

Universidade do Minho

Seminário
Governação na era digital

Sistemas de suporte à Governação Electrónica

Preservação a longo-prazo de informação digital

José Carlos Ramalho
jcr@di.uminho.pt

Miguel Ferreira
mferreira@dsi.uminho.pt
http://www.dsi.uminho.pt/~ferreira





27 de Fevereiro de 2008
Amarante

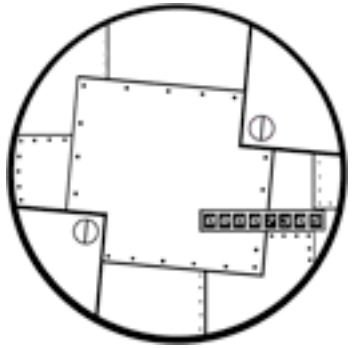
Conteúdo

- **Contextualização**
- **Projecto RODA**
- **Exemplos** de obsolescência tecnológica
- **Conceitos**/definições
- **Estratégias** de preservação
  - Preservação de tecnologia
  - Preservação do objecto conceptual

Passado recente

Digitarq (2003-presente)

- Gestão de metainformação (baseada em EAD)
- Gestão de objectos digitais



RODA (2006-2008)

- Gestão de metainformação
- Gestão de objectos digitais
- Protocolos e Políticas de Preservação Digital

CRAV: Sala de Leitura Virtual (2006-2008)

- Gestão de Pedidos
- Workflow documental

Contextualização

- Vivemos numa **era digital**!
  - Grande parte da **produção intelectual** é criada em **formatos digitais**
  - **Facilidade** de **criação** e **disseminação**
  - **Qualidade** dos resultados obtidos
- Administração pública
  - **Todos** os organismos da AP Central e Regional dispõem de ligação à Internet (87% destes em banda larga)
  - 89% destes possui presença na Web e
  - Disponibilizam serviços como:
    - informação sobre a instituição
    - formulários
    - bases de dados
    - bens our serviços em formato digital
    - etc.

Contextualização

- Rápida **obsolescência** da tecnologia
  - Embora um objecto digital possa ser **copiado infinitas vezes sem perder qualidade**
  - Depende de um **contexto tecnológico** para que possa ser interpretado
  - Os objectos acabam **prisioneiros** dos seus próprios **suportes** e/ou **formatos**

Contextos de preservação digital

- Domínio doméstico
  - **Avaria** no disco rígido
  - Novo computador não tem **drive de disquetes**
  - Serviço de **e-mail online** que desaparece
  - O **vídeo de casamento** está em VHS e já não temos vídeo
  - **Baby-blogs**, documentos em formatos obsoletos, ...
- Porta aviões USS Carrier
  - 1/2 milhão de diagramas (eléctricos, propulsão, etc...)
  - Após **actualização do software** de CAD
  - *"Wow, this doesn't look exactly like the drawing did before"*, Brad Cumming



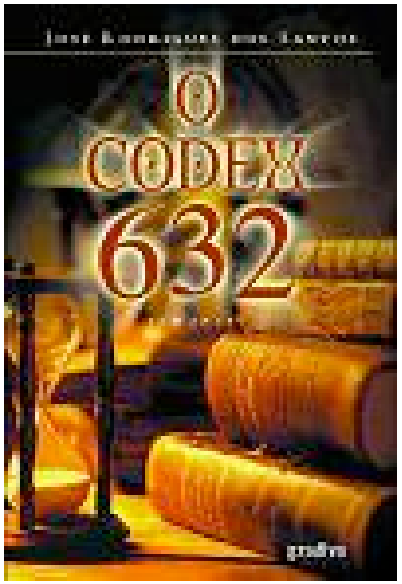
Contextos de preservação digital

- Domínio governamental
  - Adopção de **novos sistemas**
  - Preservação da **história**
  - Informação com **valor probatório**

Alguns requisitos

- How do we achieve Authenticity?
- How do we describe and classify DO?
- How can we implement digital preservation?

Autenticidade



"O Codex 632" de José Rodrigues dos Santos

Tema: Quem foi na realidade Cristophorus Colombus?

Italiano? Espanhol? Ou um português oriundo de uma família judia?

Autenticidade

É preciso confiar nas fontes : na História antiga não há discurso directo ou evidência.

EX: a Bíblia

Autenticidade

É preciso confiar nas fontes : na História antiga não há discurso directo ou evidência.

EX: a Bíblia

How do we become trustful?

Autenticidade

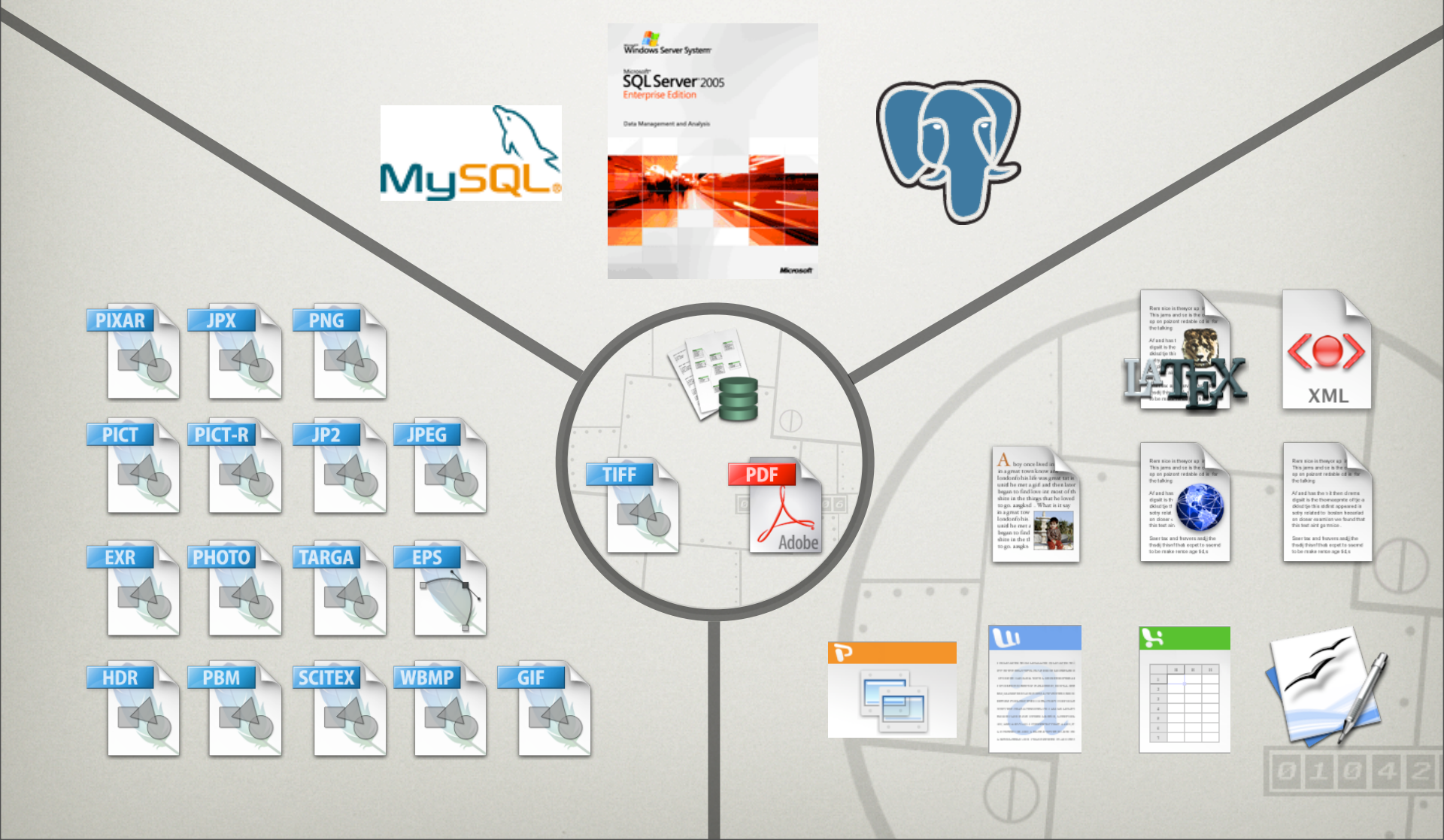
É preciso confiar nas fontes : na História antiga não há discurso directo ou evidência.

EX: a Bíblia

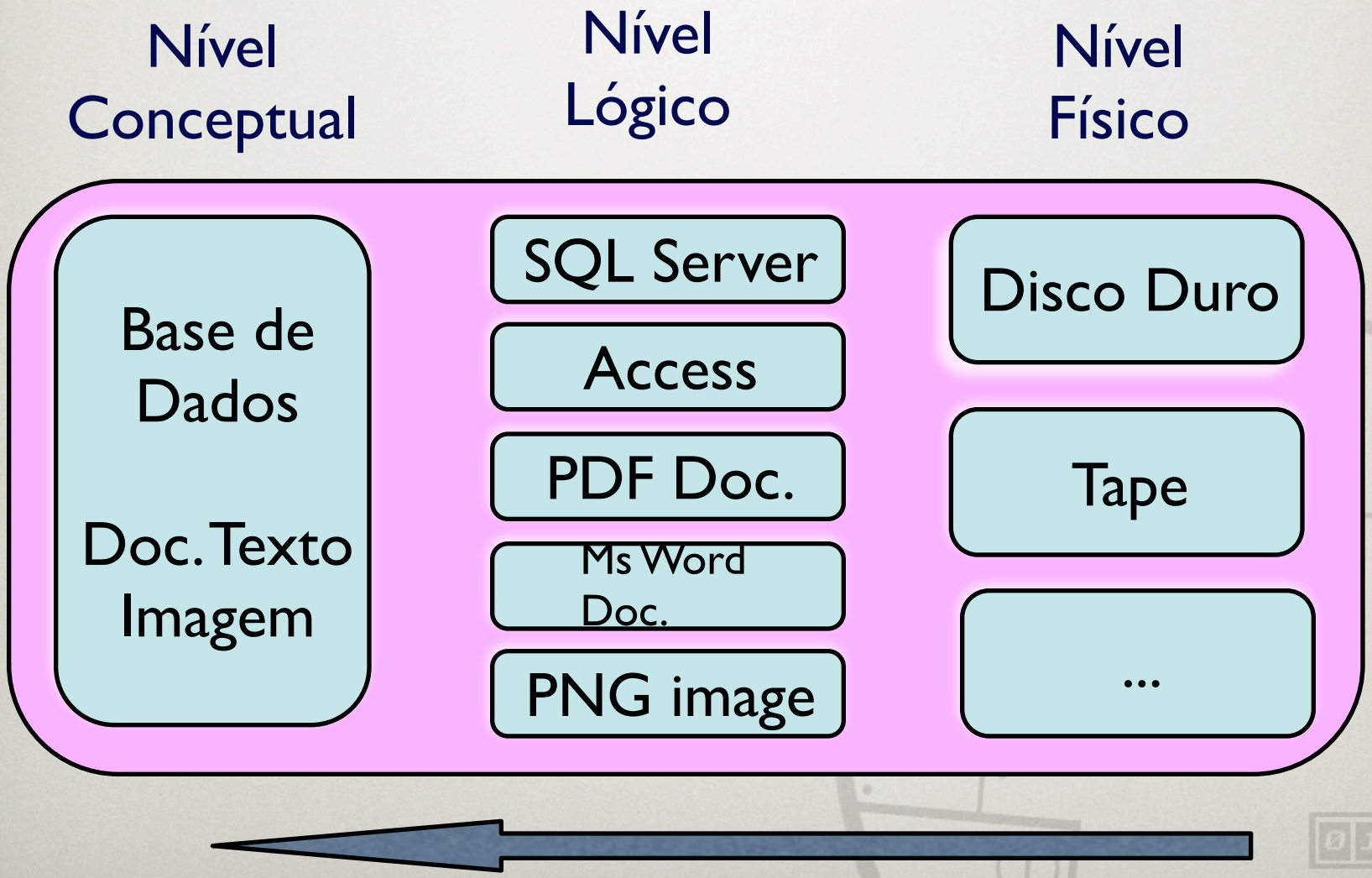
How do we become trustful?

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

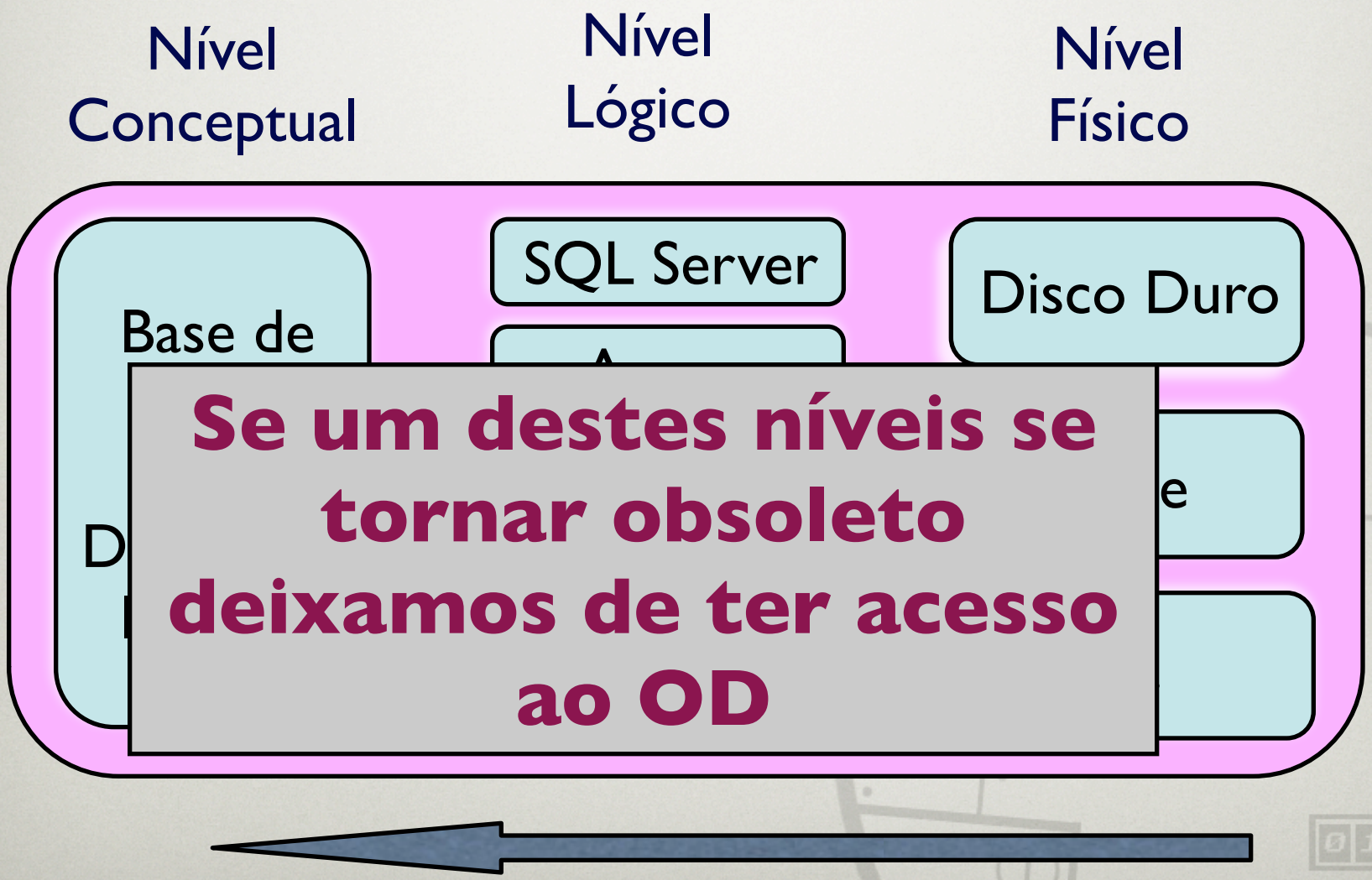
Classes de ODs



Anatomia de um OD



Anatomia de um OD



Preservação digital

Conjunto de

actividades ou **processos**

responsáveis por garantir o **acesso**

continuado a **longo-prazo** aos **objectos digitais** e restante **património** cultural existente em

formatos digitais sem que seja

comprometida a sua

autenticidade

Objecto, material ou recurso digital

Tudo aquilo que possa ser **representado através de uma sequência de dígitos binários**

- **Nascido digital**
- **Convertido** a partir de formatos analógicos (digitalizado)

- **Exemplos**:
  - bases de dados
  - folhas de cálculo
  - documentos de texto
  - páginas Web
  - correio electrónico
  - vídeo, audio, imagens
  - modelos 3D
  - software
  - etc…



Thibodeau, K., Overview of Thechnological Approaches to Digital Preservation and Challenges in Coming Years. in *The State of Digital Preservation: An International Perspective*, (Washington D.C., July 2002), Documentation Abstracts, Inc. - Institutes for Information Science.

Analógico vs Digital



TER 24 AGO

EDIÇÃO PORTO

PÚBLICO

www.publico.pt

SINÓNIMOS E ANTÓNIMOS

O Dicionário COM MAIS 10 VOLUMES

HOJE PRIMEIRO VOLUME POR 2,2 EUROS

TRÊS MILHÕES POR DISTRIBUIR ÀS VÍTIMAS DOS INCÊNDIOS DE 2003

Verbas das campanhas de solidariedade do ano passado podem ir para casos desse ano

OS GRANDES TEMAS DE AGOSTO DESCONHECIDOS NA HISTÓRIA

A joalheira de Phnom Penh

CULTURA

Polícia à procura do DNA dos ladrões dos quadros de Munch

Assessor de Rui Rio ganha mais que Jorge Sampaio

ECONOMIA

Finanças com dúvidas no reembolso de 40 milhões de IVA ao BCP

FUTEBOL

Gilberto Madail admite chamada de Derlei à selecção nacional

A festa de Holmes no Estádio Olímpico

Britânica Kelly Holmes acaba com reinado de Maria Mutola nos 800 metros

Francis Obikwelu recebeu a medalha de prata e diz-se cansado para os 200 metros

Edivaldo Monteiro sem ilusões para as meias-finais dos 400 metros barreiras

Como Angola defrontou os gigantes norte-americanos do basquetebol

Analógico vs Digital



TER 24 AGO

EDIÇÃO PORTO

PÚBLICO

www.publico.pt

SINÓNIMOS E ANTÓNIMOS

O Dicionário COM MAIS 10 VOLUMES

HOJE PRIMEIRO VOLUME POR 2,2 EUROS

TRÊS MILHÕES POR DISTRIBUIR ÀS VÍTIMAS DOS INCÊNDIOS DE 2003

Verbas das campanhas de solidariedade do ano passado podem ir para casos deste ano

A joalheira de Phnom Penh

CULTURA

Polícia à procura do DNA dos ladrões dos quadros de Munch

Assessor de Rui Rio ganha mais que Jorge Sampaio

ECONOMIA

Finanças com dúvidas no reembolso de 40 milhões de IVA ao BCP

FUTEBOL

Gilberto Madail admite chamada de Derlei à selecção nacional

1809

A festa de Holmes no Estádio Olímpico

Britânica Kelly Holmes acaba com reinado de Maria Mutola nos 800 metros

Francis Obikwelu recebeu a medalha de prata e diz-se cansado para os 200 metros

Edivaldo Monteiro sem ilusões para as meias-finais dos 400 metros barreiras

Como Angola defrontou os gigantes norte-americanos do basquetebol



Analógico vs Digital



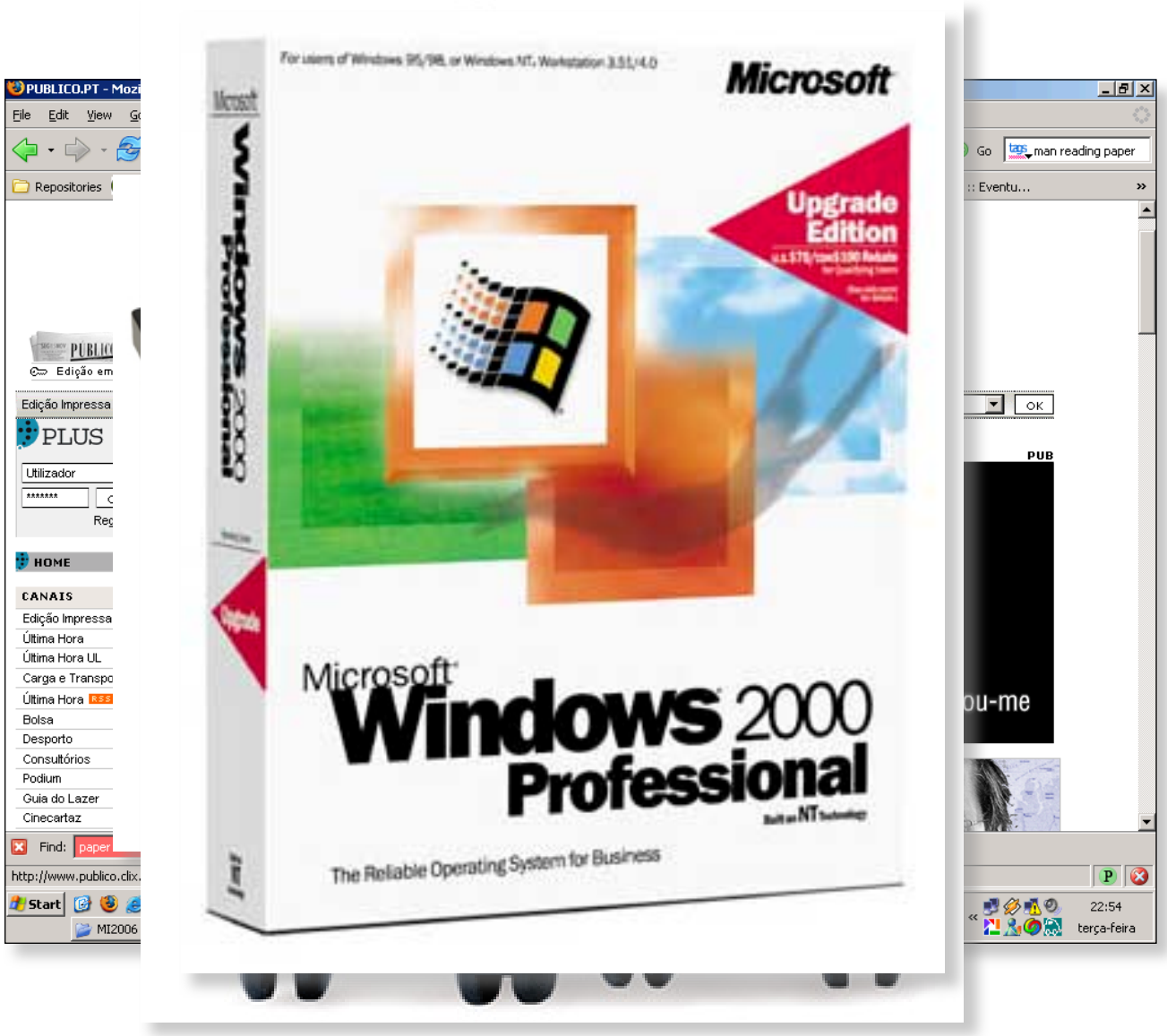
Analógico vs Digital



Analógico vs Digital



Analógico vs Digital



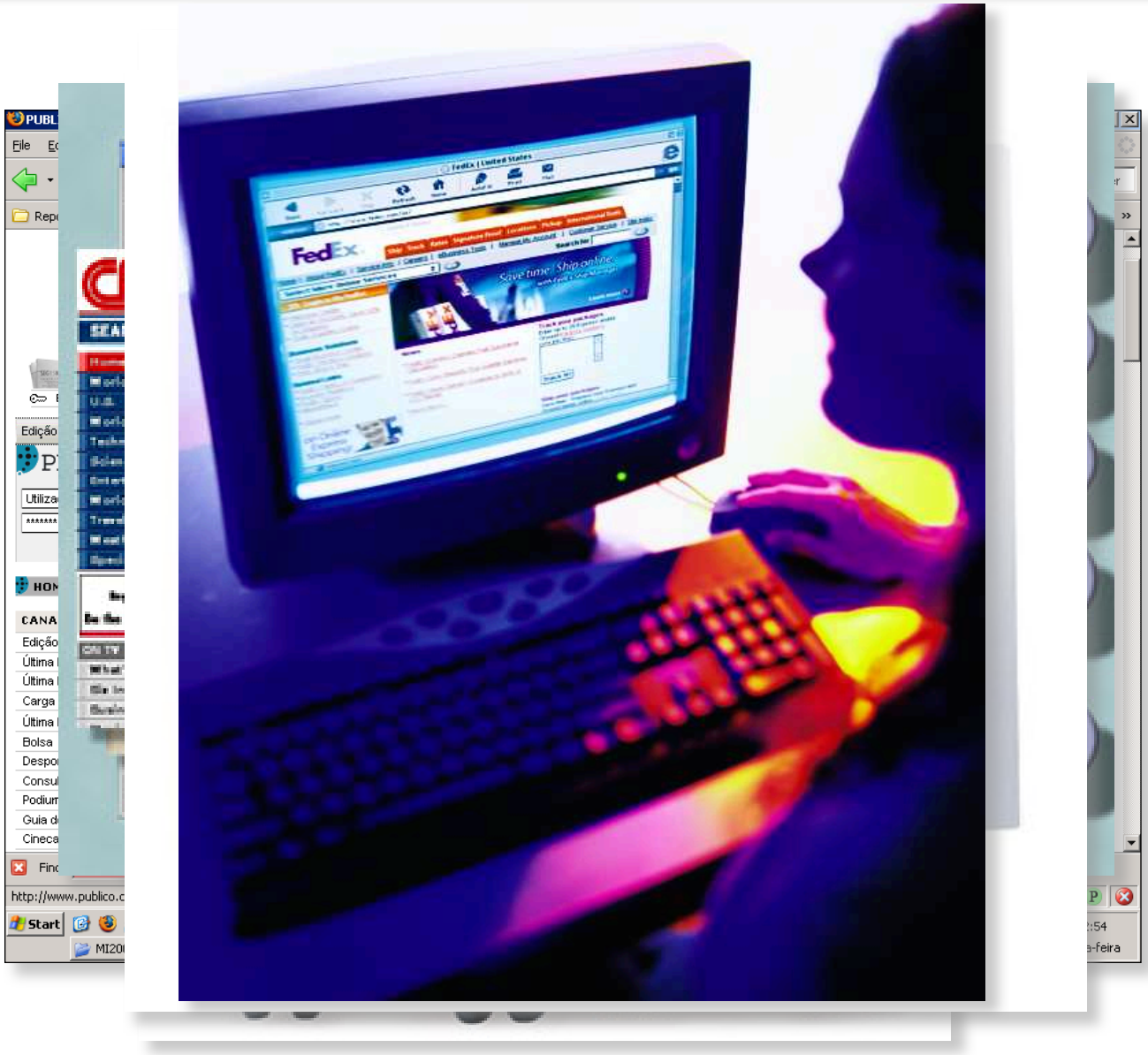
Analógico vs Digital



Analógico vs Digital



Analógico vs Digital



Estratégias de preservação





Preservação de tecnologia

- **Preservar** a informação nos seus **formatos** lógicos e/ou físicos **originais**
- Utilização da **tecnologia originalmente associada** a estes **objectos** para garantir o acesso aos mesmos
- Foco no objecto **físico/lógico**

Preservação da informação

- Preservar as **características essenciais** dos objectos de forma **independente** do hardware e do software
- Foco no objecto **conceptual**

Refrescamento

Transferência de **informação** de um **suporte físico** de armazenamento **para outro mais estável** antes que o primeiro se deteriore ou se torne irremediavelmente obsoleto



Não é tanto uma estratégia, é mais uma **precondição** necessária para a preservação!

Estratégias de preservação

Preservação de tecnologia



Preservação da informação



Estratégias de preservação

Preservação de tecnologia





Universidade do Minho

Seminário
Governação na era digital

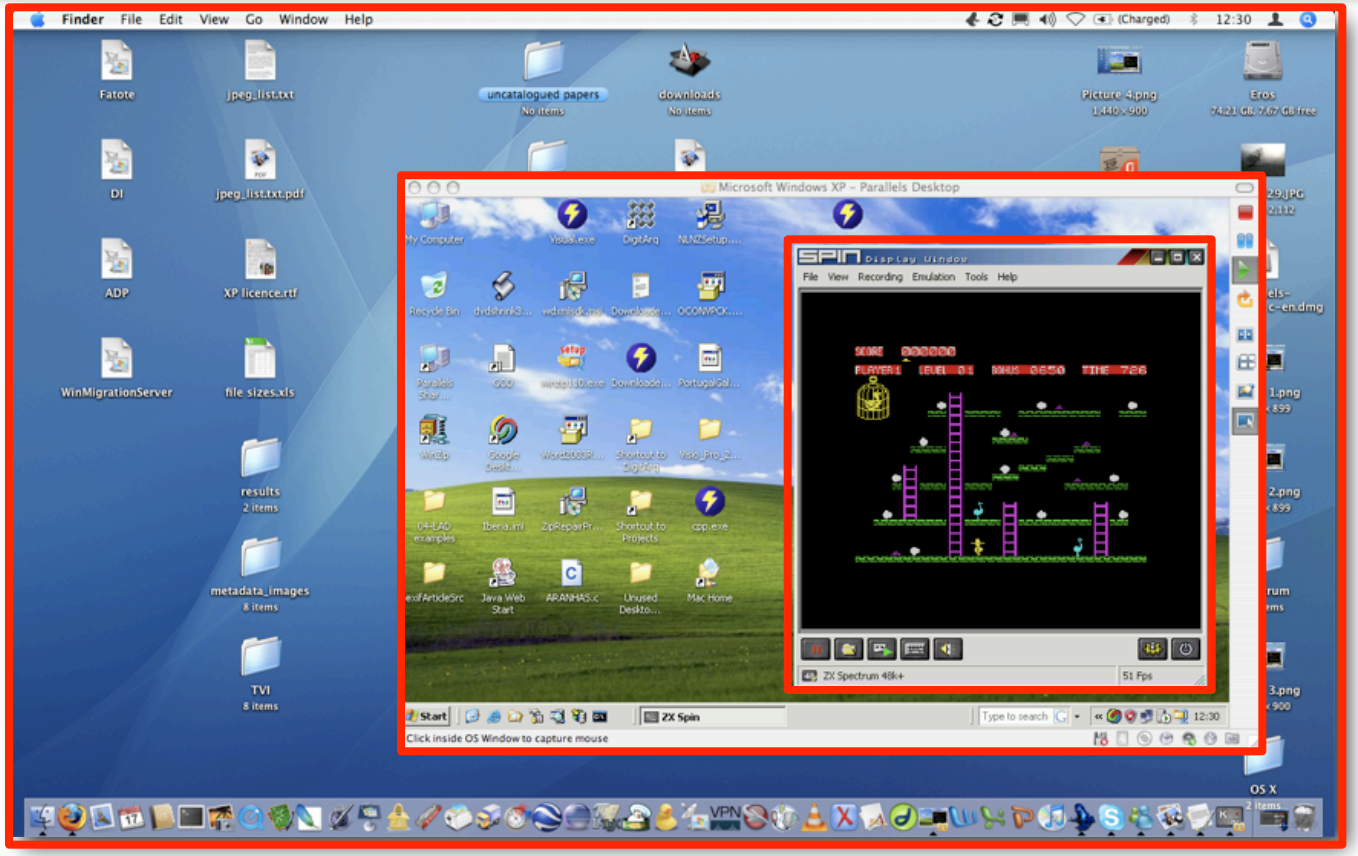
Preservação de tecnologia

- **Museus** de hardware
  - **Preservação** do **contexto** de criação/interpretação **original**
  - Dificuldade em **manter o hardware operacional** a longo-prazo (oneroso)
  - **Restrições** ao nível do **acesso** e **reutilização** da informação
- **Emuladores**
  - **Software** capaz de **reproduzir** o **comportamento** de uma **plataforma** de hardware e/ou software
  - A construção de emuladores é um processo complexo e muito dependente da publicação de informação por parte do fabricante de hardware ou software
  - Preservar um sistema operativo inteiro apenas para visualizar um documento parece uma tarefa exageradamente pesada
  - Não garante a reutilização da informação

Emulação [exemplo]



Emulação [exemplo]





Estratégias de preservação

Preservação de tecnologia



Preservação da informação



Estratégias de preservação

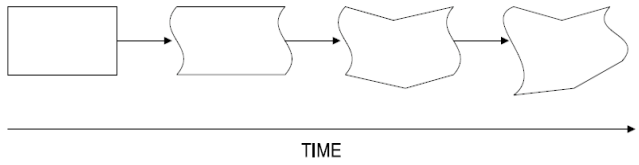
Preservação da informação



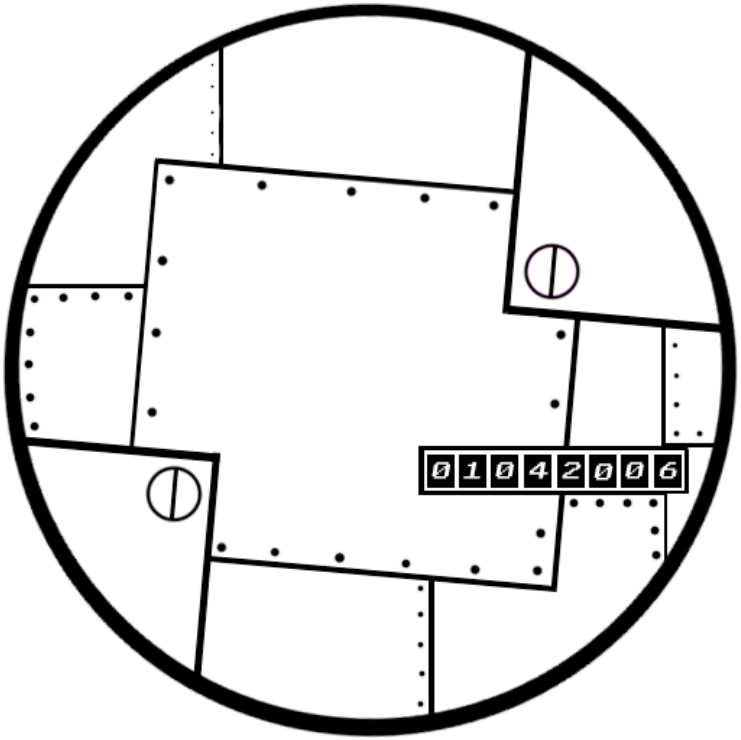
Migração de formatos

Transferência periódica de **material** digital de uma **configuração** de hardware/software para outra mais **actual**

- **Vantagens**
  - **Disseminação** dos materiais em **formatos actualizados** que os consumidores são capazes de reconhecer
  - **Sem necessidade** de preservar o **software original**
  - A estratégia **mais aplicada** na maioria dos domínios
    - e.g. doméstico, organizacional, etc.
- **Desvantagens**
  - Possível **perda de informação** durante a conversão
  - Necessidade de **diligência continua**
  - **Dispendiosa** a longo-prazo



RODA



Repositório de Objectos Digitais Autênticos

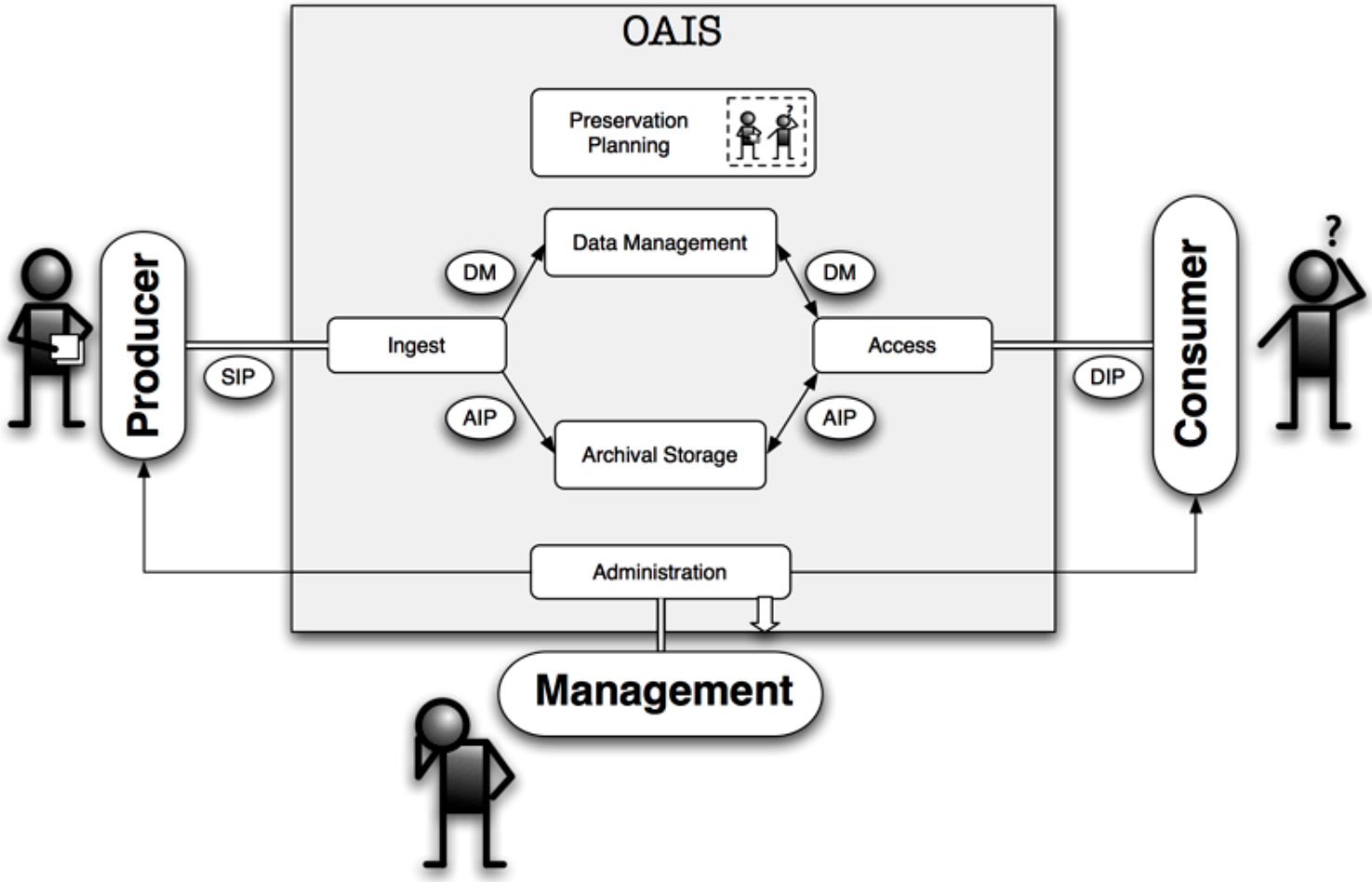
O que é?

- Projecto de I&D iniciado em 2006
  - Final do desenvolvimento previsto para fins de 2008
  - Parceria entre a DGARQ e a UMINHO
- Repositório digital
- Guardar, preservar, dar acesso a objectos digitais
  - Imagens,
  - Documentos de texto,
  - Bases de dados relacionais
- Preservação por longos períodos de tempo
  - Na ordem das dezenas de anos
- Assegurar o valor probatório da informação custodiada

A quem se destina?

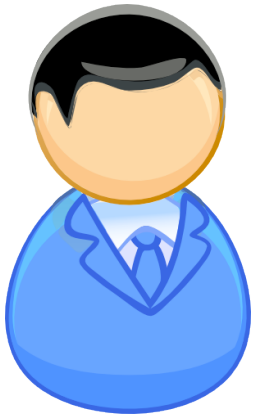
- Organismos **públicos**
  - clientes privilegiados
- Organismos **privados**
  - estabelecimento de contratos de incorporação
- **Arquivos pessoais** - Não nesta fase!
- Todos os organismos que queiram **preservar objectos digitais produzidos** no decorrer da sua **actividade**
  - Valor patrimonial
  - Valor probatório
  - Incorporação obrigatória

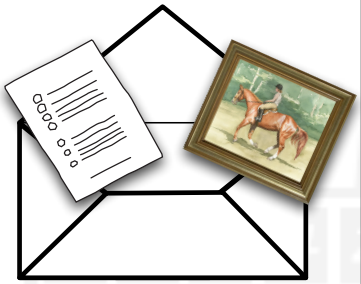
Arquitectura geral



Como funciona?

- Contrato pré-ingestão
  - Arquivo e o produtor acordam o formato do SIP (informação de depósito)
  - O processo de ingestão é completamente descrito e acordado entre ambas as partes
- Produtor prepara SIPs
  - Metainformação descritiva
  - Objectos digitais
  - Envelope XML (METS)





Estrutura do SIP (ex.)

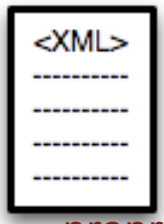


uma imagem

Estrutura do SIP (ex.)



uma imagem



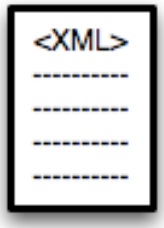
propriedades:

- data
- hardware
- ...

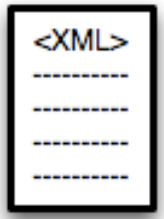
Estrutura do SIP (ex.)



uma imagem



Metainformação Técnica:
- cor
- dimensões
- ...

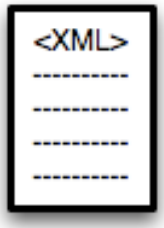


propriedades:
- data
- hardware
- ...

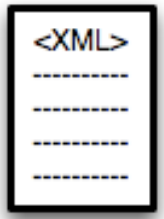
Estrutura do SIP (ex.)



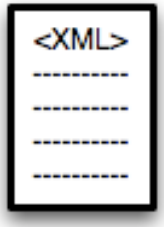
uma imagem



Metainformação Técnica:
- cor
- dimensões
- ...



propriedades:
- data
- hardware
- ...

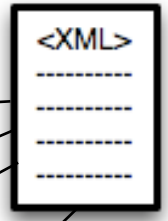


Metainformação Descritiva:
- produtor
- colecção
- ...

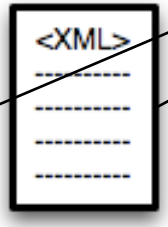
Estrutura do SIP (ex.)



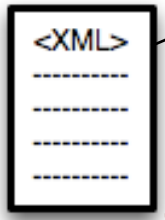
uma imagem



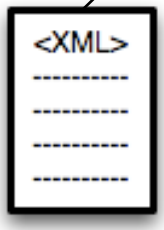
Manifesto



Metainformação Técnica:
- cor
- dimensões
- ...



propriedades:
- data
- hardware
- ...



Metainformação Descritiva:
- produtor
- colecção
- ...

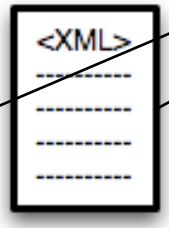
Estrutura do SIP (ex.)



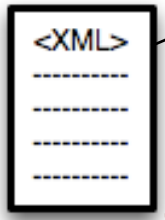
uma imagem



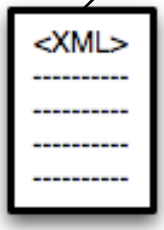
Manifesto



Metainformação Técnica:
- cor
- dimensões
- ...



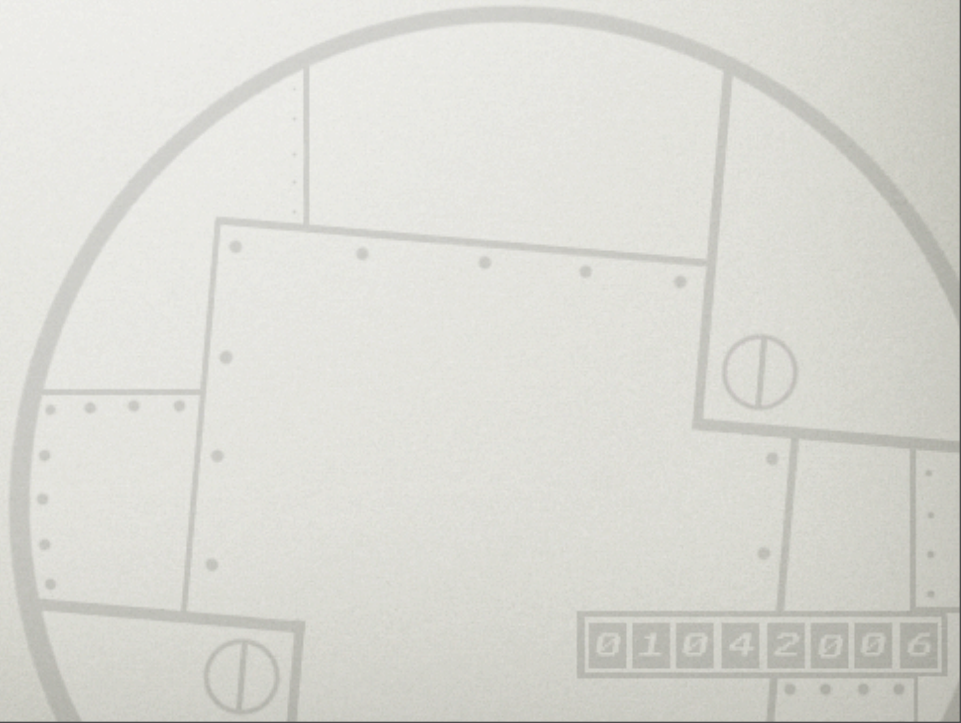
propriedades:
- data
- hardware
- ...



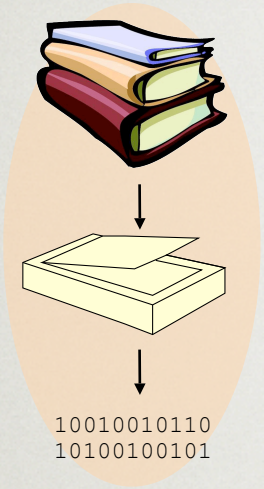
Metainformação Descritiva:
- produtor
- colecção
- ...

Ficheiro Compactado

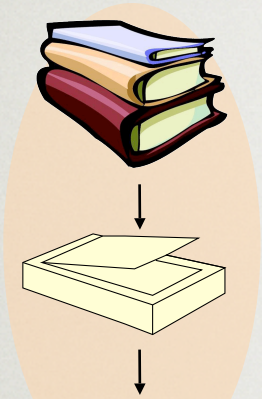
Estrutura do SIP

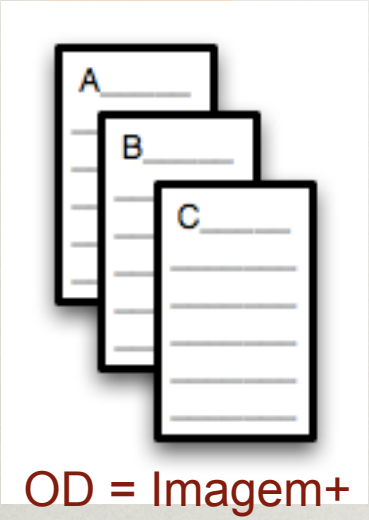


Estrutura do SIP



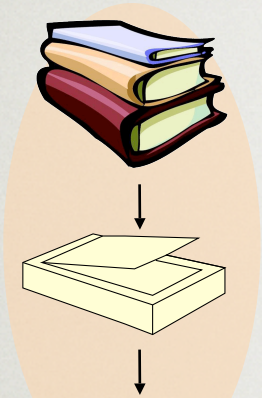
Estrutura do SIP

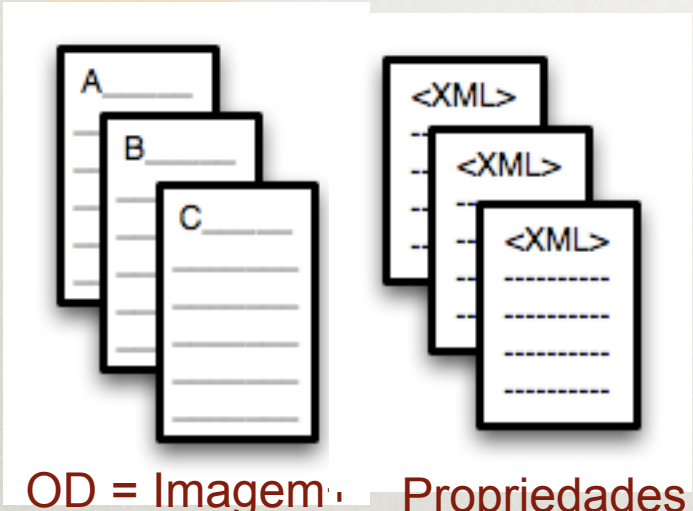




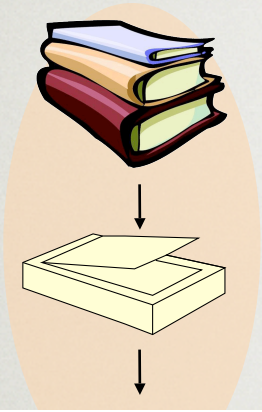
OD = Imagem+

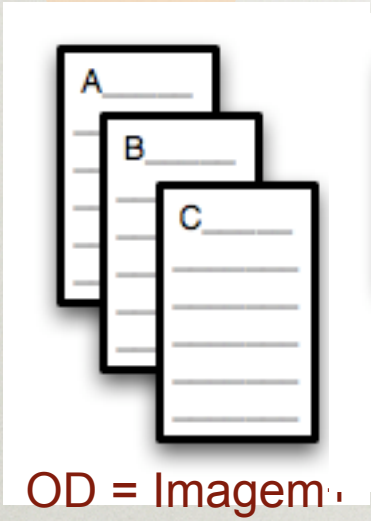
Estrutura do SIP



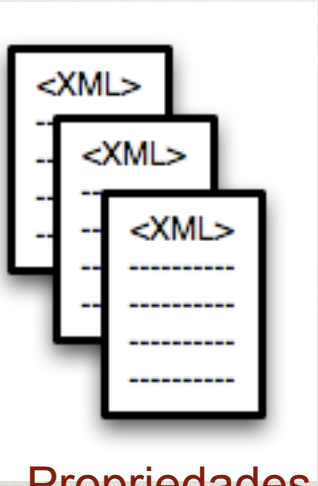


Estrutura do SIP

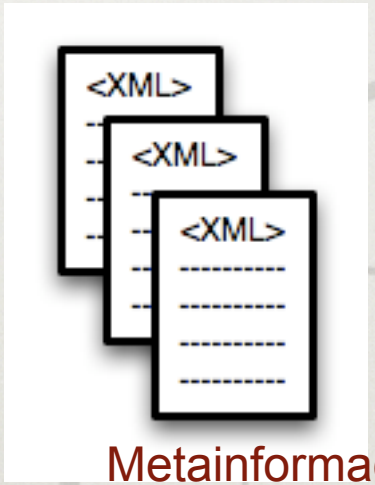




OD = Imagem

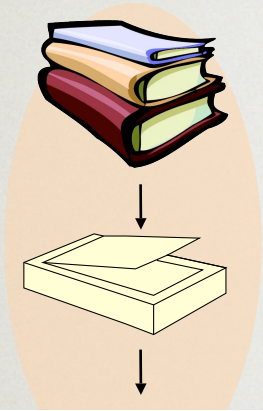


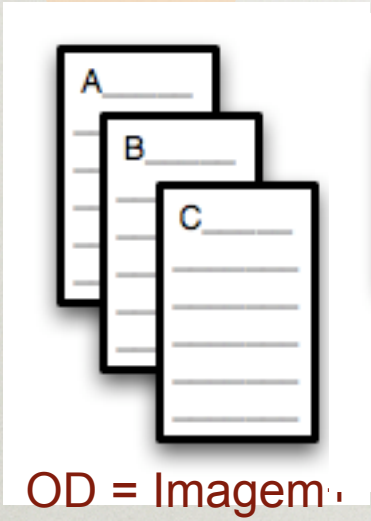
Propriedades



Metainformação Técnica

Estrutura do SIP





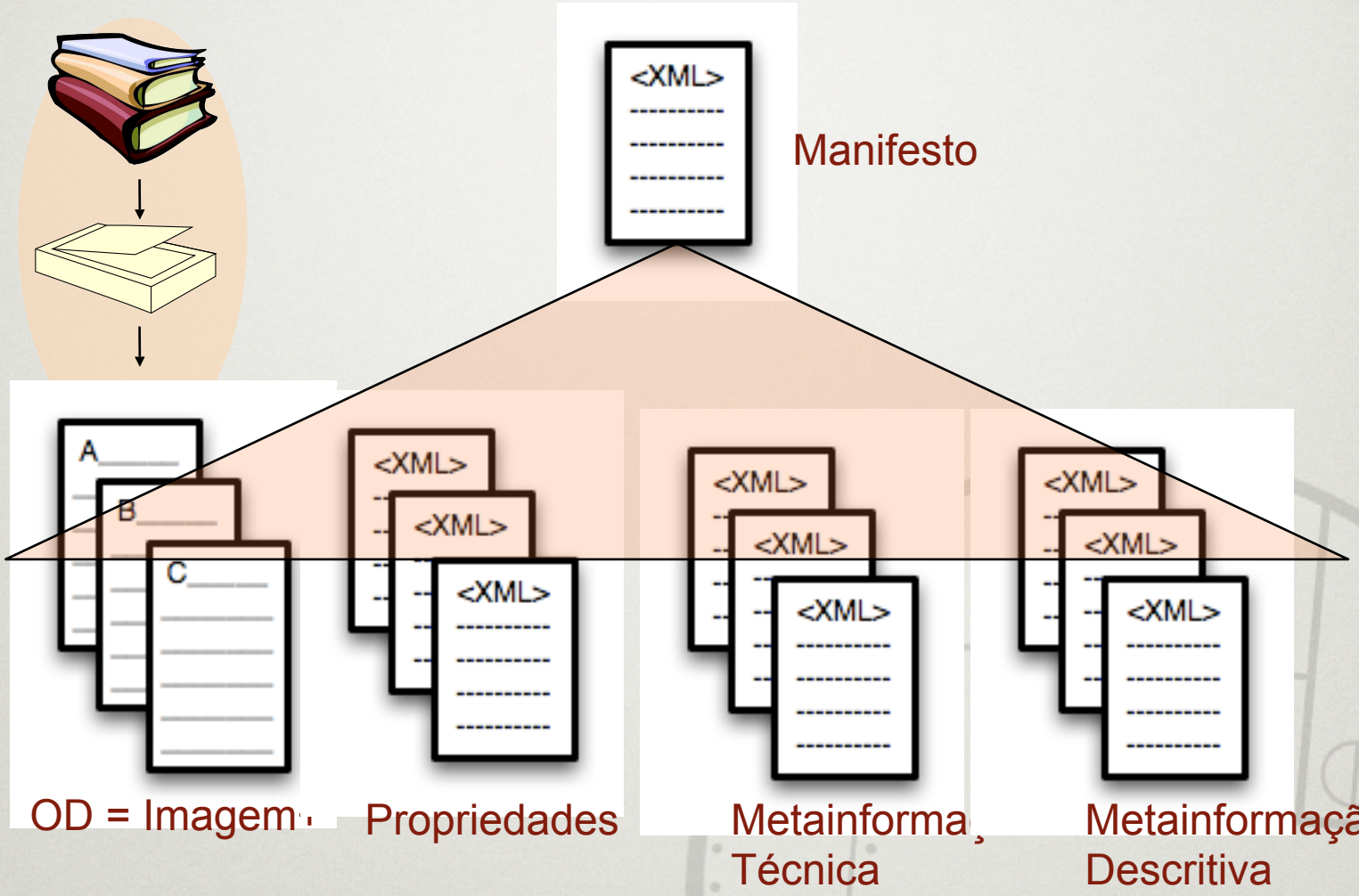
OD = Imagem

Propriedades

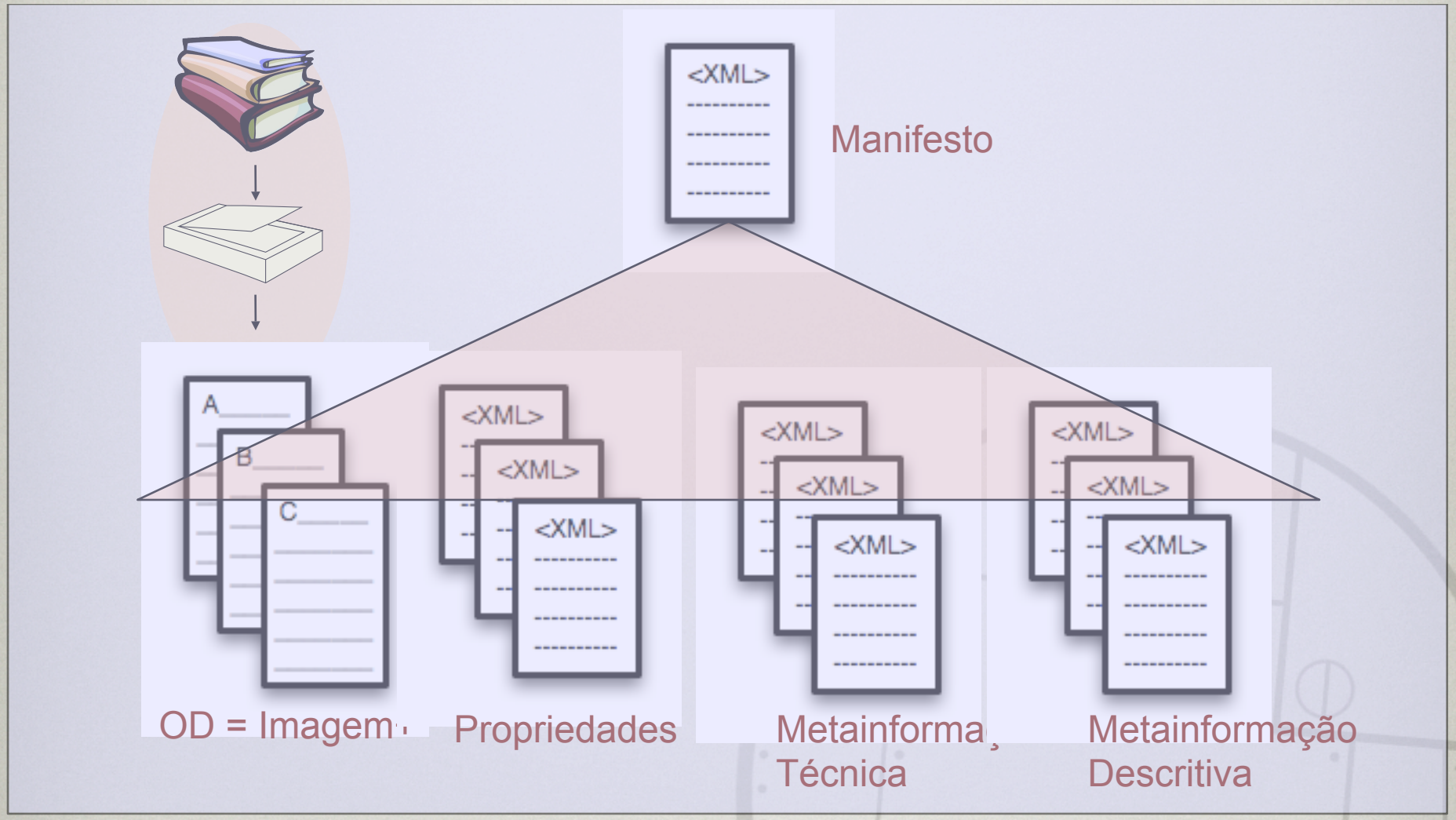
Metainforma
Técnica

Metainformação
Descritiva

Estrutura do SIP



Estrutura do SIP



Ficheiro Compactado

Ingestão

- Validação automática do SIP
  - testes.. (verificação de integridade)
  - testes.. (verificação de vírus)
  - testes.. (geração de PREMIS)
  - testes.. (normalização)
  - testes.. (geração de MI técnica)
  - testes.. (geração de PREMIS)
  - testes..
- Validação manual da SIP
  - tratamento arquivístico



Acesso

- Pesquisa básica
- Pesquisa avançada
- Navegação
- Visualização das representações
- Descarregamento

Processos de preservação

- Nível físico
  - Tecnologia de armazenamento de elevada disponibilidade baseada em NAS
  - Refrescamento de suportes (hot-swap)
  - Plano de contingência (recuperação perante desastres)
- Nível lógico
  - Replicação local (backups)
  - Replicação remota
- Nível conceptual/semântico
  - Normalização de formatos durante a ingestão
  - Migração de formatos obsoletos
  - Modelo económico



Universidade do Minho

Seminário
Governação na era digital

Processos de preservação

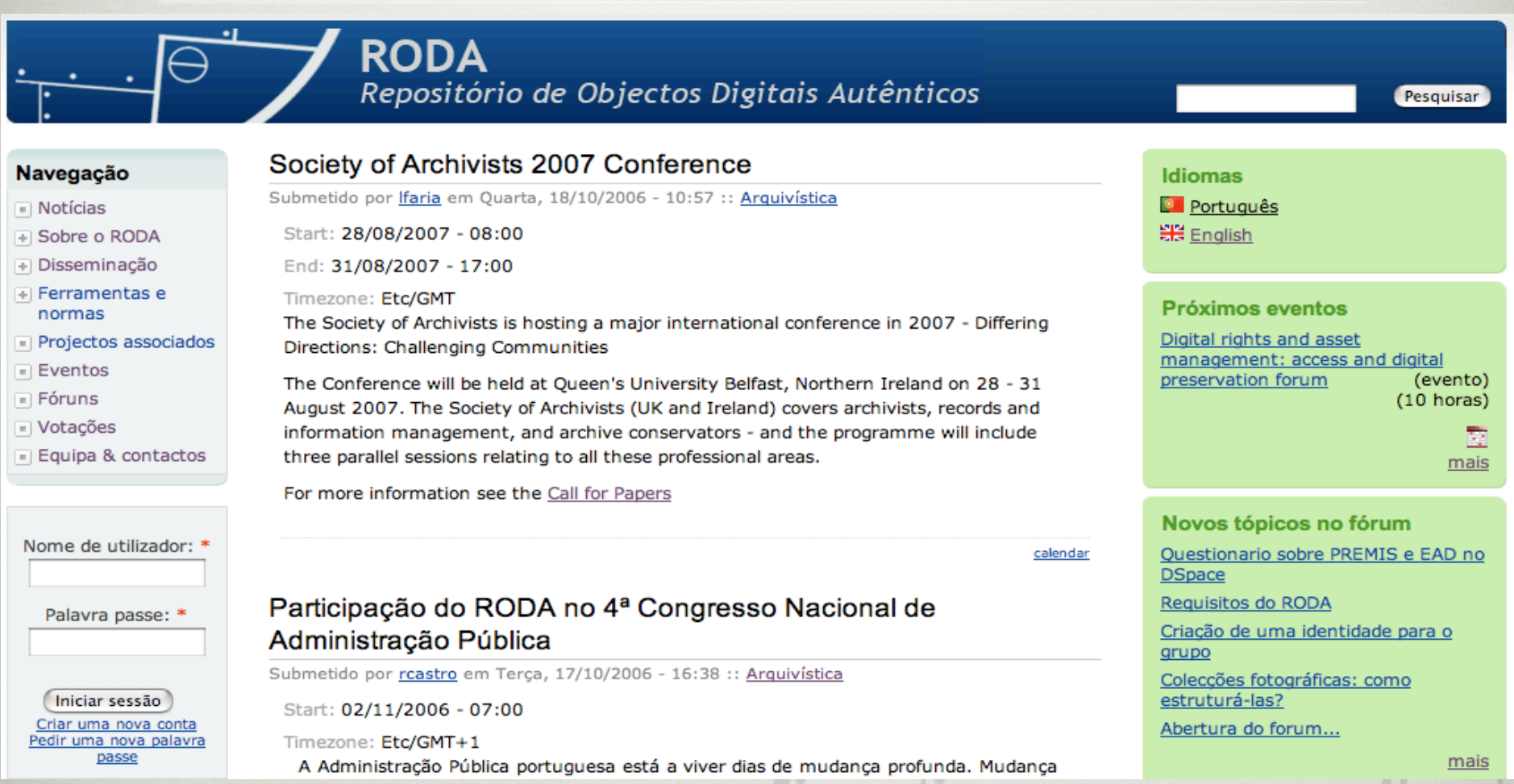
- Nível social
  - Todos os utilizadores têm de estar autenticados
  - Toda a sua actividade é registada
  - O sistema procura evitar a ocorrência de erros humanos
- Nível económico
  - Modelo de sustentabilidade económica para o repositório
  - Custos de preservação fazem parte do orçamento anual da DGARQ
- Nível económico
  - Definição de um plano de sucessão

Normas suportadas

- OAIS
- METS
- ISADg
- EAD
- NIZO
- TRAC
- DRAMBORA


Veja como nossos irmãos estão trabalhando para transferir nossos escritos para CD-ROM.

HTTP://RODA.IANTT.PT



... para mais informação...



Disponível em

http://hdl.handle.net/1822/5820