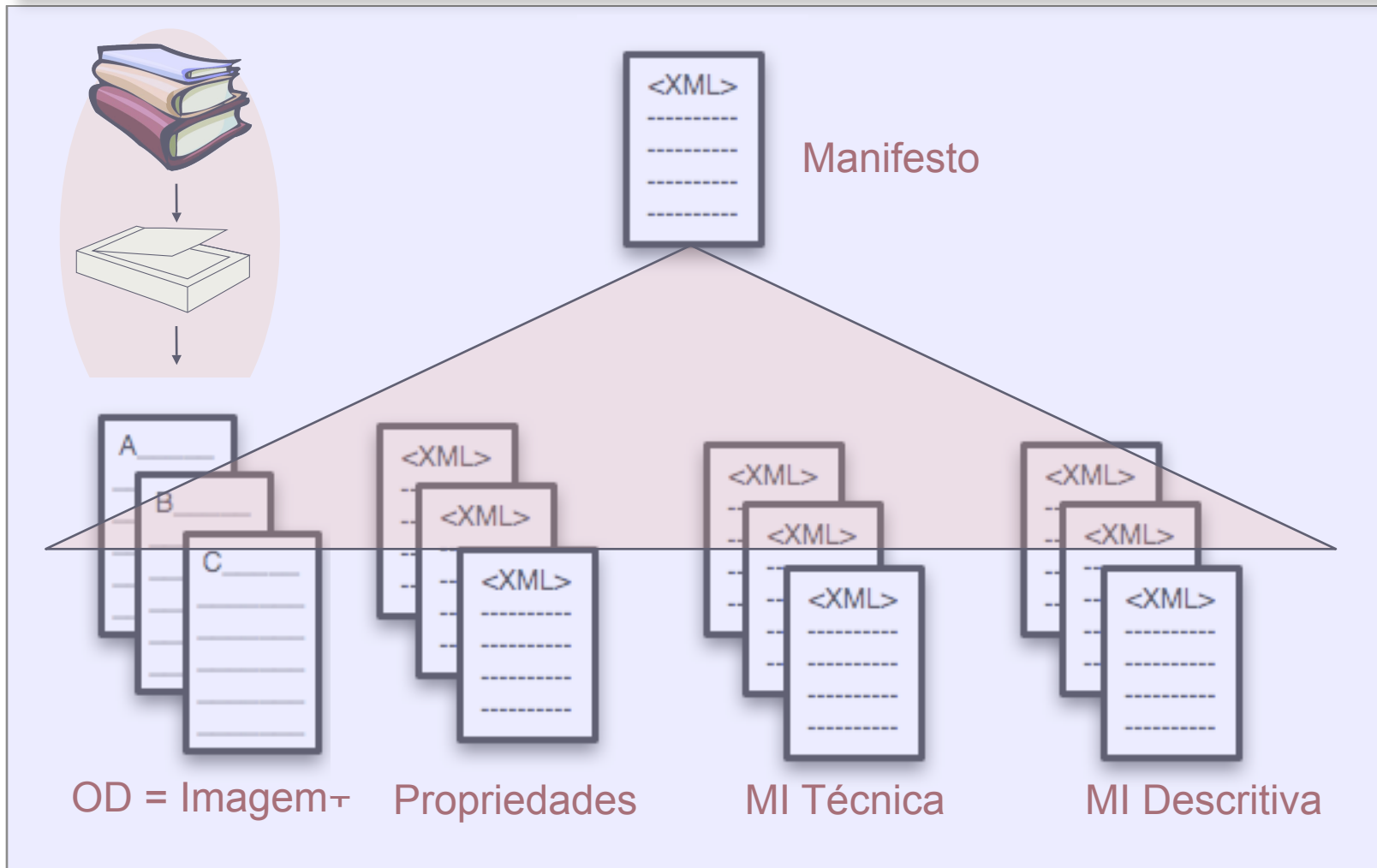
# Estrutura dum SIP (+complexo)





Universidade  
do Minho

# Estrutura dum SIP (+complexo)



**Ficheiro Compactado**



# Especificação de um SIP

- **Metainformação Descritiva**
  - Dublin Core
  - EAD: Encoded Archival Description
  - Outras: Fedora = FOXML + DC + RDF
- **Metainformação Técnica**
  - Depende do tipo de objecto
  - Imagens: NISO MIX (Z39.87)
  - Bases de Dados Relacionais: DBML
- **Especificação do Manifesto**
  - METS: Metadata Encoding and Transmission Standard
  - <http://www.loc.gov/standards/mets/>



Universidade  
do Minho

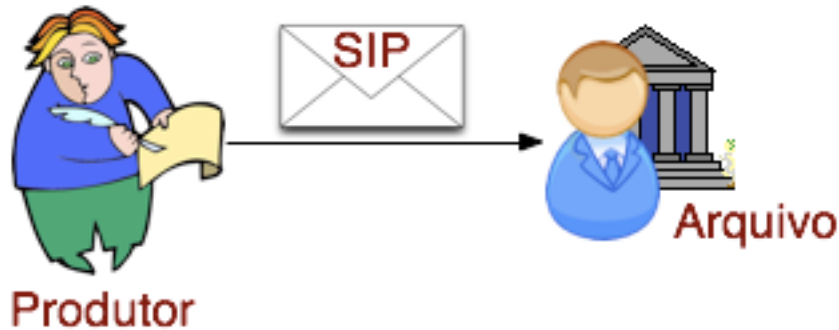
Pesquisa de Informação e Inteligência Artificial

# Etapas do processo de Ingestão



Universidade  
do Minho

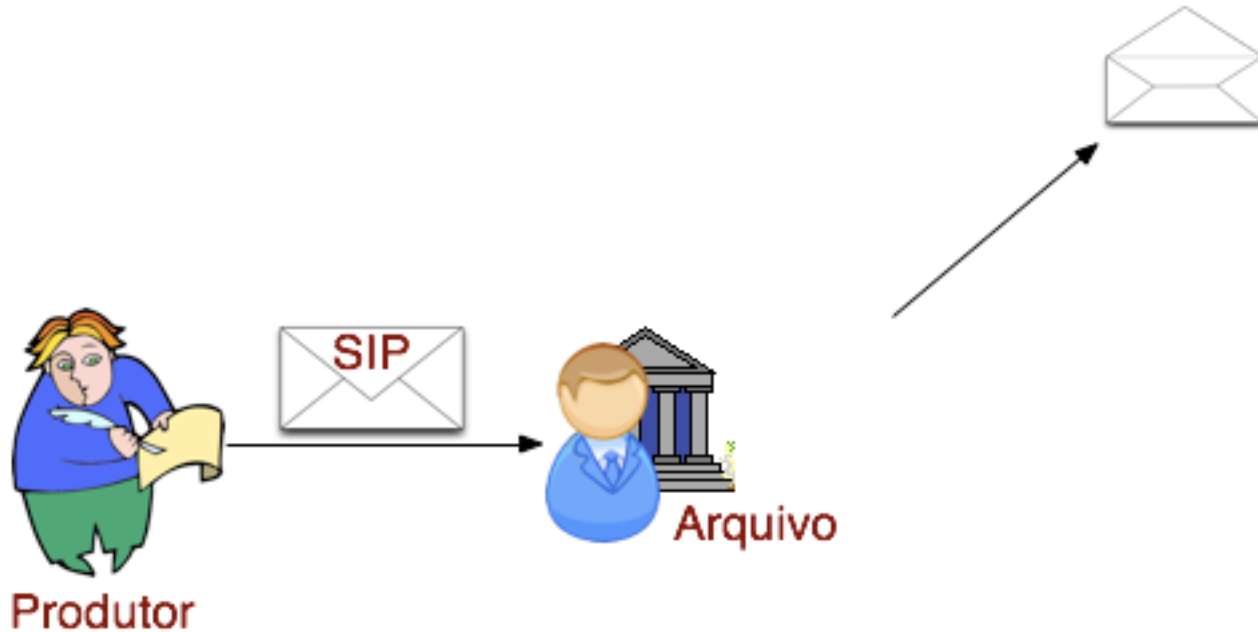
# Etapas do processo de Ingestão





Universidade  
do Minho

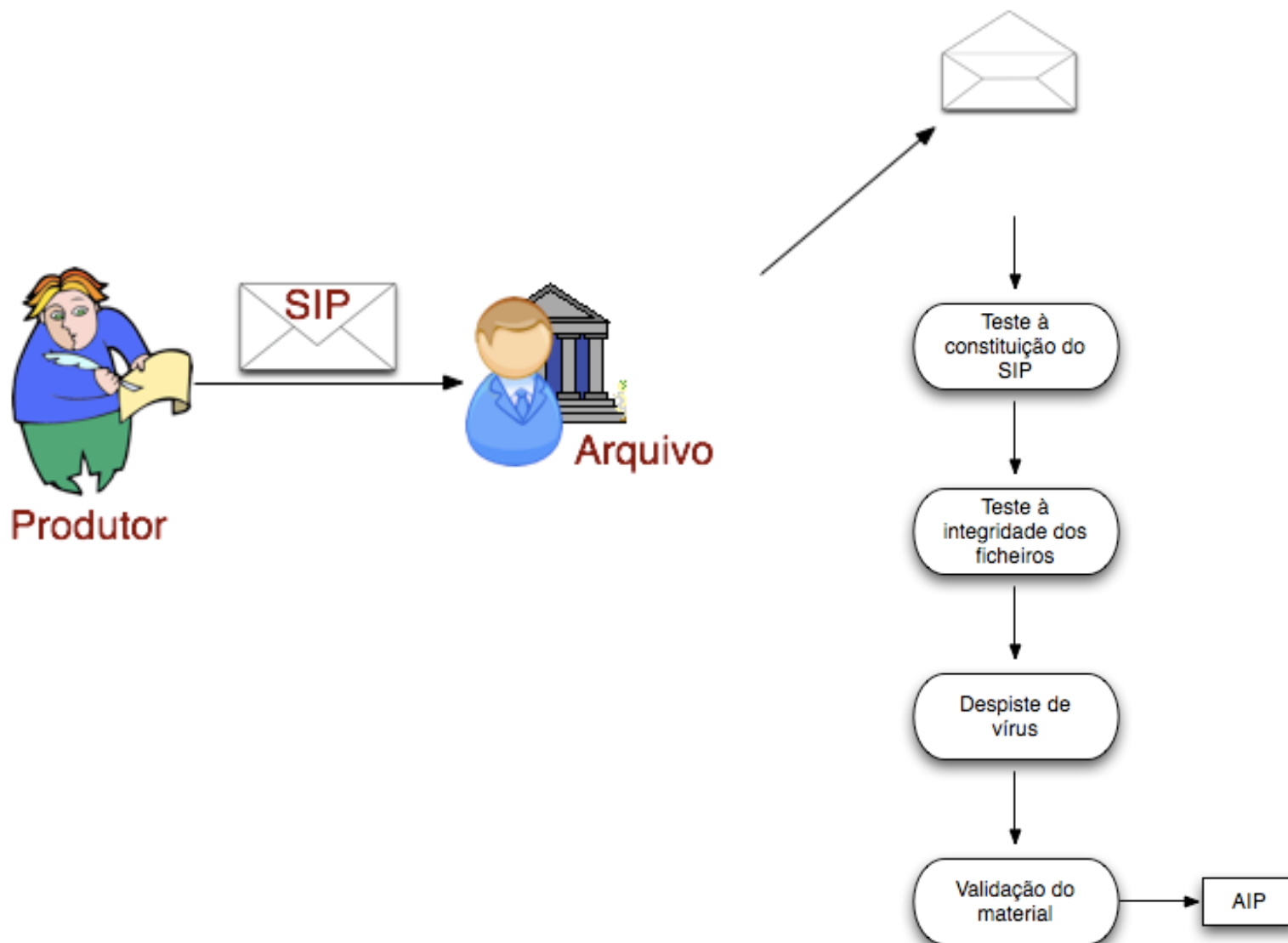
# Etapas do processo de Ingestão





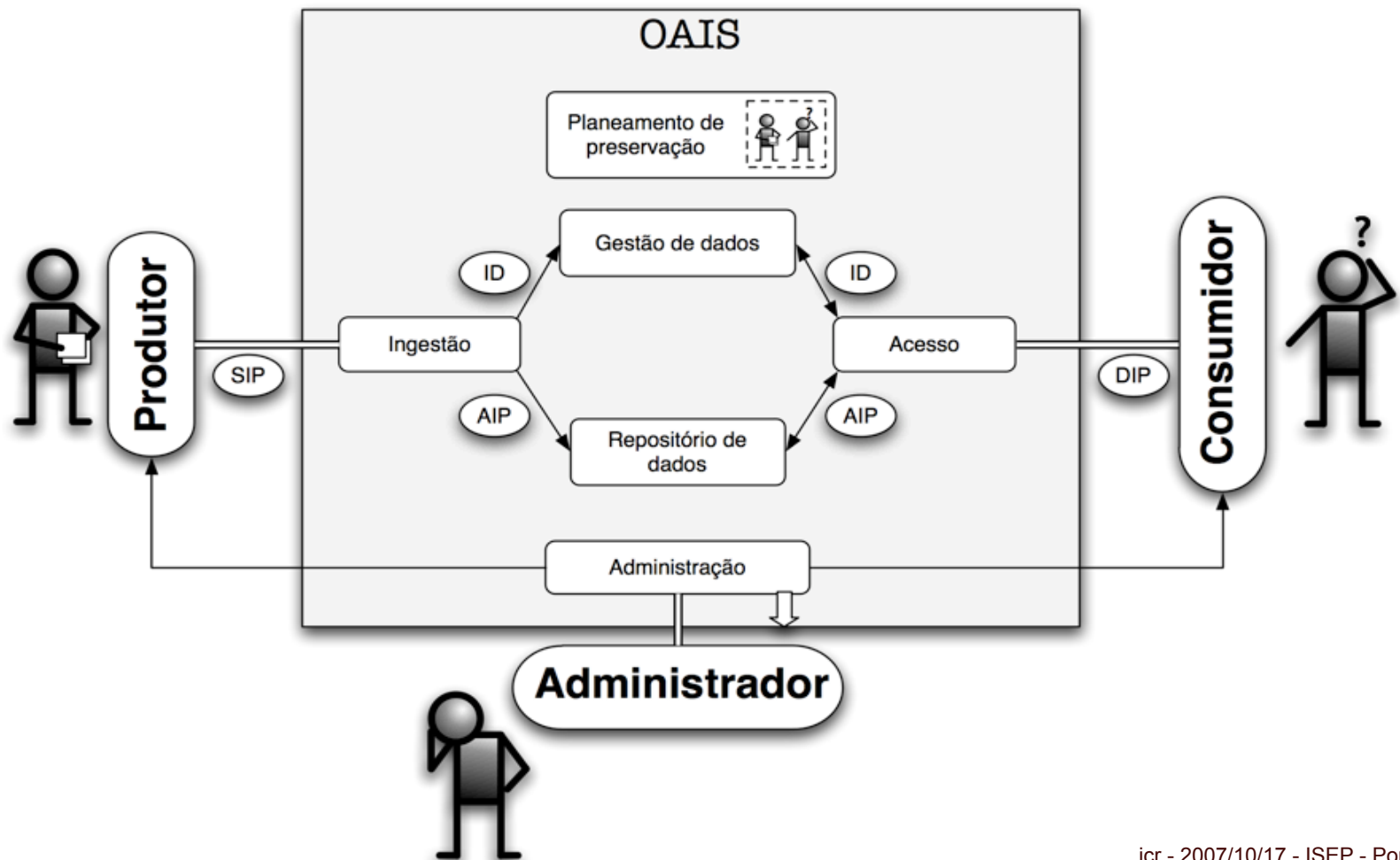
Universidade  
do Minho

# Etapas do processo de Ingestão





# OAIS: relembrando o modelo





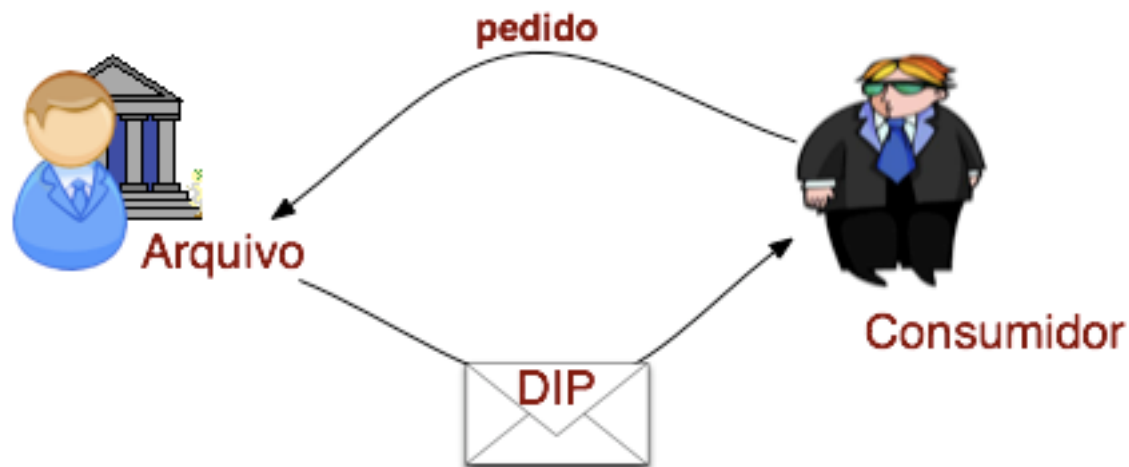
# Gestão

- Auditorias
  - Edição e actualização de metainformação
- **Preservação**
  - Políticas de Migração
  - Políticas de Refrescamento
- Manipulação de ODs
  - Alteração da Estrutura
  - Associações entre o domínio intelectual e o domínio físico
- ❖ Tudo isto suportado por interfaces amigáveis



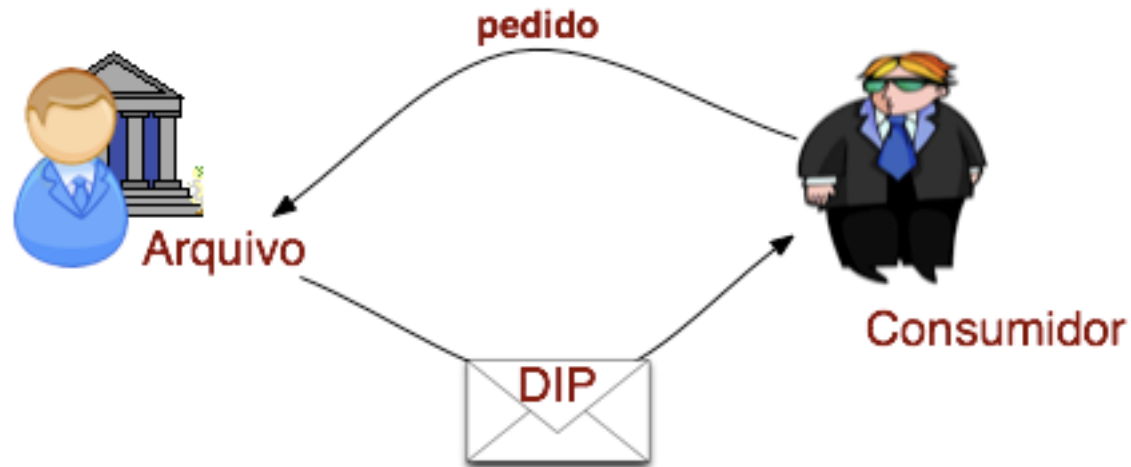
Universidade  
do Minho

# Disseminação





# Disseminação



- DIP
  - ★ Website
  - ★ Pacote semelhante ao SIP
  - ★ Front-end dum motor de pesquisa



Universidade  
do Minho

# Tecnologia para Implementação

fedora



digit Arq



ejprints



Tools and technology  
for Open Repositories

<MyCoRe>

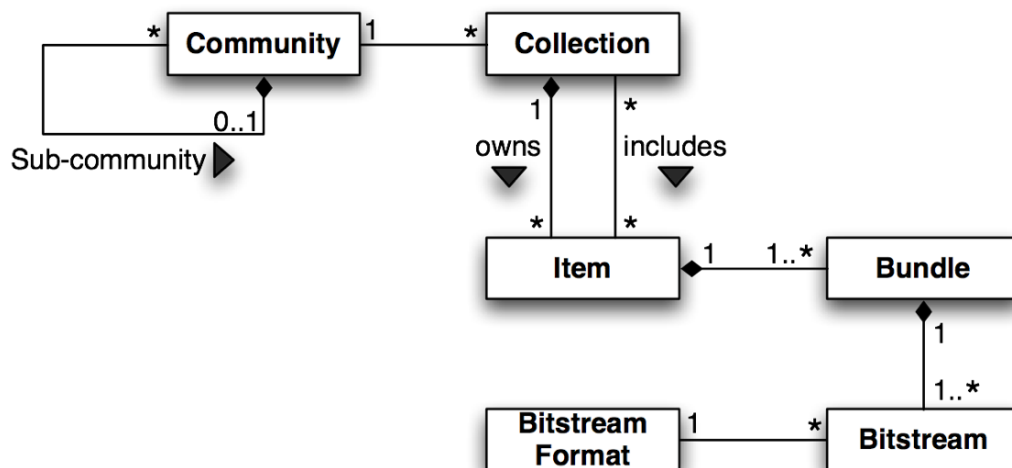
Archimède





# Tecnologias Diferentes

## DSpace





Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

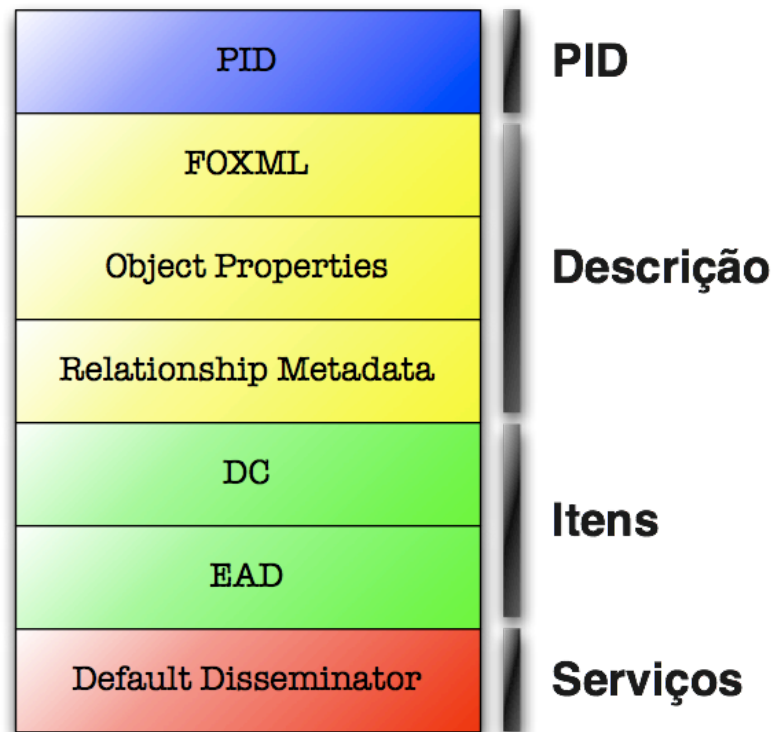
# Tecnologias Diferentes



Universidade  
do Minho

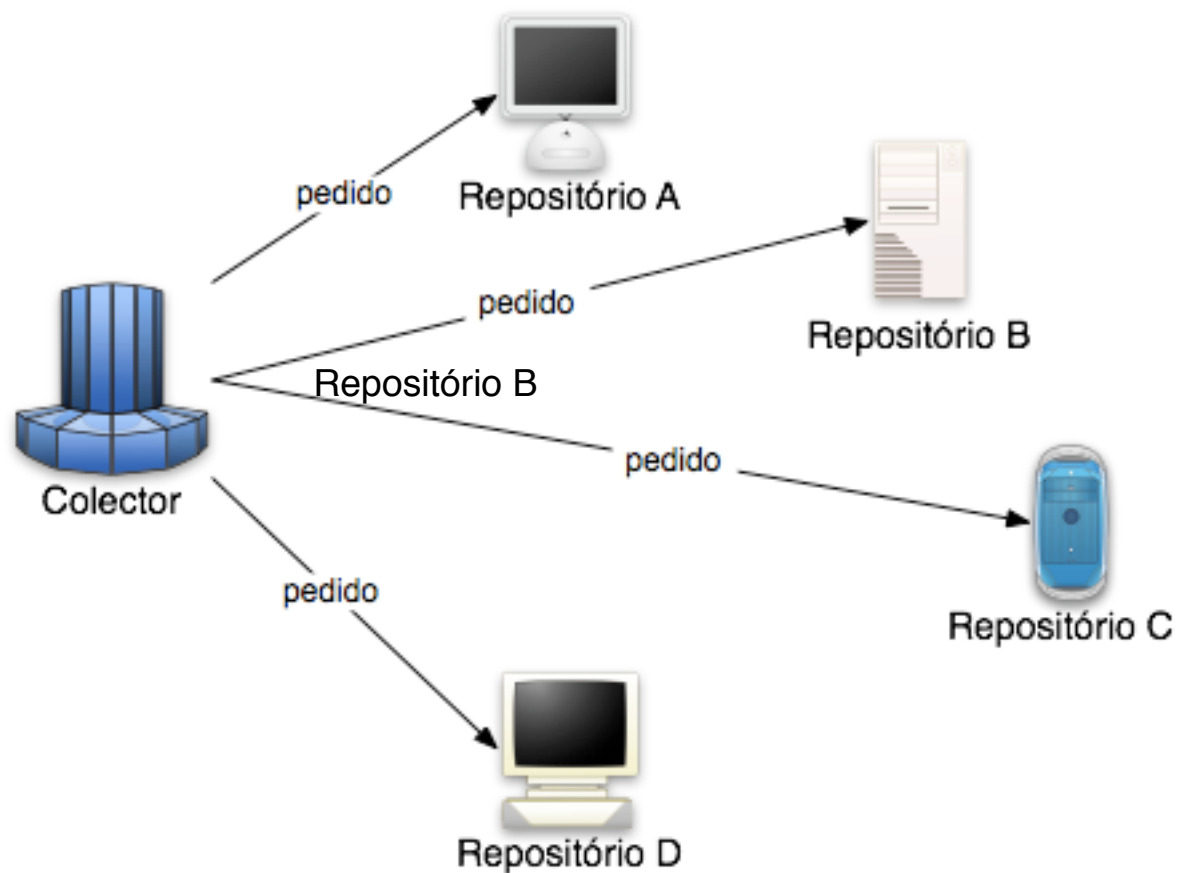
# Tecnologias Diferentes

## Fedora





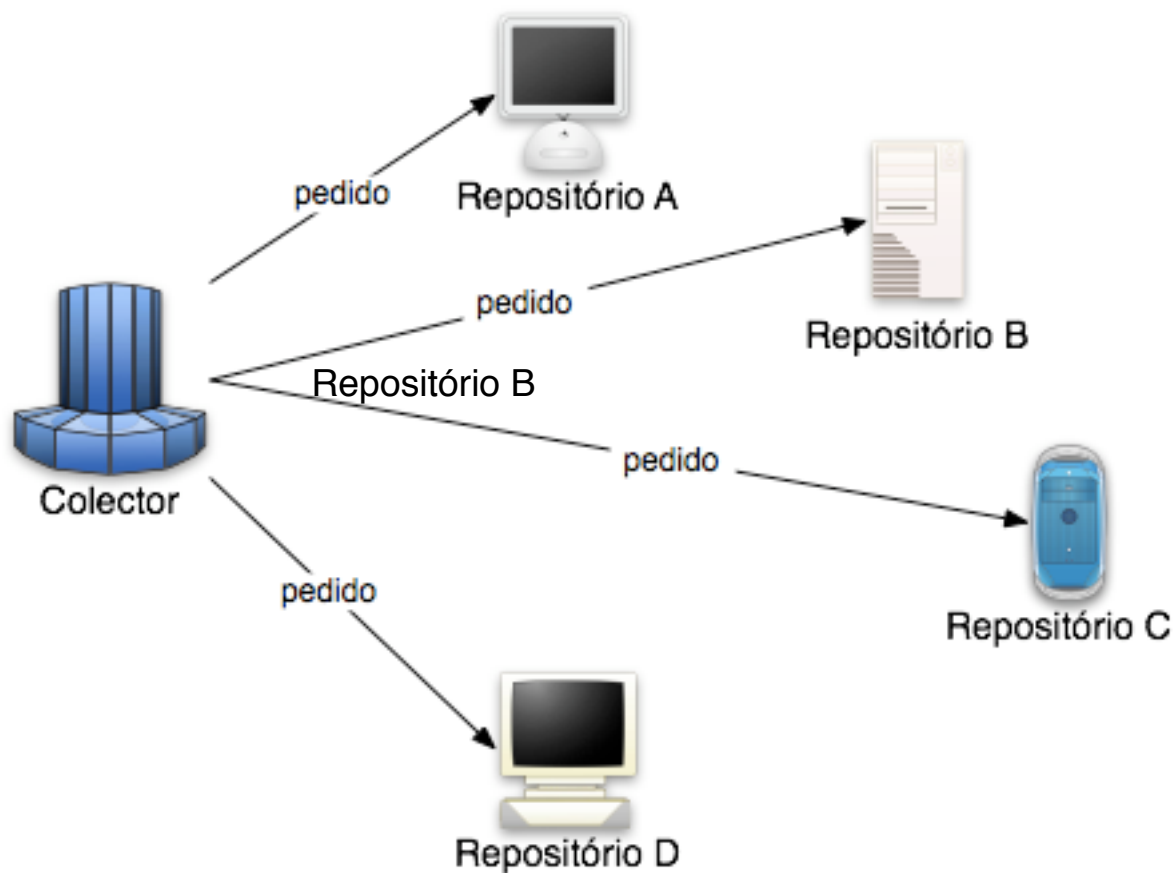
# Federação de Repositórios





Universidade  
do Minho

# Federação de Repositórios



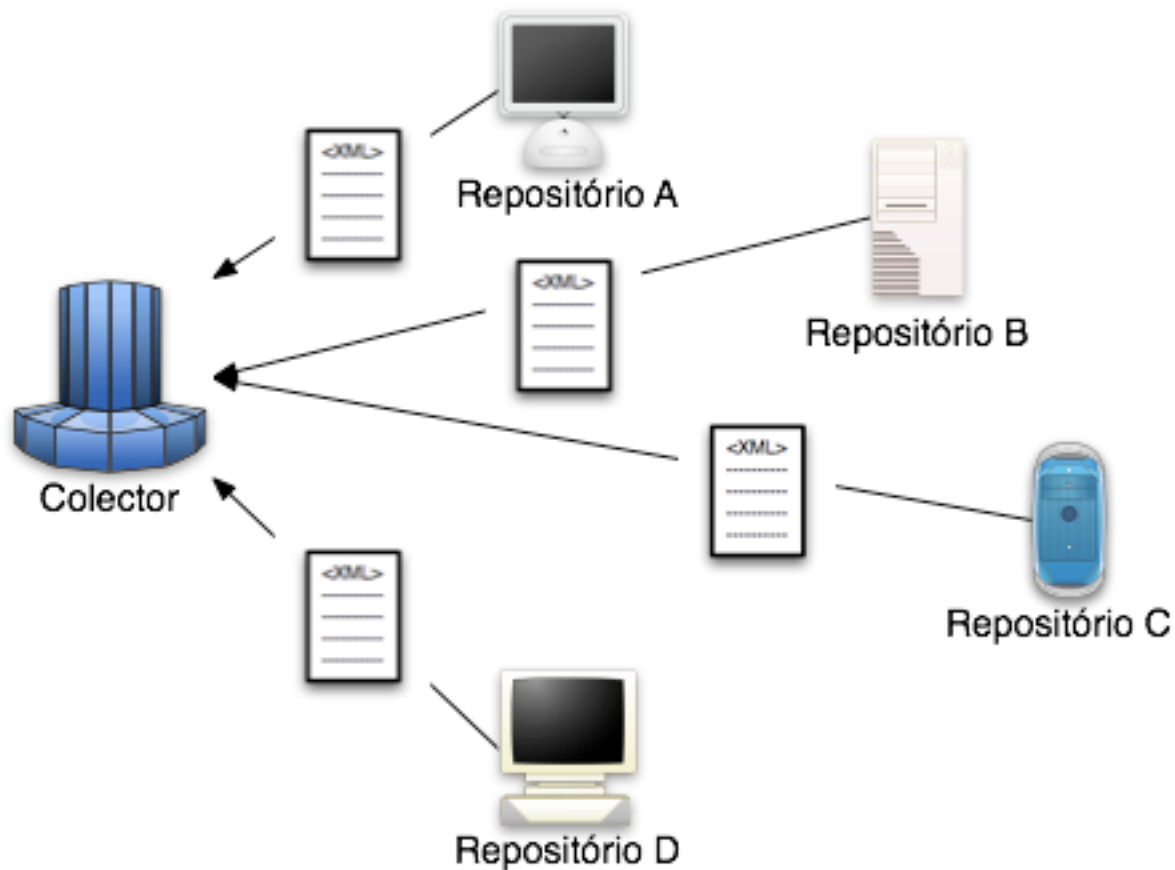
## Protocolo OAI-PMH

Open Archives Initiative Protocol for Metadata Harvesting



Universidade  
do Minho

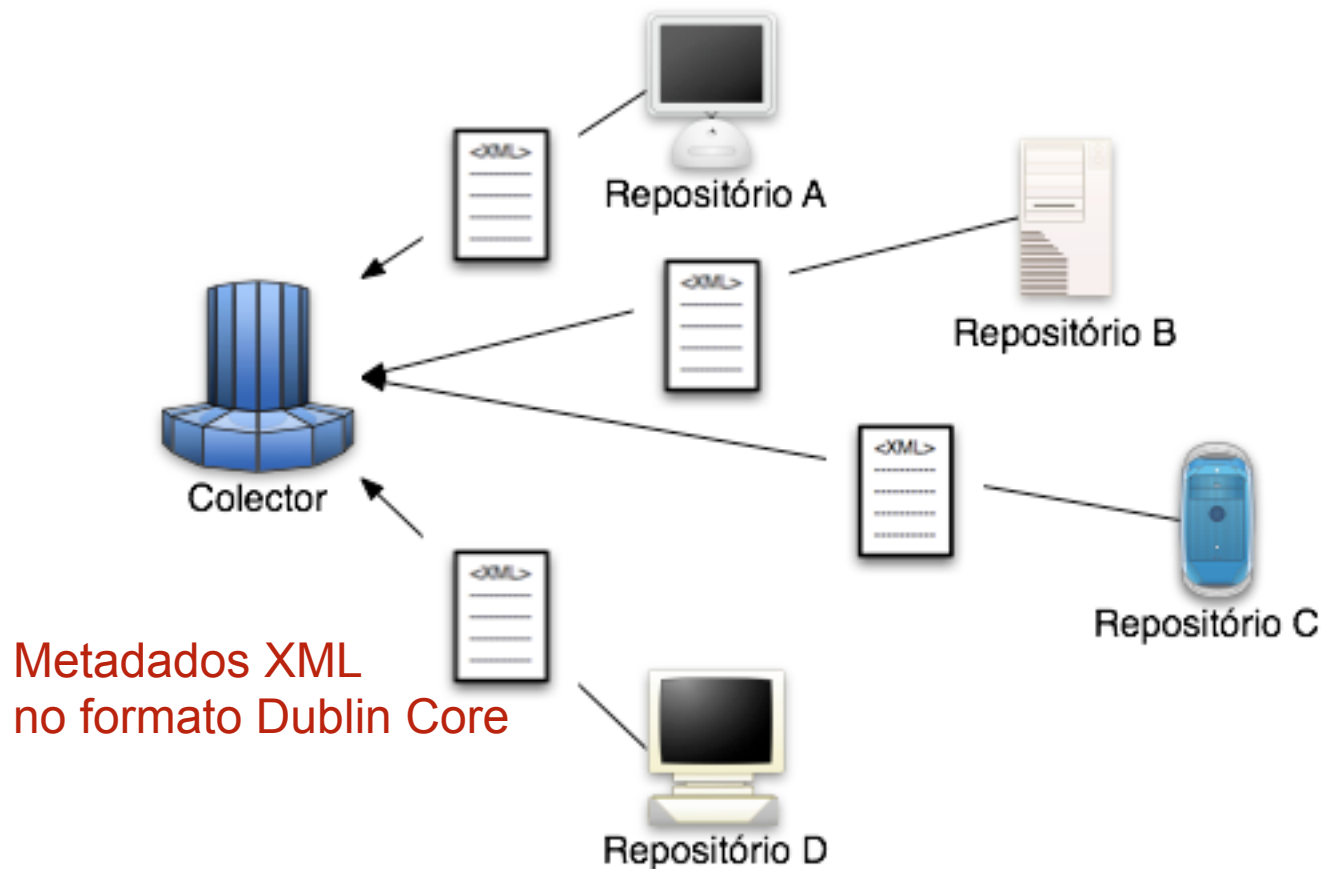
# Federação de Repositórios





Universidade  
do Minho

# Federação de Repositórios





Universidade  
do Minho

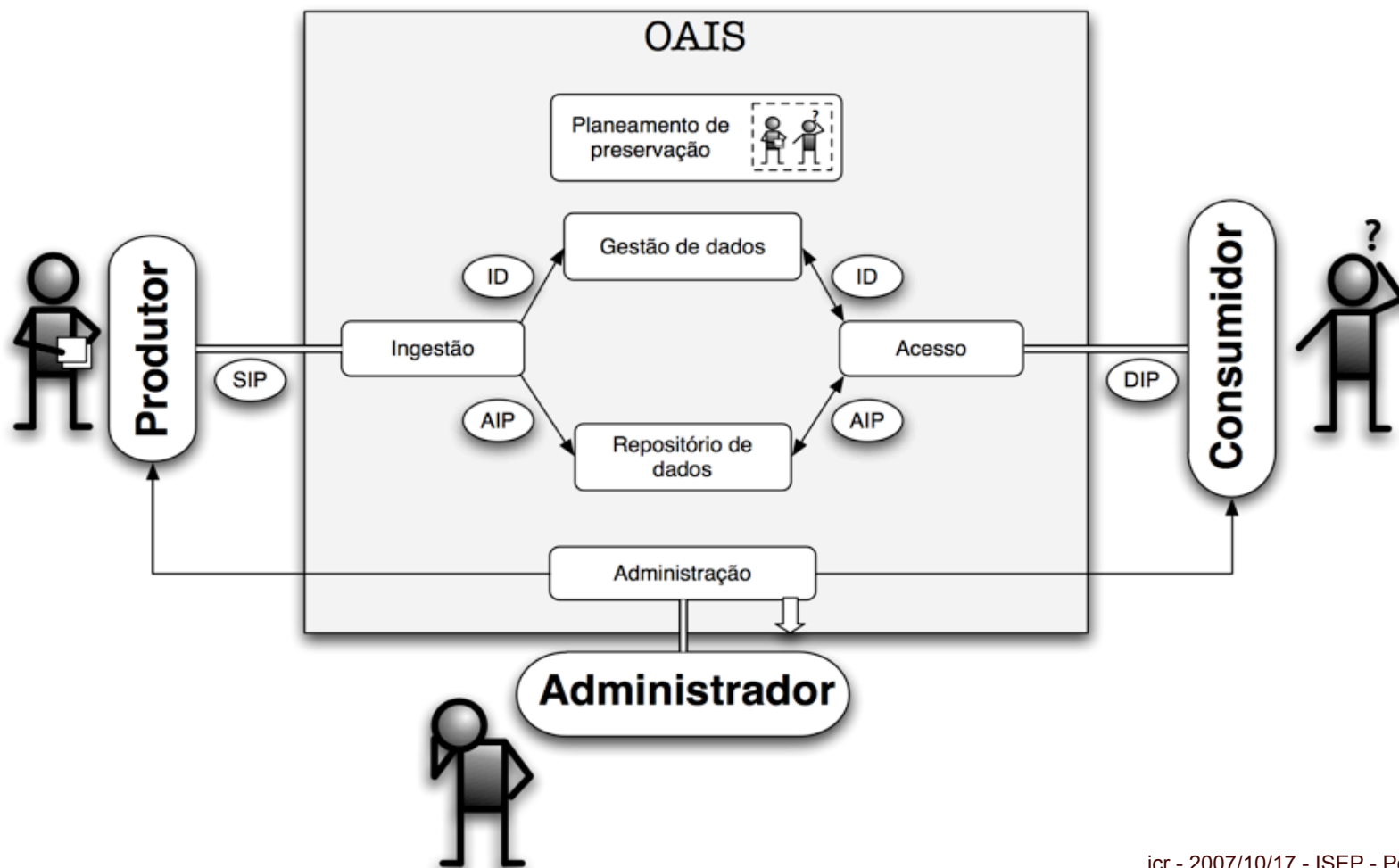
# Federação de Repositórios



- OAlster
  - Google
  - Yahoo!
- 
- CERN
  - Dspace
  - GNU Eprints
  - Hyperjournal
  - arXiv



# Voltando ao início: OAIS





# Áreas para Investigação

- Preservação digital
- Construção e Especificação de SIPs
- Modelos de Armazenamento para Repositórios
- Web Semântica e Federação de Repositórios



Universidade  
do Minho

Pesquisa de Informação e Inteligência Artificial

2007-10-17



# Questões?

---